Gulf University for Science & Technology
Department of Economics & Finance
ECON 380: Business Statistics
Spring 2012

Dr. Fida Karam
Room N1-115
PHONE: 2530-7351
karam.f@gust.edu.kw

Problem Set I:
Descriptive Statistics and Normal Distribution

1. The following data were obtained for the number of minutes spent listening to recorded music for a sample of 5 individuals on one particular day: 10, 20, 12, 17, and 16. Compute the mean, the median, the variance, the standard deviation and the coefficient of variation.

   ANS:
   Mean: $\bar{x} = \frac{\Sigma x_i}{n} = \frac{75}{5} = 15$
   Median: *10, 12, 16, 17, 20*
   Median = 16 (middle value)
   Variance: $s^2 = 16$
   Standard Deviation: $s=4$
   Coefficient of variation: $((4/15)\times100)\% = 26.67\ \%$

2. Consider a sample with data values of 10, 20, 21, 17, 16, and 12. Compute the mean, the median, the variance, the standard deviation and the coefficient of variation.

   ANS:
   Mean: $\bar{x} = \frac{\Sigma x_i}{n} = \frac{96}{6} = 16$
   Median: *10, 12, 16, 17, 20, 21*
   Median $= \frac{16+17}{2} = 16.5$
   Variance: $s^2 = 18.8$
   Standard Deviation: $s=4.34$
   Coefficient of variation: $((4.34/16)\times100)\% = 27.125\ \%$

1 2 3 4 5

1 2 $\boxed{3\ 4}$ 5 6

2

3.5

3. The following data represent the daily demand (y in thousands of units) and the unit price (x in dollars) for a product.

x_i	4	6	11	3	16
y_i	50	50	40	60	30

a. Compute and interpret the sample covariance for the above data.
b. Compute the standard deviation for the daily demand.
c. Compute the standard deviation for the unit price.
d. Compute and interpret the sample correlation coefficient.

ANS:

$$\Sigma x_i = 40 \quad \bar{x} = \frac{40}{5} = 8 \quad \Sigma y_i = 230 \quad \bar{y} = \frac{230}{5} = 46$$

$$\Sigma(x_i - \bar{x})(y_i - \bar{y}) = -240 \quad \Sigma(x_i - \bar{x})^2 = 118 \quad \Sigma(y_i - \bar{y})^2 = 520$$

$$s_{xy} = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{n-1} = \frac{-240}{5-1} = -60$$

$$s_x = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{118}{5-1}} = 5.4314$$

$$s_y = \sqrt{\frac{\Sigma(y_i - \bar{y})^2}{n-1}} = \sqrt{\frac{520}{5-1}} = 11.4018$$

$$r_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{-60}{(5.4314)(11.4018)} = -0.969$$

There is a strong negative linear relationship.

4. The following data represent the daily supply (y in thousands of units) and the unit price (x in dollars) for a product.

x_i	6	11	15	21	27
y_i	6	9	6	17	12

a. Compute and interpret the sample covariance.
b. Compute and interpret the sample correlation coefficient.

ANS:

$\Sigma x_i = 80 \quad \bar{x} = \frac{80}{5} = 16 \quad \Sigma y_i = 50 \quad \bar{y} = \frac{50}{5} = 10$

$\Sigma(x_i - \bar{x})(y_i - \bar{y}) = 106 \quad \Sigma(x_i - \bar{x})^2 = 272 \quad \Sigma(y_i - \bar{y})^2 = 86$

$$s_{xy} = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{n-1} = \frac{106}{5-1} = 26.5$$

$$s_x = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{272}{5-1}} = 8.2462$$

$$s_y = \sqrt{\frac{\Sigma(y_i - \bar{y})^2}{n-1}} = \sqrt{\frac{86}{5-1}} = 4.6368$$

$$r_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{26.5}{(8.2462)(4.6368)} = 0.693$$

A positive linear relationship.

5. A simple random sample of 5 months of sales data provided the following information:

Month:	1	2	3	4	5
Units Sold:	94	100	85	94	92

$\bar{x}$

a. Calculate a point estimate of the population mean number of units sold per month.
b. Calculate a point estimate of the population standard deviation.

s

ANS:

a. $\bar{x} = \frac{\Sigma x_i}{n} = \frac{465}{5} = 93$

b.

	x_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
	94	*+1*	*1*
	100	*+7*	*49*
	85	*-8*	*64*
	94	*+1*	*1*
	92	*-1*	*1*
Totals	*465*	*0*	*116*

$$s = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{116}{4}} = 5.39$$

6. A Harris poll used a survey of 1008 adults to learn about how people feel about the economy. Responses were as follows:

595 adults	The economy is growing.
332 adults	The economy is staying about the same.
81 adults	The economy is shrinking.

Calculate point estimates of the following population parameters.

a. The proportion of all adults who feel the economy is growing.
b. The proportion of all adults who feel the economy is staying about the same.
c. The proportion of all adults who feel the economy is shrinking.

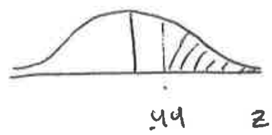
ANS:

a. *595/1008 = 0.59*
b. *332/1008 = 0.33*
c. *81/1008 = 0.08*

7. Given that z is a standard normal random variable, compute the following probabilities.

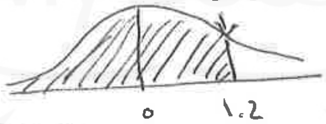
a. $P(0 \leq z \leq .83)$ = 0.2967
$P(0 \leq z \leq 1.57)$ = b. $P(-1.57 \leq z \leq 0)$ = 0.4418
c. $P(z > .44)$ 0.5 − .17 = 0.3300
d. $P(z \geq -.23)$ = $P(-0.23 \leq z \leq 0) + 0.5$
e. $P(z < 1.20)$
f. $P(z \leq -.71)$
= 0.091 + 0.5

c.



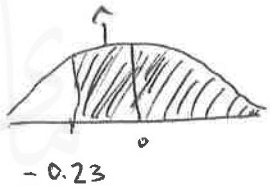
e

table



d.

table

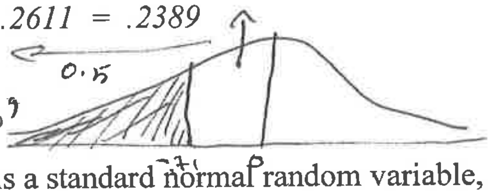


0.0910 + .5000 = .591

ANS:

a. *.2967*
b. *.4418*
c. *.5000 - .1700 = .3300*
d. *.0910 + .5000 = .5910*
e. *.3849 + .5000 = .8849* table
f. *.5000 - .2611 = .2389*

0.5 − 0.2611 = 0.2389



8. Given that z is a standard normal random variable, compute the following probabilities.

a. $P(-1.98 \leq z \leq .49)$
b. $P(.52 \leq z \leq 1.22)$
c. $P(-1.75 \leq z \leq -1.04)$

ANS:

a. $.4761 + .1879 = .6640$
b. $.3888 - .1985 = .1903$
c. $.4599 - .3508 = .1091$

9. Given that z is a standard normal random variable, find z_0 for each situation.
   a. $P(0 \leq z \leq z_0) = 0.4750$ (The area between 0 and z_0 is .4750).
   b. $P(0 \leq z \leq z_0) = 0.2291$ (The area between 0 and z_0 is .2291).
   c. $P(z \geq z_0) = 0.1314$ (The area to the right of z_0 is .1314).
   d. $P(z \leq z_0) = 0.6700$ (The area to the left of z_0 is .6700).

ANS:

a. *Using the table of areas for the standard normal probability distribution, the area of .4750 corresponds to* $z_0 = 1.96$.
b. *Using the table, the area of .2291 corresponds to* $z_0 = .61$.
c. *Look in the table for an area of* $.5000 - .1314 = .3686$. *This provides* $z_0 = 1.12$.
d. *Look in the table for an area of* $.6700 - .5000 = .1700$. *This provides* $z_0 = .44$.

μ

10. Attendance at a rock concert is normally distributed with a mean of 28,000 persons and a standard deviation of 4000 persons. What is the probability, that:
   a. At least 28000 persons will attend?
   b. Less than 14000 persons will attend?
   c. Between 17000 and 25000 persons will attend?
   d. Suppose the number who actually attended was X_1 and the probability of achieving this level of attendance or higher was found to be 5%. What is X_1?

ANS:

a. $P(X \geq 28{,}000) = P(z \geq 0) = 0.5.$
b. $P(X \leq 14{,}000) = P(z \leq -3.5) = 0.5 - 0.5 = 0.$
c. $P(17{,}000 \leq X \leq 25{,}000) = P(-2.75 \leq z \leq -0.75) = 0.497 - 0.2734 = 0.2236.$
d. $P(X \geq X_1) = 0.05;\ P(0 \leq z \leq z_1) = 0.5 - 0.05 = 0.45;\ z_1 = 1.645;\ X_1 = 34{,}580.$

11. The time a salesperson takes to travel from customer A to customer B varies but can be described by a normal probability function with mean 45 minutes and standard deviation 6 minutes.

μ

σ

a. What proportion of the journeys takes less than 35 minutes?
b. What proportion of the journeys takes over 60 minutes?
c. How long should the salesperson allow for a journey if they want to be 70 per cent sure of not being late?

ANS:

a. $P(X \leq 35) = P(z \leq -1.67) = 0.5 - 0.4525 = 0.0475.$
b. $P(X \geq 60) = P(z \geq 2.5) = 0.5 - 0.4938 = 0.0062.$
c. $P(X \leq X_1) = 0.7;\ P(0 \leq z \leq z_1) = 0.7 - 0.5 = 0.2;\ z_1 = 0.52;\ X_1 = 48.12.$

12. Assume that the test scores from a college admissions test are normally distributed, with a mean of 450 and a standard deviation of 100.
    a. What percentage of the people taking the test score between 400 and 500?
    b. Suppose someone receives a score of 630. What percentage of the people taking the test score better? What percentage score worse?
    c. If a particular university will not admit anyone scoring below 480, what percentage of the persons taking the test would be acceptable to the university?
    d. If the top 2.5% of test scores receive tuition discounts, what is the lowest score eligible for a discount?
    e. What are the minimum and the maximum values of the middle 95% of test scores?

ANS:

a. $P(400 \leq X \leq 500) = P(-0.5 \leq z \leq 0.5) = 0.1915 + 0.1915 = 0.383$ *or 38.3%.*
b. $P(X \geq 630) = P(z \geq 1.8) = 0.5 - 0.4641 = 0.0359$ *or 3.59%.*
$P(X \leq 630) = P(z \leq 1.8) = 0.5 + 0.4641 = 0.9641$ *or 96.41%.*
c. $P(X \leq 480) = P(z \leq 0.3) = 0.5 + 0.1179 = 0.6179.$ *38.21% are acceptable.*
Or $P(X \geq 480) = P(z \geq 0.3) = 0.5 - 0.1179 = 0.3821$ *or 38.21%.*
d. $P(X \geq X_1) = 0.025;\ P(z \geq z_1) = 0.025;\ z_1 = 1.96;\ X_1 = 646.$
e. $P(X_0 \leq X \leq X_1) = 0.95;\ P(-z_1 \leq z \leq z_1) = 0.95;\ ;\ z_1 = 1.96;\ X_0 = 254$ *and* $X_1 = 646.$

Gulf University for Science & Technology
Department of Economics & Finance
ECON 380: Business Statistics
Spring 2012

Dr. Fida Karam
Room N1-115
PHONE: 2530-7351
karam.f@gust.edu.kw

Problem Set II:

Descriptive Statistics, Normal Distribution, Chapters 7 and 8

Multiple choice questions:

n s^2

1. The variance of a sample of 60 observations equals 81. The standard deviation of the sample equals
   a. 1.37
   b. 9
   c. 81
   d. 1.17

   ANS: B

$S = \sqrt{81} = 9$ ✓

2. The heights (in cm) of a sample of 30 individuals were recorded and the following statistics were calculated.

mean = 165	range = 60
median = 155	variance = 81

The coefficient of variation equals
   a. 0.0545%
   b. 5.45%
   c. 49.09%
   d. 0.4909%

   ANS: B

$s = 9$

3. The standard deviation of a sample was reported to be 9. The report indicated that $\sum (x - \bar{x})^2 = 810$. What has been the sample size?
   a. 16
   b. 9
   c. 10
   d. 11

   ANS: D

$$s^2 = \frac{\Sigma (x_i - x)^2}{n-1}$$

$$9^2 = \frac{810}{n-1} \Rightarrow n - 1 = \frac{810}{81}$$

$$n = 10 + 1 = 11$$

Exhibit 1
A researcher has collected the following sample data

5 12 6 7 8 5 , 5 5 6 7 8 12

4. Refer to Exhibit 1. The median is
   a. 6
   b. 6.5
   c. 7
   d. 8

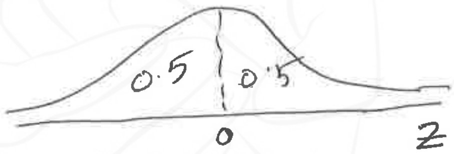
$\frac{6+7}{2} = 6.5$

ANS: B

5. Refer to Exhibit 1. The mean is
   a. 43
   b. 6.5
   c. 7.17
   d. 8

ANS: C

6. For a standard normal distribution, the probability of $z \le 0$ is
   a. Zero
   b. -0.5
   c. 0.5
   d. ~~One~~



ANS: C

7. For a standard normal distribution, the probability of $z \ge 0$ is
   a. Zero
   b. -0.5
   c. 0.5
   d. One

ANS: C

8. Z is a standard normal random variable. The $P(-2 \le Z \le -1.51)$ equals
   a. 0.4772
   b. 0.0427
   c. 0.4345
   d. 0.0228

ANS: B

μ σ

9. X is a normally distributed random variable with a mean of 19 and a standard deviation of 7. The probability that X is less than 5 is
   a. 0.5
   b. 0.5228
   c. 0.0228
   d. 0.4772

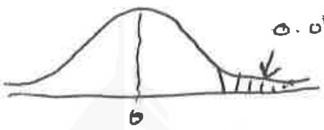
   ANS: C

شنو عكس

$0.5 - P(Z \le \frac{5-7}{})$

10. Given that Z is a standard normal random variable, what is the value of Z if the area to the right of Z is 0.0166?
    a. 0.4834
    b. 2.13
    c. 0.0000
    d. 1.96

    ANS: B



$P(0 \le Z \le Z_0)$
$= 0.5 - 0.0166$
$= 0.4834$
$Z_0 = 2.13$

11. Given that Z is a standard normal random variable. What is the value of Z if the area between -Z and Z is 0.9198?
    a. ± 1.75
    b. ± 1.96
    c. ± 2.0
    d. ± 11.6

    ANS: A

$P(0 \le Z \le Z_0) = \frac{0.9198}{2} = 0.4599$

~~= 0.4599~~

$Z_0 = \pm 1.75$

12. Given that Z is a standard normal random variable, what is the value of Z if the area to the right of Z is 0.996?
    a. 0.496
    b. -2.65
    c. +2.65
    d. Zero

    ANS: B

$P(Z_0 \le Z \le 0) = 0.996 - 0.5$
$= 0.496$
$Z_0 = -2.65$

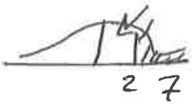
13. A simple random sample of 25 observations was taken from a large population. The sample mean and the standard deviation were determined to be 50 and 40 respectively. The standard error of the mean is
    a. 1.875
    b. 40
    c. 10
    d. 8

    ANS: D

$n = 35$ $\bar{x}$ $s = 40$

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

$s_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{40}{\sqrt{25}} = 8$



14. A population has a mean of 50 and a standard deviation of 6. A sample of 36 observations will be taken. The probability that the sample mean will be larger than 52 is
    a. 0.5228
    b. 0.9772
    c. 0.4772
    d. 0.0228

    ANS: D

$\mu = 50$ $\sigma = 6$, $n = 36$, $P(\bar{X} \geq 52)$

$z = \frac{\bar{x} - \mu}{\sigma_{\bar{x}}}$

$P(\bar{X} \geq 52) = P\left(\frac{\bar{x} - \mu}{\sigma_{\bar{x}}} \geq \frac{52-50}{1}\right)$

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{6}{\sqrt{36}} = 1$

$= P(Z \geq 2) = 0.5 - 0.4772 = 0.0228$

15. A population has a mean of 80 and a standard deviation of 21. A sample of 49 observations will be taken. The probability that the sample mean will be between 83 and 86 is
    a. 0.1359
    b. 0.8185
    c. 0.3413
    d. 0.4772

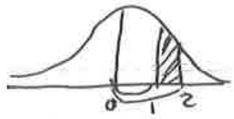
    ANS: A

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{21}{\sqrt{49}} = 3$

$\mu = 80$ $\sigma = 21$ $n = 49$

$P(83 \leq \bar{X} \leq 86) = P\left(\frac{83-80}{3} \leq \frac{\bar{x} - \mu}{\sigma_{\bar{x}}} \leq \frac{86-80}{3}\right) = P(1 \leq Z \leq 2)$

$= 0.4772 - 0.3413 = 0.1359$



16. A population has a standard deviation of 14. A random sample of 49 items from this population is selected. The sample mean is determined to be 50. At 95% confidence, the margin of error is
    a. 5
    b. 3.92
    c. 2
    d. 27.44

    ANS: B

$\sigma = 14$ $n = 49$ $\bar{X} = 50$

$M.E = Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} = 1.96 \frac{14}{\sqrt{49}} = 3.92$

17. In order to determine an interval for the mean of a population with unknown population standard deviation a sample of 81 items is selected. The mean of the sample is determined to be 25. The number of degrees of freedom for reading the t value is
    a. 1.96
    b. 81
    c. 80
    d. 24

    ANS: C

$\bar{X} = 25$ $n = 81$ $df = 81 - 1 = 80$

$df = n - 1$

18. If an interval estimate is said to be constructed at the 90% confidence level, the confidence coefficient would be
    a. 0.1
    b. 0.95
    c. 0.9
    d. 0.05

    ANS: C

19. A sample of 81 elements from a population with a standard deviation of 27 is selected. The sample mean is 80. The 99% confidence interval for μ is

a. 105.0 to 225.0
b. 175.0 to 185.0
c. 74.12 to 85.88
d. 72.272 to 87.728

ANS: D

$n = 81 \quad \sigma = 27 \quad \bar{X} = 80$

99%

$\mu = 80 \pm 2.576 \frac{27}{\sqrt{81}}$

$= 80 \pm 7.728$ or 72.272 to 87.728

20. A random sample of 25 students at a university showed an average age of 23 years and a sample standard deviation of 2 years. The 95% confidence interval for the true average age of all students in the university is

a. 22.084 to 23.916
b. 3.7744 to 5.4256
c. 22.174 to 23.826
d. 20.0 to 30.0

ANS: C

$n = 25$ ② $\quad \bar{X} = 23 \quad S = 2$ ① ③

95%

$1 - \alpha = 0.95$

$\alpha = 0.05$

$\frac{\alpha}{2} = 0.025$

t

$\mu = \bar{X} \pm t_{\alpha/2} \frac{S}{\sqrt{n}}$

$df = 25 - 1 = 24$

$t = 2.0639$

$23 \pm 2.0639 \frac{2}{\sqrt{25}}$

$= 23 \pm 0.826$

or 22.147 to 23.0826

21. A random sample of 81 statistics examinations was taken. The average score, in the sample, was 70 with a variance of 9. The 98% confidence interval for the average examination score of the population is

a. 69.34 to 70.65
b. 77.40 to 86.60
c. 69.22 to 70.78
d. 68.00 to 100.00

ANS: C

$n = 81 \quad \bar{X} = 70 \quad s^2 = 9 \Rightarrow s = \sqrt{9} = 3$

98%

$1 - \alpha = 0.98 \Rightarrow \alpha = 0.02 \Rightarrow \alpha/2 = 0.01$

$df = 81 - 1 = 80 \quad t_{\frac{\alpha}{2}} = 2.3264$

$\mu = \bar{X} \pm t_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}} = 70 \pm 2.3264 \frac{3}{\sqrt{81}}$

22. It is known that the population variance equals 90. With a 0.95 probability, the sample size that needs to be taken if the desired margin of error is 3 or less is

a. 90
b. 38
c. 38.146
d. 39

ANS: D

$\sigma^2 = 90 \quad E = 3$

$n = \left(\frac{Z_{\frac{\alpha}{2}} \sigma}{M.E}\right)^2 = \frac{Z_{\alpha/2}^2 \sigma^2}{M.E^2} = \frac{1.96^2 (90)}{3^2} = 38.146$

$= 39$

23. In a random sample of 81 observations, p= 0.3. The 97.8% confidence interval for π is

a. 0.122 to 0.278
b. 0.2002 to 0.3998
c. 0.134 to 0.266
d. 0.1855 to 0.4145

ANS: D

$n = 81 \quad p = 0.3 \quad 97.8\%$

$\pi = p \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}}$

$= 0.3 \pm 2.29 \sqrt{\frac{0.3(0.7)}{81}}$

$= 0.3 \pm 0.1145$

or 0.1855 to 0.4145

$P(0 \le Z \le Z_0) = \frac{0.978}{2} = 0.489$

$z = 2.29$

24. A machine that produces a major part for an airplane engine is monitored closely. With a .90 probability, the sample size that needs to be taken if the desired margin of error to build a confidence interval for the proportion of defective parts is .07 or less is

a. 138.06 90% $M.E = 0.07$ $n = \frac{Z^2_{\alpha/2} P^*(1-P^*)}{M.E^2}$
b. 138
c. 140 $P^* = 0.5$
d. 139 $= \frac{1.645^2 (0.5)(1-0.5)}{0.07^2}$

ANS: D

$= 138.06 = 139$

25. In a sample of 400 voters, 40 indicated they do not favor the incumbent governor. The 95% confidence interval for the proportion of voters favoring the incumbent is

a. 0.871 to 0.929 $n = 400$ $P = \frac{400-40}{400} = 0.9$
b. 0.120 to 0.280
c. 0.765 to 0.835
d. 0.071 to 0.129

ANS: A

$\pi = P \pm Z_{\alpha/2} \sqrt{\frac{P(1-P)}{n}}$

$= 0.9 \pm 1.96 \sqrt{\frac{0.9(0.1)}{400}}$

$= 0.9 \pm 0.0294$

or 0.8706 to 0.9294

Problem 1:

You are given the following information obtained from a sample of 5 observations taken from a population that has a normal distribution:

10 25 17 39 21

Develop a 95% confidence interval estimate for the population mean.

ANSWER:

$\bar{X} = 22.4$

$S^2 = 116.8$

$t = 2.7765$

Confidence Interval for μ*: 8.98 to 35.82.*

Problem 2:

The average life expectancy in France is 75 with a standard deviation of 5 years. A random sample of 64 individuals is selected.

a. What is the probability that the sample mean will be larger than 76 years?
b. What is the probability that the sample mean will be less than 73 years?
c. What is the probability that the sample mean will be between 74 and 77 years?
d. What is the probability that the sample mean will be between 73 and 74 years?
e. What is the probability that the sample mean will be larger than 75.5 years?
f. What is the probability that the sample mean will be within ± 1 of the population mean?

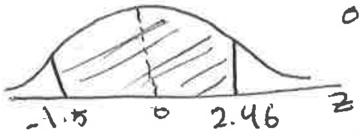
ANSWER:

a. $P(\bar{X} \geq 76) = P(z \geq 1.6) = 0.5 - 0.4452 = 0.0548.$
b. $P(\bar{X} \leq 73) = P(z \leq -3.2) = 0.5 - 0.5 = 0.$
c. $P(74 \leq \bar{X} \leq 77) = P(-1.6 \leq z \leq 3.2) = 0.4452 + 0.5 = 0.9452.$
d. $P(73 \leq \bar{X} \leq 74) = P(-3.2 \leq z \leq -1.6) = 0.5 - 0.4452 = 0.0548.$
e. $P(\bar{X} \geq 75.5) = P(z \geq 0.8) = 0.5 - 0.2881 = 0.2119.$
f. $P(-1 \leq \bar{X} - \mu \leq +1) = P(-1.6 \leq z \leq 1.6) = 2(0.4452) = 0.8904.$

Fall - 2011
ECO-380: Business Statistics
Dr. Khalid Kisswani – Dr. Fida Karam
<u>Exam I (B) – key</u>

Name: ....

1. Given that Z is a standard normal random variable, what is the probability that $-1.5 \le Z \le 2.46$?
a. 0.9263
b. 0.4332
c. 0.4931
d. 0.0599

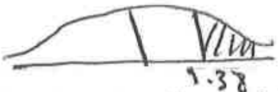


$0.4332 + 0.4931$
$= 0.9263$

2. X is a normally distributed random variable with a mean of 5 and a standard deviation of 4. The probability that X is greater than 10.52 is
a. 0.0029
b. 0.0838
c. 0.4162
d. 0.9971

$X \sim N(\overset{\mu}{5}; \overset{\sigma}{4})$ $P(X \geq 10.52)$

$= P\left(\frac{X-\mu}{\sigma} \geq \frac{10.53-5}{4}\right) = P(Z \geq 1.38)$



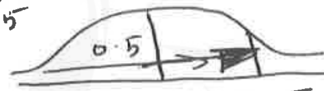
$0.5 - 0.4162$

3. X is a normally distributed random variable with a mean of 20 and a variance of 25. The probability that X is less than 30 is
a. 0.6554
b. 0.4772
c. 0.0228
d. 0.9772

$X \sim N(\overset{\mu}{20}; \overset{\sigma^2}{25})$

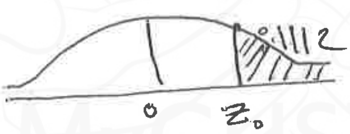
$\sigma \sqrt{25} = 5$

$P(X \le 30)$
$= P\left(\frac{X-\mu}{\sigma} \le \frac{30-20}{5}\right)$



$= P(Z \le 2) = 0.5 + 0.4772 = 0.97$

4. Z is a standard normal random variable. What is the value of Z if the area to the right of Z is 0.1112?
a. 0.3888
b. 1.22
c. -1.22
d. 3.22



$0.5 - 0.1112 = 0.3888$

$Z_\alpha = 1.22$

5. Given that Z is a standard normal random variable, what is the value of Z if the area to the right of Z is 0.8944?
a. 0.3944
b. -1.25
c. +1.25
d. Zero

$\boxed{Z_\alpha = -1.25}$ $0.8944 - 0.5 = 0.3944$

6. Random samples of size 64 are taken from an infinite population whose mean and variance are 100 and 256, respectively. The mean and the standard error of the mean are
a. 100 and 16
b. 100 and 2
c. 100 and 256
d. 100 and 32

$n = 64$ $\mu = 100$ $\sigma^2 = 256$

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{\sqrt{256}}{\sqrt{64}} = \frac{16}{8} = 2$

7. A population has a mean of <u>80</u> and a variance of $\overset{\sigma^2}{\underline{49}}$. A sample of 64 observations will be taken. The probability that the sample mean will be between 81 and 82 is
a. 0.1161
b. 0.8619
c. 0.489
d. 0.3729

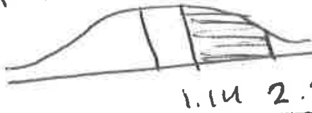
$n = 64$

$P(81 < \bar{X} < 82) = P\left(\frac{\bar{X}-\mu}{\sigma_{\bar{x}}}\right)$

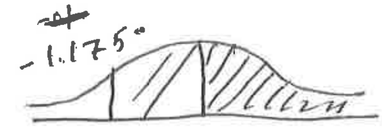
$\sigma = \frac{\sigma}{\sqrt{n}} = \frac{\sqrt{49}}{\sqrt{64}} = 0.875$

$\left(\frac{81-80}{0.875} < Z < \frac{82-80}{0.875}\right)$

$(1.14 < Z < 2.29)$



$=$ Area 2.29 − Area 1.14



$0.5 + 0.4599 = 0.9599$

8. A population has a mean of 50 and a standard deviation of 28. A sample of 49 observations will be taken. The probability that the sample mean will be greater than 43 is
a. 0.5
b. 0.0401
c. 0.4505
d. 0.9599

$\mu = 50$ $\sigma = 28$ $n = 49$ $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{28}{\sqrt{49}} = 4$

$P(\bar{X} > 43) = P\left(\frac{\bar{X} - \mu}{\sigma_{\bar{x}}} \geq \frac{43 - 50}{4}\right)$

$= P(z \geq -1.75)$

9. The standard deviation of a sample of 100 observations equals 25. The variance of the sample equals
a. 2.5
b. 25
c. 625
d. 28,461

~~n = 120~~ $n = 100$ $s = 25$ $s^2 = 25^2 = 625$

10. The mean height of a sample of 25 individuals is 64 inches and the sample variance is 625. The coefficient of variation equals
a. 11.2%
b. 1120%
c. 0.39%
d. 39%

11. The standard deviation of a sample was reported to be 15. The report indicated that $\sum (x - \bar{x})^2 = 4950$. What has been the sample size?
a. 22
b. 21
c. 15
d. 23

$S = 15$

$S^2 = \frac{\sum (x - \bar{x})^2}{n-1}$ 15^2 $225 = \frac{4950}{n-1}$

$225 = \frac{4950}{n-1} =$

Exhibit 1

The following data was collected from a simple random sample of a population.

5	12	6	8
6	7	5	12

5 5 6 6 7 8 12 12

12. Refer to Exhibit 1. The sample mean
a. cannot be determined, since the population size is unknown
b. is 7.625
c. is 81
d. is 8

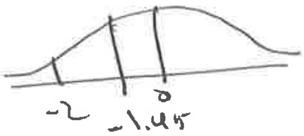
$\bar{X} = \frac{\sum x_i}{n} = 7.625$

13. Refer to Exhibit 1. The median is
a. 6
b. 6.5
c. 7
d. 8

5 5 6 6 7 8 12 12

$\frac{6+7}{2} = 6.5$

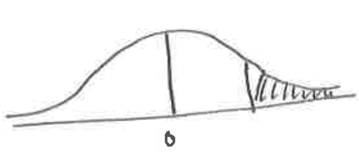
14. Z is a standard normal random variable. The P(-2 ≤ Z ≤ -1.45) equals
a. -0.0507
b. 0.0507
c. 0.0228
d. 0.9037



$0.4772 - 0.4265$

15. Given that Z is a standard normal random variable, what is the probability that Z ≤ -2.12?
a. 0.4830
b. 0.9830
c. 0.017
d. 0.966

$0.5 - 0.483 = 0.017$



St. Dev = σ = 8000

Problem 1 (show all your work) (20 points)

The average starting salary for this year's graduates at a large university (LU) is $20,000 with a ~~variance of $64,000~~. Furthermore, it is known that the starting salaries are normally distributed.

a. What is the probability that a randomly selected LU graduate will have a starting salary of at most $30,400?

$$P(X < 30{,}400) = P\left(Z < \frac{30400 - 20000}{8000}\right) = P(Z < 1.3) = 0.5 + P(0<Z<1.3) = 0.5 + 0.4032 = 0.9032 = 90.32\%$$

b. Individuals with starting salaries of less than $16,600 receive a low income tax break. What percentage of the graduates will receive the tax break?

$$P(X < 16600) = P\left(Z < \frac{16600 - 20000}{8000}\right) = P(Z < -0.43) = P(Z > 0.43) = 0.5 - P(0 < Z < 0.43) = 0.5 - 0.1664 = 0.3336 = 33.36\%$$

c. What are the minimum and the maximum starting salaries of the middle 68% of the LU graduates?

$P(-Z_0 \le Z \le Z_0) = 0.68$, so: $P(0 \le Z \le Z_0) = 0.34$, $Z_0 = 1$, $-Z_0 = -1$

$$Z_0 = \frac{X - 20000}{8000},\ 1 = \frac{X - 20000}{8000},\ X = 30000 \text{ (max)}$$

$$-Z_0 = \frac{X - 20000}{8000},\ -1 = \frac{X - 20000}{8000},\ X = 12000 \text{ (min)}$$

d. If 199 of the recent graduates have salaries of at least $32,240, how many students graduated this year from this university?

$$P(X > 32240) = P\left(Z > \frac{32240 - 20000}{8000}\right) = P(Z > 1.53) = 0.5 - P(0<Z<1.53) = 0.5 - 0.437 = 0.063 = 6.3\%$$

So, 199 graduates are 6.3%, then 199/X = 6.3% , then: total number of graduates = 3,158.7 which is <u>3159</u>

e. if a sample of 36 graduate students is taken, what is the probability that the sample mean of starting salary is at most $22,400?

$$P(\bar{X} < 22400) = P\left(Z < \frac{22400 - 20000}{\frac{8000}{\sqrt{36}}}\right) = P(Z < 1.8)$$

$$= 0.5 + P(0 < Z < 1.8) = 0.5 + 0.4641 = 0.9641$$

Gulf University for Science & Technology
Department of Economics & Finance
ECON 380: Business Statistics
Fall 2011

Dr. Fida Karam
Room N1-115
PHONE: 2530-7351
karam.f@gust.edu.kw

Problem Set VI:
Chapters 11 – 12 - 13

Multiple choice questions:

1. A sample of 20 items provides a sample mean of 15 and a sample variance of 6. Compute a 95% confidence interval estimate for the standard deviation of the population.

a. 3.47 to 12.8
b. 2.88 to 3.88
c. 1.86 to 3.58
d. 5.7 to 6.89

$n = 20$
$\bar{x} = 15$
$s^2 = 6$

$\leq \sigma^2 \leq$

Exhibit 1
On the basis of data provided by a salary survey, the variance in annual salaries for seniors in accounting firms is approximately 20 and the variance in annual salaries for managers in accounting firms is approximately 30. Assuming that the salary data were based on samples of 16 seniors and 16 managers, test the hypothesis that the population variance in the salaries for managers is greater than the population variance in salaries for seniors.

	Managers	Seniors
Sample Size	16	16
Sample Mean	520	540
Sample Variance	30	20

2. Refer to Exhibit 1. The null hypothesis is

a. $S_1^2 > S_2^2$
b. $S_1^2 \leq S_2^2$
c. $\sigma_1^2 > \sigma_2^2$
d. $\sigma_1^2 \leq \sigma_2^2$ ✓

3. Refer to Exhibit 1. The test statistic is ✓

a. 1.5
b. 0.96

c. 1
d. 4

4. Refer to Exhibit 1. The *p*-value for this test is
a. greater than 0.1
b. less than 0.1
c. between 0.025 and 0.05
d. None of these alternatives is correct.

5. Refer to Exhibit 1. At 99% confidence the null hypothesis
a. should be rejected
b. should not be rejected
c. should be revised
d. None of these alternatives is correct.

Exhibit 2
The filling variance for boxes of breakfast cereal is designed to be 0.25. A sample of 25 boxes of cereal shows a sample variance of 0.4 grams. We need to determine whether the variance in the cereal box fillings is not meeting the design specification.

6. Refer to Exhibit 2. The null hypothesis is
a. $S^2 = 0.25$
b. $S^2 \leq 0.25$
c. $\sigma^2 = 0.25$
d. $\sigma^2 \leq 0.25$

7. Refer to Exhibit 2. The test statistic is
a. 15.36
b. 38.4
c. 40
d. 24

8. Refer to Exhibit 2. The *p*-value for this test is
a. 0.05
b. between 0.025 and 0.05
c. between 0.05 and 0.1
d. 1.96

9. Refer to Exhibit 2. At 95% confidence, the null hypothesis
a. should be rejected
b. should not be rejected
c. should be revised

d. None of these alternatives is correct.

Exhibit 3:
During the first 13 weeks of the autumn schedules, the Saturday evening 8:00 p.m. to 9:00 p.m. audience proportions were recorded as: BBC1 & 2: 43%; Sky channels: 34%; and others, 23%. A sample of 400 homes two weeks after a Saturday night schedule revision yielded the following viewing audience data: BBC1 & 2: 164 homes; Sky channels; 172 homes; and others, 64 homes. Test with $\alpha = 0.01$ to determine whether the viewing audience proportions changed.

10. Refer to Exhibit 3. The expected frequency of BBC1 & 2 is
a. 172
b. 43%
c. 164
d. 64

11. Refer to Exhibit 3. The calculated value for the test statistic equals
a. 0.5444
b. 300
c. 18.42
d. 6.6615

12. Refer to Exhibit 3. The *p*-value is
a. less than .005
b. 0.01
c. between .05 and 0.1
d. greater than 0.1

13. Refer to Exhibit 3. At 95% confidence, the null hypothesis
a. should not be rejected
b. should be rejected
c. was designed wrong
d. None of these alternatives is correct.

14. The chi-square value for a one-tailed (lower tail) hypothesis test at 95% confidence and a sample size of 25 is
a. 13.848
b. 36.415
c. 39.364
d. 12.401

15. The chi-square value for a one-tailed test (upper tail) when the level of significance is 0.1 and the sample size is 15 is

a. 21.064
b. 23.685
c. 7.790
d. 6.571

16. The ANOVA procedure is a statistical approach for determining whether or not
a. the means of two samples are equal
b. the means of two or more samples are equal
c. the means of more than two samples are equal
d. the means of two or more populations are equal

17. The critical value of F for a one-tailed (upper tail) hypothesis test at 90% confidence when there is a sample size of 16 for the sample with the smaller variance, and there is a sample size of 8 for the sample with the larger sample variance is
a. 2.16
b. 2.71
c. 2.63
d. 3.51

18. In a completely randomized design involving four treatments, the following information is provided.

	Treatment 1	Treatment 2	Treatment 3	Treatment 4
Sample Size	45	20	13	19
Sample Mean	30	35	40	50

The overall mean for all treatments is

$\bar{\bar{X}} = \frac{(45,30)+(35,20)+(13,40)+(19.50)}{45+20+13+19} = 36.288$

a. 36.29
b. 38.75
c. 40
d. 24.25

Exhibit 4

k n

In a completely randomized experimental design involving four treatments, 5 observations were recorded for each of the four treatments. The following information is provided.

SSE = 600
SST = 800

19. Refer to Exhibit 4. The sum of squares between treatments (SSTR) is
a. 20
b. 800

~~SSTR =~~

~~1200~~

$\overset{800}{SST} = SSTR + \overset{600}{SSE}$

$SSTR = 800 - 600 = 200$

c. 600
d. 200

20. Refer to Exhibit 4. The number of degrees of freedom corresponding to between treatments is
a. 16
b. 3
c. 5
d. 4

K-1 = 4-1=3

nT-K

21. Refer to Exhibit 4. The number of degrees of freedom corresponding to within treatments is
a. 16
b. 59
c. 4
d. 3

nT-K = 20-4=16

22. Refer to Exhibit 4. The mean square between treatments (MSTR) is
a. 3.34
b. 16.67
c. 66.67
d. 12.00

$MSTR = \frac{SSTR}{K-1} = \frac{200}{4-1} =$

23. Refer to Exhibit 4. The mean square within treatments (MSE) is
a. 50
b. 37.5
c. 200
d. 16.67

$MSE = \frac{SSE}{nT-K} = \frac{600}{16} = 37.5$

24. Refer to Exhibit 4. The test statistic is
a. 1.78
b. 5.0
c. 0.56
d. 15

$F = \frac{MSTR}{MSE} = \frac{66.67}{37.5} = 1.77$

25. Refer to Exhibit 4. If at 95% confidence we want to determine whether or not the means of the four populations are equal, the *p*-value is
a. between 0.05 to 0.10
b. greater than 0.1
c. between 0.01 to 0.025
d. less than 0.01

$\alpha = 0.05$

$F_{0.05}(3, 16) =$

$1.77 < 3.24$

P-value

P-vaule > 0.1

K-1 = 4-1=3
nT-K = 16

Problem 1:

Guitars R. US has three stores located in three different areas. Random samples of the sales of the three stores (in $1000) are shown below.

$\bar{\bar{X}} = (5.80)+(4.83)+(3.84) = \frac{735.84}{12}$

$400 + 332 + 252$

$\frac{984}{12} = 82$

	Store 1		Store 2		Store 3
1	80	1	85	1	79
2	75	2	86	2	85
3	76	3	81	3	88
4	89	4	80		
5	80				
	$\bar{X} = 80$		$\bar{X} = 83$		$\bar{X} = 84$

a. Compute the overall mean $\bar{\bar{X}}$. 82
b. State the null and alternative hypotheses to be tested.
c. Show the complete ANOVA table for this test including the test statistic.
d. The null hypothesis is to be tested at 95% confidence. Determine the critical value for this test. What do you conclude?
e. Determine the *p*-value and use it for the test.

	df	SS	MS	F
R-	2	SSTR	MSTR	MSTR
Error	9	SSE	MSSE	
	11	SST		

ANS:

a. *82*

b. *H_o: $\mu_1 = \mu_2 = \mu_3$*

H_a: At least one mean is different from the others.

c.

Source of Variation	*SS*	*df*	*MS*	*F*
Between Groups	*36*	*2*	*18*	*0.8526*
Within Groups	*190*	*9*	*21.11*	
Total	*226*	*11*		

d. *Critical F = 4.26, do not reject Ho and conclude there is no evidence of significant difference.*

e. *p-value > 0.1, therefore do not reject H_o*

Problem 2:

In a completely randomized experimental design, 18 experimental units were used for the first treatment, 10 experimental units for the second treatment, and 15 experimental units for the third treatment. Part of the ANOVA table for this experiment is shown below.

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F
Between Treatments	36 ?	2 ?	18 ?	3.0
Error (Within Treatments)	240 ?	40 ?	6	
Total	276 ?	42 ?	$\frac{x}{6} = 3$	

$x = 3 \times 6 = 18$

a. Fill in **all** the blanks in the above ANOVA table.
b. At 95% confidence, test to see if there is a significant difference among the means.

ANS:

a.

Source of Variation	*Sum of Squares*	*Degrees of Freedom*	*Mean Square*	*F*
Between Treatments	*36*	*2*	*18*	*3.0*
Error (Within Treatments)	*240*	*40*	*6*	
Total	*276*	*42*		

b. *For F = 3, the p-value is between 0.05 and 0.1; do not reject H_0 and conclude there is not a significant difference among the means. (Also, test statistic F = 3 < 3.23.)*

3 < p-value <

$$MSTR = \frac{SSTR}{2} =$$

$$18 = \frac{SSTR}{2} = 36$$

$$MSE = \frac{SSE}{n_T - K}$$

$$6 = \frac{x}{40} \qquad x = 6 \times 40 = 240$$

Gulf University for Science & Technology
Department of Economics & Finance
ECON 380: Business Statistics
Fall 2011

Dr. Fida Karam
Room N1-115
PHONE: 2530-7351
karam.f@gust.edu.kw

Problem Set VI:
Chapters 11 – 12 - 13

Multiple choice questions:

$\bar{x}$ s^2
n

1. A sample of 20 items provides a sample mean of 15 and a sample variance of 6. Compute a 95% confidence interval estimate for the standard deviation of the population.

a. 3.47 to 12.8
b. 2.88 to 3.88
c. 1.86 to 3.58
d. 5.7 to 6.89

$n = 20 \quad \bar{x} = 15 \quad s^2 = 6$

$$\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{(1-\frac{\alpha}{2})}}$$

Exhibit 1

On the basis of data provided by a salary survey, the variance in annual salaries for seniors in accounting firms is approximately 20 and the variance in annual salaries for managers in accounting firms is approximately 30. Assuming that the salary data were based on samples of 16 seniors and 16 managers, test the hypothesis that the population variance in the salaries for managers is greater than the population variance in salaries for seniors.

	Managers		**Seniors**
Sample Size	16	>	16
Sample Mean	520		540
Sample Variance	30		20

Research
H1
σ_1^2 20

2. Refer to Exhibit 1. The null hypothesis is

a. $S_1^2 > S_2^2$
b. $S_1^2 \leq S_2^2$
c. $\sigma_1^2 > \sigma_2^2$
d. $\sigma_1^2 \leq \sigma_2^2$

3. Refer to Exhibit 1. The test statistic is

a. 1.5
b. 0.96

$$F = \frac{S_1^2}{S_2^2} = \frac{30}{20} = 1.5$$

c. 1
d. 4

4. Refer to Exhibit 1. The p-value for this test is
a. greater than 0.1
b. less than 0.1
c. between 0.025 and 0.05
d. None of these alternatives is correct.

$df_N = n_1 - 1 = 16 - 1 = 15$
$df_D = n_2 - 1 = 16 - 1 = 15$
$1.5 < 1.97$
$P.V > 0.1$

5. Refer to Exhibit 1. At 99% confidence the null hypothesis
a. ~~should be rejected~~
b. should not be rejected
c. should be revised
d. None of these alternatives is correct.

$99\% = 1 - \alpha \rightarrow \alpha = 0.01$
$P.V > \alpha$
so don not reject

Exhibit 2
The filling variance for boxes of breakfast cereal is designed to be 0.25. A sample of 25 boxes of cereal shows a sample variance of 0.4 grams. We need to determine whether the variance in the cereal box fillings is not meeting the design specification.

6. Refer to Exhibit 2. The null hypothesis is
a. $S^2 = 0.25$
b. $S^2 \leq 0.25$
c. $\sigma^2 = 0.25$
d. $\sigma^2 \leq 0.25$

7. Refer to Exhibit 2. The test statistic is
a. 15.36
b. 38.4
c. 40
d. 24

$$\chi^2 = \frac{(n-1)\, s^2}{\sigma_0^2} = \frac{(25-1)\,0.4}{0.25} = 38.4$$

8. Refer to Exhibit 2. The p-value for this test is
a. 0.05
b. between 0.025 and 0.05
c. between 0.05 and 0.1
d. 1.96

9. Refer to Exhibit 2. At 95% confidence, the null hypothesis
a. should be rejected
b. should not be rejected
c. should be revised

$95\% = 1 - \alpha$
$\alpha = 0.05$
$\chi^2_{0.05,19}$
$P.V > \alpha \rightarrow$ don't reject

	π	f_i	e_i
A	43%	164	43% (400) = 172
B	34%	172	~~136~~ 136
C	23%	64	92

d. None of these alternatives is correct.

Exhibit 3:
During the first 13 weeks of the autumn schedules, the Saturday evening 8:00 p.m. to 9:00 p.m. audience proportions were recorded as: BBC1 & 2: 43%; Sky channels: 34%; and others, 23%. A sample of 400 homes two weeks after a Saturday night schedule revision yielded the following viewing audience data: BBC1 & 2: 164 homes; Sky channels; 172 homes; and others, 64 homes. Test with $\alpha = 0.01$ to determine whether the viewing audience proportions changed.

10. Refer to Exhibit 3. The expected frequency of BBC1 & 2 is
a. 172
b. 43%
c. 164
d. 64

π BBC Sky other — e_i BBC1 — 43% + 400 BBC1

11. Refer to Exhibit 3. The calculated value for the test statistic equals
a. 0.5444
b. 300
c. 18.42
d. 6.6615

43%

12. Refer to Exhibit 3. The *p*-value is
a. less than .005
b. 0.01
c. between .05 and 0.1
d. greater than 0.1

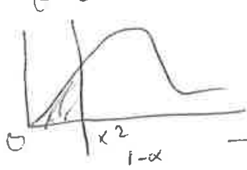
df = K − 1 = 3 − 1 = 2
18.42 > 10.597
PV < 0.005

13. Refer to Exhibit 3. At 95% confidence, the null hypothesis
a. should not be rejected
b. should be rejected
c. was designed wrong
d. None of these alternatives is correct.

$\alpha = 0.05$
PV < 0.05 → rejec

14. The chi-square value for a one-tailed (lower tail) hypothesis test at 95% confidence and a sample size of 25 is
a. 13.848
b. 36.415
c. 39.364
d. 12.401

n = 25



→ $x^2_{0.95,\,24} = 13.848$

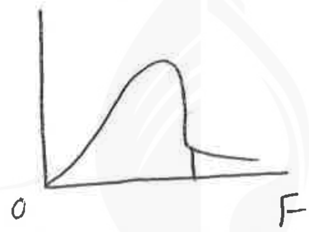
15. The chi-square value for a one-tailed test (upper tail) when the level of significance is 0.1 and the sample size is 15 is

→ $X^2_{0.1,\,14} = 21.064$

a. 21.064
b. 23.685
c. 7.790
d. 6.571

16. The ANOVA procedure is a statistical approach for determining whether or not
a. the means of two samples are equal
b. the means of two or more samples are equal
c. the means of more than two samples are equal
d. the means of two or more populations are equal

17. The critical value of F for a one-tailed (upper tail) hypothesis test at 90% confidence when there is a sample size of 16 for the sample with the smaller variance, and there is a sample size of 8 for the sample with the larger sample variance is
a. 2.16
b. 2.71
c. 2.63
d. 3.51



$\alpha = 1 - 90\% = 0.1$

$n = 16$

$n_1 = 8$

$n_2 = 16$

$dfD = 15$

$F_{0.1(7,15)} = 2.16$ $\quad dfN = 7$

18. In a completely randomized design involving four treatments, the following information is provided.

	Treatment 1	Treatment 2	Treatment 3	Treatment 4
Sample Size n	45	20	13	19
Sample Mean $\bar{x}$	30	35	40	50

$\div \; n_T$

The overall mean for all treatments is
a. 36.29
b. 38.75
c. 40
d. 24.25

$$\bar{\bar{x}} = \frac{45(30) + 20(35) + 13(40) + 19(50)}{30 + 35 + 40 + 50} = 36.27$$

Exhibit 4
In a completely randomized experimental design involving four treatments, 5 observations were recorded for each of the four treatments. The following information is provided.

$n_1 \; n_2 \; n_3 \; n_4$
$5 \quad 5 \quad 5 \quad 5 \qquad k = 4$

SSE = 600
SST = 800

S S TR $\quad$ SSTR =

19. Refer to Exhibit 4. The sum of squares between treatments (SSTR) is
a. 20
b. 800

$$SSTR = SST - SSE = 800 - 600 = 200$$

c. 600
d. 200

20. Refer to Exhibit 4. The number of degrees of freedom corresponding to between treatments is
a. 16
b. 3
c. 5
d. 4

$4 - 1 = 3$

21. Refer to Exhibit 4. The number of degrees of freedom corresponding to within treatments is
a. 16
b. 59
c. 4
d. 3

$4 + 4 + 4 + 4 = 16$

M S TR

22. Refer to Exhibit 4. The mean square between treatments (MSTR) is
a. 3.34
b. 16.67
c. 66.67
d. 12.00

$MSTR = \frac{200}{3} = 66.67$

M M S

23. Refer to Exhibit 4. The mean square within treatments (MSE) is
a. 50
b. 37.5
c. 200
d. 16.67

$MSE = \frac{600}{16} = 37.5$

24. Refer to Exhibit 4. The test statistic is
a. 1.78
b. 5.0
c. 0.56
d. 15

$F = \frac{66.67}{37.5} = 1.78$

25. Refer to Exhibit 4. If at 95% confidence we want to determine whether or not the means of the four populations are equal, the *p*-value is
a. between 0.05 to 0.10
b. greater than 0.1
c. between 0.01 to 0.025
d. less than 0.01

$\alpha = 0.05$

$dfN = 3$

$dfD = 16$

$F = 1.78$

$1.78 < 2.96$ (5)

Pvalue > 0.1

Problem 1:

Guitars R. US has three stores located in three different areas. Random samples of the sales of the three stores (in $1000) are shown below.

$\bar{\bar{x}} = 82$

Store 1	Store 2	Store 3
80	85	79
75	86	85
76	81	88
89	80	
80		

$n_1 = 5$ $n_2 = 4$ $n_3 = 3$

$\bar{x}_1 = 80$ $\bar{x}_2 = 83$ $\bar{x}_3 = 84$

a. Compute the overall mean $\bar{\bar{X}}$.

b. State the null and alternative hypotheses to be tested.

c. Show the complete ANOVA table for this test including the test statistic.

d. The null hypothesis is to be tested at 95% confidence. Determine the critical value for this test. What do you conclude?

e. Determine the p-value and use it for the test.

ANS:

a. 82

b. H_o: $\mu_1 = \mu_2 = \mu_3$

H_a: *At least one mean is different from the others.*

c.

Source of Variation	*SS*	*df*	*MS*	*F*
Between Groups	*36*	*2*	*18*	*0.8526*
Within Groups	*190*	*9*	*21.11*	
Total	*226*	*11*		

d. *Critical F = 4.26, do not reject Ho and conclude there is no evidence of significant difference.*

e. *p-value > 0.1, therefore do not reject* H_o

Problem 2:

In a completely randomized experimental design, 18 experimental units were used for the first treatment, 10 experimental units for the second treatment, and 15 experimental units for the third treatment. Part of the ANOVA table for this experiment is shown below.

$K = 3$ DFN $= 3 - 1 = 2$

n_1 n_2 n_3

18 10 15 DFD $= 40$

Source of Variation	**Sum of Squares**	**Degrees of Freedom**	**Mean Square**	**F**
Between Treatments	36 ?	2 ?	18 ?	3.0
Error (Within Treatments)	240 ?	40 ?	6	
Total	276 ?	42 ?		

a. Fill in **all** the blanks in the above ANOVA table.
b. At 95% confidence, test to see if there is a significant difference among the means.

ANS:

a.

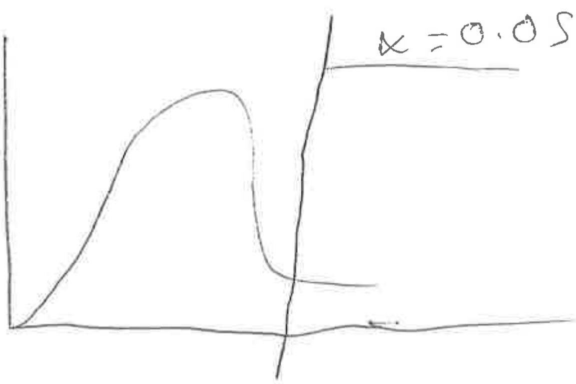
Source of Variation	*Sum of Squares*	*Degrees of Freedom*	*Mean Square*	*F*
Between Treatments	*36*	*2*	*18*	*3.0*
Error (Within Treatments)	*240*	*40*	*6*	
Total	*276*	*42*		

b. For F = 3, the p-value is between 0.05 and 0.1; do not reject H_0 and conclude there is not a significant difference among the means. (Also, test statistic F = 3 < 3.23.)

$F = 3$

$df_N = 2$

$df_D = 40$



$F_{0.05}(2, 40) = 3.23$

$F \leq 3.23$

don't reject

Gulf University for Science & Technology
Department of Economics & Finance
ECON 380: Business Statistics
Fall 2011

Dr. Fida Karam
Room N1-115
PHONE: 2530-7351
karam.f@gust.edu.kw

Problem Set III:
Chapters 9 and 10

Multiple choice questions:

1. Your partner claims that the average yearly rate of return of your joint project is less than 20.0%. You plan on taking a sample to test his claim. The correct set of hypotheses is

a. H_0: $\mu < 20.0\%$ H_a: $\mu \geq 20.0\%$
b. H_0: $\mu \leq 20.0\%$ H_a: $\mu > 20.0\%$
c. H_0: $\mu > 20.0\%$ H_a: $\mu \leq 20.0\%$
d. H_0: $\mu \geq 20.0\%$ H_a: $\mu < 20.0\%$

ANS: B

2. The manager of a bookstore is considering a new bonus plan in order to increase sales. Currently, the mean sales rate per salesperson is 30 books per month. The correct set of hypotheses for testing the effect of the bonus plan is

a. H_0: $\mu < 30$ H_a: $\mu \leq 30$
b. H_0: $\mu \leq 30$ H_a: $\mu > 30$
c. H_0: $\mu > 30$ H_a: $\mu \leq 30$
d. H_0: $\mu \geq 30$ H_a: $\mu < 30$

ANS: B

3. For a two-tailed test at 87.64% confidence; Z =

a. 1.54
b. 1.96
c. 1.645
d. 1.16

ANS: A

4. For a one-tailed test (lower test), a sample of 20 at 99% confidence, t =
a. 2.576
b. 2.5395
c. -2.8609
d. -2.5395

ANS: D

5. For a one-tailed test (upper tail), a sample size of 18 at 90% confidence, t =
a. 1.7396
b. 1.645
c. -1.3334
d. 1.3334

ANS: D

6. For a two-tailed test, a sample size of 11 at 90% confidence, t =
a. 1.645
b. 1.3722
c. 1.8125
d. -1.3722

ANS: C

7. For a one-tailed test (lower tail) at 94.63% confidence, Z =
a. -1.61
b. -1.93
c. 1.93
d. 1.61

ANS: A

8. For a one-tailed test (upper tail) 85.31% confidence, Z =
a. 1.96
b. -1.05
c. 1.05
d. 1.45

ANS: C

Exhibit 1

$n = 64$ $\bar{x} = 50$ $s = 16$ $H_0: \mu = 54$
$H_a: \mu \neq 54$

9. Refer to Exhibit 1. The test statistic equals
a. -4
b. -3
c. -2
d. -1

ANS: C

10. Refer to Exhibit 1. The *p*-value is between
a. .005 to .01
b. .01 to .025
c. .02 to .05
d. .01 to .05

ANS: C

11. Refer to Exhibit 1. If the test is done at 95% confidence, the null hypothesis should
a. not be rejected
b. be rejected
c. Not enough information is given to answer this question.
d. None of these alternatives is correct.

ANS: B

12. Refer to Exhibit 1. The critical value for t, when $\alpha = 0.05$, is
a. 1.96
b. 1.645
c. 1.669
d. 1.998

ANS: D

13. Refer to Exhibit 1. The null hypothesis will be rejected if the test statistic t is
a. $\geq t_\alpha$
b. $\leq t_\alpha$
c. $\leq -t_\alpha$ or $\geq t_\alpha$
d. $\leq -t_{\alpha/2}$ or $\geq t_{\alpha/2}$

ANS: D

Exhibit 2
The average price of toothbrushes charged by a toothbrush producer has been \$2.5. Recently, the company has undertaken several efficiency measures in order to reduce prices. Management is interested in determining whether their efficiency measures have actually reduced prices. A random sample of 81 shops is selected and the average price is determined to be \$2.4. Furthermore, assume that the standard deviation of the population is \$0.18.

14. Refer to Exhibit 2. The standard error has a value of
a. 0.18
b. 7
c. 2.5
d. 0.02

ANS: D

15. Refer to Exhibit 2. The value of the test statistic for this hypothesis test is
a. 1.96
b. 1.645
c. -5
d. 5

ANS: C

16. Refer to Exhibit 2. The *p*-value for this problem is
a. 0.4938
b. 0.0000
c. 0.0124
d. 0.5

ANS: B

17. Refer to Exhibit 2. If $\alpha = 0.05$, the null hypothesis should
a. not be rejected
b. be rejected
c. Not enough information is given to answer this question.
d. None of these alternatives is correct.

ANS: B

Exhibit 3
A sample of 40 petrol stations in city A yielded a mean price for unleaded petrol of €1.54 per litre. A sample of 35 petrol stations in city Y yielded a mean price of €1.22 per litre. Assume that prior studies indicate a population standard deviation of €1 in city X and €0.8 in city Y.

	μ_1 X	μ_2 Y
n	40	35
$\bar{X}$	1.54	1.22
$z \leftarrow \sigma$	1	0.8

18. Refer to Exhibit 3. The point estimate for the difference between the means of the two populations is
a. 0.1
b. 0.08
c. -5
d. 0.32

$\bar{X}_1 - \bar{X}_2 = 1.54 - 1.22 = 0.32$

ANS: D

19. Refer to Exhibit 3. The standard error of the difference between the two population means is
a. 0.32
b. 0.1
c. 0.08
d. 0.2

$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{1^2}{40} + \frac{0.8^2}{35}} = 0.2$ ✓

ANS: D

20. Refer to Exhibit 3. The 95% confidence interval for the difference between the two population means is
a. -0.072 to 0.712
b. -3.92 to 3.92
c. -13.84 to 1.84
d. -24.228 to 12.23

$\mu_1 - \mu_2 = 0.32 \pm 1.96(0.2) = 0.32 \pm .392$

-0.072 to 0.712

ANS: A

21. Refer to Exhibit 3. The test statistic for the equality between the two population means is
a. -.47
b. -.65
c. 16
d. 1.6

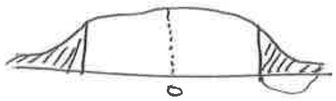
$H_0: \mu_1 - \mu_2 = 0$

$H_1: \mu_1 - \mu_2 \neq$

$Z = \frac{0.32 - 0}{0.2} = 1.6$

ANS: D

22. Refer to Exhibit 3. The *p*-value for the difference between the two population means is
a. .0548
b. .1096
c. .4987
d. .9987



2 times the area

P-value = 2(0.5 - 0.44

= 0.1096

ANS: B

Exhibit 4
The following information was obtained from independent random samples.
Assume normally distributed populations with equal variances.

	Sample 1	Sample 2
Sample Mean $\bar{X}$	45	42
Sample Variance S^2	85	90 → t table
Sample Size n	10	12

23. Refer to Exhibit 4. The point estimate for the difference between the two population means is
a. 0
b. 2
c. 3
d. 15

$\bar{X}_1 - \bar{X}_2 = 45 - 42 = 3$

ANS: C

24. Refer to Exhibit 4. The standard error of $\bar{x}_1 - \bar{x}_2$ is
a. 3.0
b. 4.0
c. 8.372
d. 19.48

$S_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} = \sqrt{\frac{85}{10} + \frac{90}{12}} = 4$

ANS: B

25. Refer to Exhibit 4. The degrees of freedom for the t-distribution are
a. 22
b. 21
c. 20
d. 19

$$df = \frac{\left(\frac{85}{10} + \frac{90}{12}\right)^2}{\frac{1}{10-1}\left(\frac{85}{10}\right)^2 + \frac{1}{12-1}\left(\frac{90}{12}\right)^2} = 19.48$$

round always.

ANS: D

26. Refer to Exhibit 4. The 95% confidence interval for the difference between the two population means is
a. -5.372 to 11.372
b. -5 to 3
c. -4.86 to 10.86
d. -2.65 to 8.65

$\mu_1 - \mu_2 = 3 \pm 2.093(4) = 3 \pm 8.372$ or -5.372 to 11.372

ANS: A

27. Refer to Exhibit 4. To perform the following test: H0: μ1 - μ2 ≤ 0 H1: μ1 - μ2 > 0., the test statistic is
a. 0.75
b. -5
c. 4.86
d. -2.65

$H_0: \mu_1 - \mu_2 \leq 0$

$H_1: \mu_1 - \mu_2 > 0$

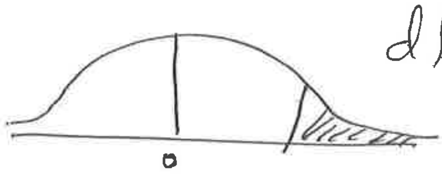
ANS: A

$t = \frac{3 - 0}{4} = 0.75$

28. Refer to Exhibit 4. The *p*-value for the difference between the two population means

a. 0.05 to 0.1
b. -0.25 to -0.1
c. 0.1 to 0.25
d. 0.05 to 0.25

ANS: C



$df = 19$

t

$0.6876 < 0.75 < 1.3277$

$0.25 > P.v > 0.1$

0.25%

0.25

Gulf University for Science & Technology
Department of Economics & Finance

Dr. Fida Karam
Room N1-115
PHONE: 2530-7351
karam.f@gust.edu.kw

ECON 380: Business Statistics

Fall 2011

Problem Set III:
Chapters 9 and 10

Multiple choice questions:

1. Your partner claims that the average yearly rate of return of your joint project is less than 20.0%. You plan on taking a sample to test his claim. The correct set of hypotheses is

Ho:

a. H_0: μ < 20.0% H_a: $\mu \geq$ 20.0%
b. H_0: $\mu \leq$ 20.0% H_a: μ > 20.0%
c. H_0: μ > 20.0% H_a: $\mu \leq$ 20.0%
d. H_0: $\mu \geq$ 20.0% H_a: μ < 20.0%

ANS: B

2. The manager of a bookstore is considering a new bonus plan in order to increase sales. Currently, the mean sales rate per salesperson is 30 books per month. The correct set of hypotheses for testing the effect of the bonus plan is

research.

a. H_0: μ < 30 H_a: $\mu \leq$ 30
b. H_0: $\mu \leq$ 30 H_a: $\mu \geq$ 30
c. H_0: μ > 30 H_a: $\mu \leq$ 30
d. H_0: $\mu \geq$ 30 H_a: μ < 30

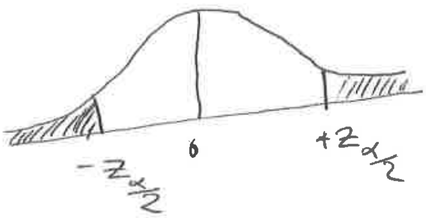
ANS: B

3. For a two-tailed test at 87.64% confidence; Z = $\frac{87.64}{100} = \frac{0.8764}{2} = \boxed{}$

go to table.

a. 1.54
b. 1.96
c. 1.645
d. 1.16

ANS: A



$P(0 \leq Z \leq Z_{\alpha/2}) = \frac{0.87}{2}$

$= 0.4388$

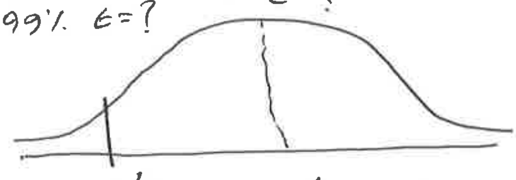
$0.99 - 1 = 0.01$

$t_{0.01, 19}$

4. For a one-tailed test (lower test), a sample of 20 at 99% confidence, t =
a. 2.576
b. 2.5395
c. -2.8609
d. -2.5395

$n = 20$

99% $t = ?$



$1 - \alpha = 99\%$

$\alpha = 0.01$

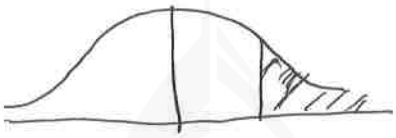
$df = 20 - 19$

$t_{0.01, 19} = 2.539$

ANS: D chose negative because lower tale test

5. For a one-tailed test (upper tail), a sample size of 18 at 90% confidence, t =
a. 1.7396
b. 1.645
c. -1.3334
d. 1.3334

ANS: D



$1 - \alpha = 90$

$\alpha = 0.1$

$df = 18 - 1 = 17$

$t_{0.1, 17} = 1.3334$

6. For a two-tailed test, a sample size of 11 at 90% confidence, t =
a. 1.645
b. 1.3722
c. 1.8125
d. -1.3722

ANS: C

$n = 11$

$1 - \alpha = 90\%$

$\alpha = 0.1$

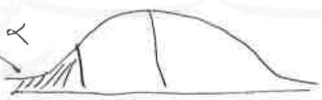
$\alpha/2 = 0.05$

$df = 11 - 1 = 10$

$t_{0.05, 10} = \pm 1.812$

7. For a one-tailed test (lower tail) at 94.63% confidence, Z =
a. -1.61
b. -1.93
c. 1.93
d. 1.61

ANS: A



~~$P(0 \le Z \le Z_0) =$~~

$\alpha = 1 - 0.963 = 0.0537$

$P(-Z \le Z \le 0) =$

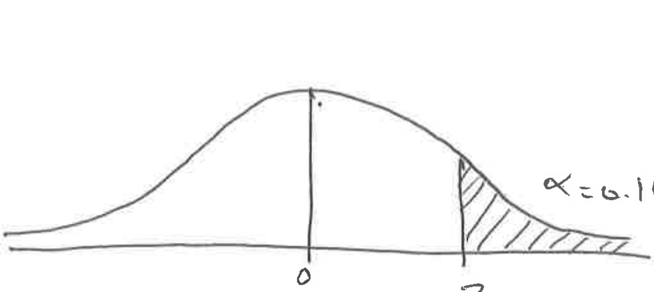
$= 0.5 - 0.0537$

$= 0.4463$

$-Z_\alpha = -1.61$

8. For a one-tailed test (upper tail) 85.31% confidence, Z =
a. 1.96
b. -1.05
c. 1.05
d. 1.45

ANS: C



$1 - \alpha = 85.31\%$

$\alpha = 1 - 0.8531 = 0.1469$

$P(0 \le Z \le Z_\alpha) = 0.5 - 0.1469$

Exhibit 1

n = 64 $\bar{x}$ = 50 s = 16 H_0: μ = 54

H_a: μ ≠ 54

9. Refer to Exhibit 1. The test statistic equals

$t = \frac{50 - 54}{16/\sqrt{64}} = -2$

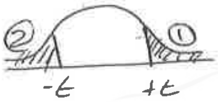
a. -4
b. -3
c. -2
d. -1

ANS: C

10. Refer to Exhibit 1. The *p*-value is between

double because to areass

a. .005 to .01
b. .01 to .025
c. .02 to .05
d. .01 to .05



df = 64−1 = 63

1.96 < 2 < 2.3264

0.025 > Area > 0.01

0.05 > pvalue > 0.02

ANS: C

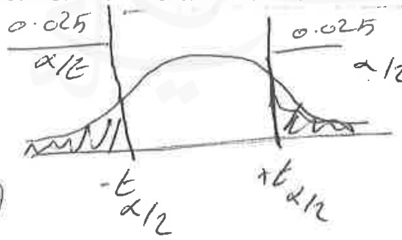
11. Refer to Exhibit 1. If the test is done at 95% confidence, the null hypothesis should
a. not be rejected
b. be rejected
c. Not enough information is given to answer this question.
d. None of these alternatives is correct.

.95 = 1−α = α = 0.05

P-value < α → reject H_0

ANS: B

12. Refer to Exhibit 1. The critical value for t, when α = 0.05, is
a. 1.96
b. 1.645
c. 1.669
d. 1.998



df = 63

$t_{0.025, 63} = 1.96$

ANS: ~~D~~ A

13. Refer to Exhibit 1. The null hypothesis will be rejected if the test statistic t is
a. ≥ t_α
b. ≤ t_α
c. ≤$-t_\alpha$ or ≥ t_α
d. ≤$-t_{\alpha/2}$ or ≥ $t_{\alpha/2}$

ANS: D

−2 < −1.96

$< -t_{\alpha/2}$

$> +t_{\alpha/2}$

Exhibit 2

The average price of toothbrushes charged by a toothbrush producer has been \$2.5. Recently, the company has undertaken several efficiency measures in order to reduce prices. Management is interested in determining whether their efficiency measures have actually reduced prices. A random sample of 81 shops is selected and the average price is determined to be \$2.4. $\bar{X}$ Furthermore, assume that the standard deviation of the population is \$0.18. S

14. Refer to Exhibit 2. The standard error has a value of

a. 0.18 $n=81$ $\bar{X}=2.4$ $\sigma=0.18 \rightarrow Z$ $H_0: \mu \geq 2.5$

b. 7 $H_1: \mu < 2.5$

c. 2.5 $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

d. 0.02 (reserach)

ANS: D $= \frac{0.18}{\sqrt{81}} = 0.02$

15. Refer to Exhibit 2. The value of the test statistic for this hypothesis test is

a. 1.96

b. 1.645 $Z = \frac{2.4-2.5}{0.02} = -5$

c. -5

d. 5

ANS: C

16. Refer to Exhibit 2. The *p*-value for this problem is

a. 0.4938

b. 0.0000

c. 0.0124

d. 0.5

$P\text{-value} = 0.5 - P(-5 \leq Z \leq 0)$

$= 0.5 - 0.5 = 0$

ANS: B

17. Refer to Exhibit 2. If α = 0.05, the null hypothesis should

a. not be rejected

b. be rejected lower

c. Not enough information is given to answer this question.

d. None of these alternatives is correct.

ANS: B

Lsf ①

Exhibit 3

A sample of 40 petrol stations in city A yielded a mean price for unleaded petrol of €1.54 per litre. A sample of 35 petrol stations in city Y yielded a mean price of €1.22 per litre. Assume that prior studies indicate a population standard deviation of €1 in city X and €0.8 in city Y.

18. Refer to Exhibit 3. The point estimate for the difference between the means of the two populations is
a. 0.1
b. 0.08
c. -5
d. 0.32

ANS: D

19. Refer to Exhibit 3. The standard error of the difference between the two population means is
a. 0.32
b. 0.1
c. 0.08
d. 0.2

ANS: D

20. Refer to Exhibit 3. The 95% confidence interval for the difference between the two population means is
a. -0.072 to 0.712
b. -3.92 to 3.92
c. -13.84 to 1.84
d. -24.228 to 12.23

ANS: A

21. Refer to Exhibit 3. The test statistic for the equality between the two population means is
a. -.47
b. -.65
c. 16
d. 1.6

ANS: D

22. Refer to Exhibit 3. The *p*-value for the difference between the two population means is
a. .0548
b. .1096
c. .4987
d. .9987

ANS: B

Exhibit 4
The following information was obtained from independent random samples.
Assume normally distributed populations with equal variances.

	Sample 1	**Sample 2**
Sample Mean	45	42
Sample Variance	85	90
Sample Size	10	12

23. Refer to Exhibit 4. The point estimate for the difference between the two population means is
a. 0
b. 2
c. 3
d. 15

ANS: C

24. Refer to Exhibit 4. The standard error of $\bar{x}_1 - \bar{x}_2$ is
a. 3.0
b. 4.0
c. 8.372
d. 19.48

ANS: B

25. Refer to Exhibit 4. The degrees of freedom for the t-distribution are
a. 22
b. 21
c. 20
d. 19

ANS: D

26. Refer to Exhibit 4. The 95% confidence interval for the difference between the two population means is
a. -5.372 to 11.372
b. -5 to 3
c. -4.86 to 10.86
d. -2.65 to 8.65

ANS: A

27. Refer to Exhibit 4. To perform the following test: H0: $\mu 1 - \mu 2 \leq 0$ H1: $\mu 1 - \mu 2 > 0$., the test statistic is
a. 0.75
b. -5
c. 4.86
d. -2.65

ANS: A

28. Refer to Exhibit 4. The *p*-value for the difference between the two population means

a. 0.05 to 0.1
b. -0.25 to -0.1
c. 0.1 to 0.25
d. 0.05 to 0.25

ANS: C

$\hat{y} = b_0 + b_1 x$

Gulf University for Science & Technology
Department of Economics & Finance
ECON 380: Business Statistics
Fall 2011

Dr. Fida Karam
Room N1-115
PHONE: 2530-7351
karam.f@gust.edu.kw

Problem Set V:
Chapter 14

Problem 1:

In a manufacturing process, the assembly line speed (meter per minute) was thought to affect the number of defective parts found during the inspection process. To test this theory, managers devised a situation in which the same batch of parts was inspected visually at a variety of line speeds. They collected the following data.

Line Speed x	Number of Defective Parts Found Y
20	21
20	19
40	15
30	16
60	14
40	17

$\hat{y}$ a. Develop the least squares estimated regression equation, $Y = \beta_0 + \beta_1 X + \varepsilon$ (or develop a least-squares regression line) and explain what the slope of the line indicates.

b. Calculate the estimated standard deviation (standard error) of b_1.

c. At 95% confidence, perform a t test and determine whether or not the slope is significantly different from zero.

d. Perform an F test and determine if the line speed and the number of defective parts are related. Let $\alpha = 0.05$.

e. Compute the coefficient of correlation and comment on the strength of relationship between x and y.

f. Construct a 95% confidence interval for β_1

g. Predict the number of defective parts when line speed = 70.

ANSWER:

a. $\hat{y} = 22.18 - 0.148x$. The slope indicates that as x goes up by 1, y goes down by \$0.148.

b. $s_{b1} = 0.0439$

c. H0: $\beta_1 = 0$; H1: $\beta_1 \neq 0$

$t = -3.37$; df = 4

Upper tail area between 0.01 and 0.025, *P-value* between 0.02 and 0.05

P-value < α, reject H0, x is a significant variable or x and y are linearly related.

Same conclusion with the critical value approach $t < -2.7765$ (-2.7765 is the critical value of t with df = 4 and $\alpha/2 = 0.025$).

d. H0: $\beta_1 = 0$; H1: $\beta_1 \neq 0$

$F = 11.33$; $df_N = 1$ and $df_D = 4$

P-value between 0.025 and 0.05

P-value < α, reject H0, or x and y are linearly related.

Same conclusion with the critical value approach $F < 7.71$ (7.71 is the critical value of F with $df_N = 1$, $df_D = 4$ and $\alpha = 0.05$).

e. $r^2 = .739$; 74% of the variability in y is explained by the linear relationship between x and y.

$r_{xy} = -0.86$. Negative linear relationship between x and y.

f. -0.148 ± 0.1219 or -0.2699 to -0.0261

g. 11.82

Problem 2:

An important application of regression analysis in accounting is in the estimation of cost. By collecting data on volume and cost and using the least squares method to develop an estimated regression equation relating volume and cost, an accountant can estimate the cost associated with a particular manufacturing volume. Consider the following sample of production volumes and total cost data for a manufacturing operation.

Production volume (units)	Total costs (euros)
400	4000
450	5000
550	5400
600	5900
700	6400
750	7000

a. Develop an estimated regression equation that could be used to predict the total cost for a given production volume.
b. What is the variable cost per unit produced?

c. Compute the coefficient of determination. What percentage of the variation in total cost can be explained by production volume?
d. Compute the coefficient of correlation and comment on the strength of relationship between x and y.
e. The company's production schedule shows that 500 units must be produced next month. What is the estimated total cost for this operation?
f. Perform an F test and determine if production volume and total costs are related. Let $\alpha = 0.05$.
g. At 95% confidence, perform a t test and determine whether or not the slope is significantly different from zero.
h. Construct a 95% confidence interval for β_1.

ANSWER:

a. $\hat{y} = 1246.67 + 7.6x$

b. 7.6 euros, the slope of the estimated regression line. The slope indicates that as production volume goes up by 1 unit, variable and total costs go up by 7.6 units.

c. $r^2 = .9587$; 95.87% of the variability in y is explained by the linear relationship between x and y.

d. $r_{xy} = 0.98$, strong positive linear relationship between production volume and total costs.

e. 5046.67 euros.

f. H0: $\beta_1 = 0$; H1: $\beta_1 \neq 0$

   $F = 92.83$; $df_N = 1$ and $df_D = 4$

   P-value is less than 0.01

   P-value < α, reject H0, or production volume and total costs are linearly related.

   Same conclusion with the critical value approach $F > 7.71$ (7.71 is the critical value of F with $df_N = 1$, $df_D = 4$ and $\alpha = 0.05$).

g. H0: $\beta_1 = 0$; H1: $\beta_1 \neq 0$

   $t = 9.62$; $df = 4$

   Upper tail area less than 0.0005, *P-value* less than 0.001.

   P-value < α, reject H0, production volume is a significant variable or production volume and total costs are linearly related.

   Same conclusion with the critical value approach $t > 2.7765$ (2.7765 is the critical value of t with $df = 4$ and $\alpha/2 = 0.025$).

f. 7.6 ± 1.6841 or 5.9159 to 9.2841

Multiple choice questions:

Exhibit 1

The following information regarding a dependent variable (Y) and an independent variable (X) is provided.

Y	X	$(x-\bar{x})$	$(x-\bar{x})^2$	$(y_i-\bar{y})$	$(y_i-\bar{y})^2$	$(x-\bar{x})(y-\bar{y})$
1	1	-2	4	-2	4	4
2	2	-1	1	-1	1	1
3	3	0	0	0	0	0
4	4	1	1	1	1	1
5	5	2	4	2	4	4
$\bar{y}=3$	$\bar{x}=3$	0	10	0	10	10

1. Refer to Exhibit 1. The least squares estimate of the Y intercept is $b_0 = \bar{Y} - b_1\bar{X}$

   b_0

   $b_0 = 3 - 1(3) = 0$

   $b_1 = \frac{(x_i-\bar{x})(y_i-\bar{y})}{(x_i-\bar{x})^2}$

   $b_1 = 1$

   $b_1 = \frac{10}{10} =$

   a. 1
   b. 0
   c. -1
   d. 3

2. Refer to Exhibit 1. The least squares estimate of the slope is b_1

   perfect fit, x is the same y

   a. 1
   b. -1
   c. 0
   d. 3

   $R^2 \quad \frac{SSR}{SST}$

3. Refer to Exhibit 1. The coefficient of correlation is

   a. 0
   b. -1
   c. 0.5
   d. 1

4. Refer to Exhibit 1. The coefficient of determination is $r_1 = 1$

   a. 0
   b. -1
   c. 0.5
   d. 1

5. If all the points of a scatter diagram lie on the least squares regression line, then the coefficient of correlation for these variables based on this data is

a. 0

b. 1

c. **either 1 or -1, depending upon whether the relationship is positive or negative**

d. could be any value between -1 and 1

Exhibit 2

For the following data the value of SSE = 18.

Y Dependent Variable	x Independent Variable	$x-\bar{x}$	$(x-\bar{x})^2$	$(y-\bar{y})$	$(y-\bar{y})^2$	$(x-\bar{x})(y-\bar{y})$
25	14	0	0	3	49	0
27	16	2	4	-1	1	-2
33	12	-2	4	5	25	-10
27	14	0	0	-1	1	0
$\bar{y} = 28$	$\bar{x} = 14$		8		36	-12

6. Refer to Exhibit 2. The slope of the regression equation is b_1

$\frac{-12}{8} = -1.5$

a. 1.5

b. 0.67

c. 49

d. **-1.5**

7. Refer to Exhibit 2. The y intercept is b_0

$b_0 = 28 - (-1.5)(14) = 49$

a. -1.5

b. **49**

c. 7

d. 1.5

8. Refer to Exhibit 2. The total sum of squares (SST) equals

a. **36**

b. 8

c. 18

d. -12

~~SST~~ $\hat{y} = b_0 + b_1(x)$

$49 + (-1.5 \cdot 25) = 11.5 - 28 = -$

$8.5 - 28 =$

$10.5 - 28 = -2$

$8.5 - 28 = -1$

28

272.25

380.25

748

380.25

~~SST=~~

~~$SSR = (\hat{y} - \bar{y})^2$~~

$(y - \bar{y})^2 =$

$R^2 = \frac{SSR}{SSE}$

$SST = SSr + SSE$

$36 = X + 18$

$X = 18 - 36$

~~36 - 18 = 18~~

9. Refer to Exhibit 2. The coefficient of determination (r^2) equals
   a. 0.78
   b. -0.98
   c. 0.5
   d. -0.5

$\frac{SSR}{SST} =$

$\frac{18}{36} = 0.5$

$SST = SSR + SSE$

$36 = SSR \quad 18$

$36 - 18 = 18$

10. Refer to Exhibit 2. The estimated standard error of the slope equals
   a. 1
   b. 1.125
   c. 1.06
   d. 0.5

$S_{b_1} = \sqrt{\frac{MSE}{\Sigma(x_i - \bar{x})^2}} = \sqrt{\frac{18/(4-2)}{8}} = 1.06$

$MSE = \frac{18}{4-2} = 9$

$S_{b_1} = \sqrt{\frac{9}{8}} = 1.060$

11. Refer to Exhibit 2. The t statistic for testing the significance of the slope is
   a. 1.42
   b. 1.96
   c. -1.42
   d. 0.555

$\frac{b_1}{S_{b_1}} = \frac{-1.5}{1.06} = -1.415$

12. Refer to Exhibit 2. The critical t value for testing the significance of the slope at 95% confidence is
   a. 1.96
   b. 4.3027
   c. 2.92
   d. 1.645

$t_{\alpha/2} =$

$0.025, 2$

$t_{0.05, 2}$

$\alpha = 0.05$

$df = n-2$

$= 4-2 = 2$

13. Refer to Exhibit 2. Based on the above estimated regression equation, if the independent variable is 30, then the estimated value of y is
   a. 4
   b. 54
   c. 94
   d. 30

$\hat{y} = 49 + (-1.5)(30) = 4$

14. Refer to Exhibit 2. Based on the above estimated regression equation, if the independent variable is 30 and the actual value of y is 3, then the residuals equal
   a. 1
   b. -1

error

$3 - 4 = -1$

$y - \hat{y} = 3 - 4 = -1$

$3 - 4 = -1$

c. 3

d. can't be found

Exhibit 3

The following information regarding a dependent variable Y and an independent variable X is provided

$\Sigma X = 25$ $\qquad$ $\Sigma(Y-\overline{Y})(X-\overline{X}) = -100$

$\Sigma Y = 75$ $\qquad$ $\Sigma(X-\overline{X})^2 = 50$

$n = 5$ $\qquad$ $\Sigma(Y-\overline{Y})^2 = 1000$

SSE = 100

15. Refer to Exhibit 3. The slope of the regression equation is

a. -2

$b_1 = \frac{-100}{50} = -2$

b. 2

c. 0.5

d. -100

16. Refer to Exhibit 3. The total sum of squares (SST) is

a. -156

$SST = \Sigma(y-\bar{y})^2$

b. 234

c. 1870

d. 1000

17. Refer to Exhibit 3. The sum of squares due to regression (SSR) is

a. 1000

$SSR = \Sigma(\hat{y}-\bar{y})^2 = SST - SSE$

b. 50

$1000 - 100 = 900$

c. 100

d. 900

18. Refer to Exhibit 3. The mean square due to error (MSE) is

a. 100

$MSE = \frac{SSE}{n-2} = \frac{100}{5-2} = 33.33$

b. 33.33

c. 1.746

d. 2.120

19. Refer to Exhibit 3. The F statistic for testing the significance of the slope is

a. 900

b. 33.33

c. **27**

d. 0.555

$H_0: B_1 = 0$

$H_1: B_1 \neq 0$

$F = \frac{MSR}{MSE} = \frac{900}{33.33} = 27$

20. Refer to Exhibit 3. The critical value of F for testing the significance of the slope at 95% confidence is

a. 17.44

b. **10.13**

c. 1.746

d. 2.120

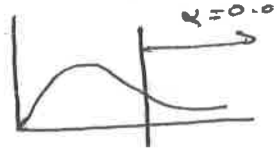
always 1 in chapter 14

$\alpha = 0.05$

$df_N = 1$

$df_D = n-2 = 3$

$\alpha = 0.05$



upper tale always

$F_{0.05}(1, 3) = 10.13$

table

21. Refer to Exhibit 3. The sample coefficient of correlation equals

a. -0.9

b. 0.9

c. 0.95

d. **-0.95**

(2) $r_{xy} = -\sqrt{R^2} = -\sqrt{0.9} = -0.95$

(1) $R^2 = \frac{SSR}{SST} = \frac{900}{1000} = 0.9$

22. Refer to Exhibit 3. The t statistic for testing the significance of the slope is

a. -2

b. 2.44

c. **-2.44**

d. 0.555

$H_0: B_1 = 0$

$H_1: B_1 \neq 0$

$t = \frac{b_1}{s_{b_1}} = \frac{-2}{0.82} = -2.44$

$s_{b_1} = \sqrt{\frac{33.33}{50}} = 0.82$

23. Refer to Exhibit 3. The critical t value for testing the significance of the slope at 90% confidence is

a. 3.1825

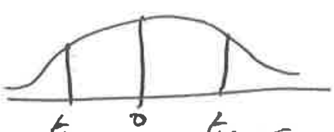
b. **2.3534**

c. 1.96

d. 1.6377

$\alpha = 0.1$ $\quad \alpha/2 = 0.05$ $\quad df = n-2 = 3$

$t_{\alpha/2} = 2.3534$



$-t_{0.05}$ $\quad 0$ $\quad t_{0.05}$

24. For a simple regression model, SST = 500 and SSE = 50. The coefficient of determination is ----- (R^2: 0.9) and the coefficient of correlation is -----

$R_{xy} = \text{Sign } b_1 \sqrt{0.9}$

$= \pm\sqrt{0.9} = \pm 0.95$

SST = 500

SSE = 50

$R^2 = \frac{SSR}{SST} = \frac{500-50}{500} = 0.9$

a. 0.9; -0.95

b. -0.9; -0.95

c. 0.9; -0.95

d. 0.9; -0.95 or +0.95 depending upon whether the relationship is positive or negative.

26. For a simple regression model, the covariance between the dependent and independent variables is 0.50 and the standard deviation of the independent variable is 0.5. The slope of the estimated regression line is

a. 1

b. -2

c. 10

d. 2

Gulf University for Science & Technology
Department of Economics & Finance
ECON 380: Business Statistics
Fall 2011

Dr. Fida Karam
Room N1-115
PHONE: 2530-7351
karam.f@gust.edu.kw

Problem Set V:
Chapter 14

Problem 1:

In a manufacturing process, the assembly line speed (meter per minute) was thought to affect the number of defective parts found during the inspection process. To test this theory, managers devised a situation in which the same batch of parts was inspected visually at a variety of line speeds. They collected the following data.

Line Speed	Number of Defective Parts Found
20	21
20	19
40	15
30	16
60	14
40	17

a. Develop the least squares estimated regression equation, $Y = \beta_0 + \beta_1 X + \varepsilon$ (or develop a least-squares regression line) and explain what the slope of the line indicates.
b. Calculate the estimated standard deviation (standard error) of b_1.
c. At 95% confidence, perform a t test and determine whether or not the slope is significantly different from zero.
d. Perform an F test and determine if the line speed and the number of defective parts are related. Let $\alpha = 0.05$.
e. Compute the coefficient of correlation and comment on the strength of relationship between x and y.
f. Construct a 95% confidence interval for β_1
g. Predict the number of defective parts when line speed = 70.

ANSWER:

a. $\hat{y} = 22.18 - 0.148x$. The slope indicates that as x goes up by 1, y goes down by \$0.148.

b. $s_{b1} = 0.0439$

c. H0: $\beta_1 = 0$; H1: $\beta_1 \neq 0$

$t = -3.37$; $df = 4$

Upper tail area between 0.01 and 0.025, *P-value* between 0.02 and 0.05

P-value < α, reject H0, x is a significant variable or x and y are linearly related.

Same conclusion with the critical value approach $t < -2.7765$ (-2.7765 is the critical value of t with $df = 4$ and $\alpha/2 = 0.025$).

d. H0: $\beta_1 = 0$; H1: $\beta_1 \neq 0$

$F = 11.33$; $df_N = 1$ and $df_D = 4$

P-value between 0.025 and 0.05

P-value < α, reject H0, or x and y are linearly related.

Same conclusion with the critical value approach $F < 7.71$ (7.71 is the critical value of F with $df_N = 1$, $df_D = 4$ and $\alpha = 0.05$).

e. $r^2 = .739$; 74% of the variability in y is explained by the linear relationship between x and y.

$r_{xy} = -0.86$. Negative linear relationship between x and y.

f. -0.148 ± 0.1219 or -0.2699 to -0.0261

g. 11.82

Problem 2:

An important application of regression analysis in accounting is in the estimation of cost. By collecting data on volume and cost and using the least squares method to develop an estimated regression equation relating volume and cost, an accountant can estimate the cost associated with a particular manufacturing volume. Consider the following sample of production volumes and total cost data for a manufacturing operation.

Production volume (units)	**Total costs (euros)**
400	4000
450	5000
550	5400
600	5900
700	6400
750	7000

a. Develop an estimated regression equation that could be used to predict the total cost for a given production volume.
b. What is the variable cost per unit produced?

c. Compute the coefficient of determination. What percentage of the variation in total cost can be explained by production volume?
d. Compute the coefficient of correlation and comment on the strength of relationship between x and y.
e. The company's production schedule shows that 500 units must be produced next month. What is the estimated total cost for this operation?
f. Perform an F test and determine if production volume and total costs are related. Let $\alpha = 0.05$.
g. At 95% confidence, perform a t test and determine whether or not the slope is significantly different from zero.
h. Construct a 95% confidence interval for β_1.

ANSWER:

a. $\hat{y} = 1246.67 + 7.6x$

b. 7.6 euros, the slope of the estimated regression line. The slope indicates that as production volume goes up by 1 unit, variable and total costs go up by 7.6 units.

c. $r^2 = .9587$; 95.87% of the variability in y is explained by the linear relationship between x and y.

d. $r_{xy} = 0.98$, strong positive linear relationship between production volume and total costs.

e. 5046.67 euros.

f. H0: $\beta_1 = 0$; H1: $\beta_1 \neq 0$

   $F = 92.83$; $df_N = 1$ and $df_D = 4$

   P-value is less than 0.01

   P-value < α, reject H0, or production volume and total costs are linearly related.

   Same conclusion with the critical value approach $F > 7.71$ (7.71 is the critical value of F with $df_N = 1$, $df_D = 4$ and $\alpha = 0.05$).

g. H0: $\beta_1 = 0$; H1: $\beta_1 \neq 0$

   $t = 9.62$; $df = 4$

   Upper tail area less than 0.0005, *P-value* less than 0.001.

   P-value < α, reject H0, production volume is a significant variable or production volume and total costs are linearly related.

   Same conclusion with the critical value approach $t > 2.7765$ (2.7765 is the critical value of t with $df = 4$ and $\alpha/2 = 0.025$).

f. 7.6 ± 1.6841 or 5.9159 to 9.2841

Multiple choice questions:

Exhibit 1

The following information regarding a dependent variable (Y) and an independent variable (X) is provided.

Y	X
1	1
2	2
3	3
4	4
5	5

1. Refer to Exhibit 1. The least squares estimate of the Y intercept is
   a. 1
   b. 0
   c. -1
   d. 3

2. Refer to Exhibit 1. The least squares estimate of the slope is
   a. 1
   b. -1
   c. 0
   d. 3

3. Refer to Exhibit 1. The coefficient of correlation is
   a. 0
   b. -1
   c. 0.5
   d. 1

4. Refer to Exhibit 1. The coefficient of determination is
   a. 0
   b. -1
   c. 0.5
   d. 1

5. If all the points of a scatter diagram lie on the least squares regression line, then the coefficient of correlation for these variables based on this data is

   a. 0

   b. 1

   c. either 1 or -1, depending upon whether the relationship is positive or negative

   d. could be any value between -1 and 1

Exhibit 2

For the following data the value of SSE = 18.

Y Dependent Variable	x Independent Variable
25	14
27	16
33	12
27	14

6. Refer to Exhibit 2. The slope of the regression equation is

   a. 1.5

   b. 0.67

   c. 49

   d. -1.5

7. Refer to Exhibit 2. The y intercept is

   a. -1.5

   b. 49

   c. 7

   d. 1.5

8. Refer to Exhibit 2. The total sum of squares (SST) equals

   a. 36

   b. 8

   c. 18

   d. -12

9. Refer to Exhibit 2. The coefficient of determination (r^2) equals

a. 0.78

b. -0.98

c. 0.5

d. -0.5

10. Refer to Exhibit 2. The estimated standard error of the slope equals

a. 1

b. 1.125

c. 1.06

d. 0.5

11. Refer to Exhibit 2. The t statistic for testing the significance of the slope is

a. 1.42

b. 1.96

c. -1.42

d. 0.555

12. Refer to Exhibit 2. The critical t value for testing the significance of the slope at 95% confidence is

a. 1.96

b. 4.3027

c. 2.92

d. 1.645

13. Refer to Exhibit 2. Based on the above estimated regression equation, if the independent variable is 30, then the estimated value of *y* is

a. 4

b. 54

c. 94

d. 30

14. Refer to Exhibit 2. Based on the above estimated regression equation, if the independent variable is 30 and the actual value of *y* is 3, then the residuals equal

a. 1

b. - 1

c. 3

d. can't be found

Exhibit 3

The following information regarding a dependent variable Y and an independent variable X is provided

$\Sigma X = 25$ $\quad\Sigma(Y-\overline{Y})(X-\overline{X}) = -100$

$\Sigma Y = 75$ $\quad\Sigma(X-\overline{X})^2 = 50$

$n = 5$ $\quad\Sigma(Y-\overline{Y})^2 = 1000$

SSE = 100

15. Refer to Exhibit 3. The slope of the regression equation is

a. -2

b. 2

c. 0.5

d. -100

16. Refer to Exhibit 3. The total sum of squares (SST) is

a. -156

b. 234

c. 1870

d. 1000

17. Refer to Exhibit 3. The sum of squares due to regression (SSR) is

a. 1000

b. 50

c. 100

d. 900

18. Refer to Exhibit 3. The mean square due to error (MSE) is

a. 100

b. 33.33

c. 1.746

d. 2.120

19. Refer to Exhibit 3. The F statistic for testing the significance of the slope is

a. 900

b. 33.33

c. 27

d. 0.555

20. Refer to Exhibit 3. The critical value of F for testing the significance of the slope at 95% confidence is

a. 17.44

b. 10.13

c. 1.746

d. 2.120

21. Refer to Exhibit 3. The sample coefficient of correlation equals

a. -0.9

b. 0.9

c. 0.95

d. -0.95

22. Refer to Exhibit 3. The t statistic for testing the significance of the slope is

a. -2

b. 2.44

c. -2.44

d. 0.555

23. Refer to Exhibit 3. The critical t value for testing the significance of the slope at 90% confidence is

a. 3.1825

b. 2.3534

c. 1.96

d. 1.6377

24. For a simple regression model, SST = 500 and SSE = 50. The coefficient of determination is ----- and the coefficient of correlation is -----

a. 0.9; -0.95

b. -0.9; -0.95

c. 0.9; -0.95

d. 0.9; -0.95 or +0.95 depending upon whether the relationship is positive or negative.

25. For a simple regression model, the covariance between the dependent and independent variables is 0.50 and the standard deviation of the independent variable is 0.5. The slope of the estimated regression line is

a. 1

b. -2

c. 10

d. 2

Gulf University for Science & Technology
Department of Economics & Finance
ECON 380: Business Statistics
Fall 2011

Dr. Fida Karam
Room N1-115
PHONE: 2530-7351
karam.f@gust.edu.kw

Problem Set V:
Chapter 14

Problem 1:

In a manufacturing process, the assembly line speed (meter per minute) was thought to affect the number of defective parts found during the inspection process. To test this theory, managers devised a situation in which the same batch of parts was inspected visually at a variety of line speeds. They collected the following data.

Line Speed	Number of Defective Parts Found
20	21
20	19
40	15
30	16
60	14
40	17

a. Develop the least squares estimated regression equation, $Y = \beta_0 + \beta_1 X + \varepsilon$ (or develop a least-squares regression line) and explain what the slope of the line indicates.
b. Calculate the estimated standard deviation (standard error) of b_1.
c. At 95% confidence, perform a t test and determine whether or not the slope is significantly different from zero.
d. Perform an F test and determine if the line speed and the number of defective parts are related. Let $\alpha = 0.05$.
e. Compute the coefficient of correlation and comment on the strength of relationship between x and y.
f. Construct a 95% confidence interval for β_1
g. Predict the number of defective parts when line speed = 70.

ANSWER:

a. $\hat{y} = 22.18 - 0.148x$. The slope indicates that as x goes up by 1, y goes down by \$0.148.

b. $s_{b1} = 0.0439$

c. H0: $\beta_1 = 0$; H1: $\beta_1 \neq 0$

$t = -3.37$; df = 4

Upper tail area between 0.01 and 0.025, *P-value* between 0.02 and 0.05

P-value < α, reject H0, x is a significant variable or x and y are linearly related.

Same conclusion with the critical value approach $t < -2.7765$ (-2.7765 is the critical value of t with df = 4 and $\alpha/2 = 0.025$).

d. H0: $\beta_1 = 0$; H1: $\beta_1 \neq 0$

$F = 11.33$; $df_N = 1$ and $df_D = 4$

P-value between 0.025 and 0.05

P-value < α, reject H0, or x and y are linearly related.

Same conclusion with the critical value approach $F < 7.71$ (7.71 is the critical value of F with $df_N = 1$, $df_D = 4$ and $\alpha = 0.05$).

e. $r^2 = .739$; 74% of the variability in y is explained by the linear relationship between x and y.

$r_{xy} = -0.86$. Negative linear relationship between x and y.

f. -0.148 ± 0.1219 or -0.2699 to -0.0261

g. 11.82

Problem 2:

An important application of regression analysis in accounting is in the estimation of cost. By collecting data on volume and cost and using the least squares method to develop an estimated regression equation relating volume and cost, an accountant can estimate the cost associated with a particular manufacturing volume. Consider the following sample of production volumes and total cost data for a manufacturing operation.

Production volume (units)	Total costs (euros)
400	4000
450	5000
550	5400
600	5900
700	6400
750	7000

a. Develop an estimated regression equation that could be used to predict the total cost for a given production volume.
b. What is the variable cost per unit produced?

c. Compute the coefficient of determination. What percentage of the variation in total cost can be explained by production volume?
d. Compute the coefficient of correlation and comment on the strength of relationship between x and y.
e. The company's production schedule shows that 500 units must be produced next month. What is the estimated total cost for this operation?
f. Perform an F test and determine if production volume and total costs are related. Let $\alpha = 0.05$.
g. At 95% confidence, perform a t test and determine whether or not the slope is significantly different from zero.
h. Construct a 95% confidence interval for β_1.

ANSWER:

a. $\hat{y} = 1246.67 + 7.6x$

b. 7.6 euros, the slope of the estimated regression line. The slope indicates that as production volume goes up by 1 unit, variable and total costs go up by 7.6 units.

c. $r^2 = .9587$; 95.87% of the variability in y is explained by the linear relationship between x and y.

d. $r_{xy} = 0.98$, strong positive linear relationship between production volume and total costs.

e. 5046.67 euros.

f. H0: $\beta_1 = 0$; H1: $\beta_1 \neq 0$

$F = 92.83$; $df_N = 1$ and $df_D = 4$

P-value is less than 0.01

P-value $< \alpha$, reject H0, or production volume and total costs are linearly related.

Same conclusion with the critical value approach $F > 7.71$ (7.71 is the critical value of F with $df_N = 1$, $df_D = 4$ and $\alpha = 0.05$).

g. H0: $\beta_1 = 0$; H1: $\beta_1 \neq 0$

$t = 9.62$; $df = 4$

Upper tail area less than 0.0005, *P-value* less than 0.001.

P-value $< \alpha$, reject H0, production volume is a significant variable or production volume and total costs are linearly related.

Same conclusion with the critical value approach $t > 2.7765$ (2.7765 is the critical value of t with $df = 4$ and $\alpha/2 = 0.025$).

f. 7.6 ± 1.6841 or 5.9159 to 9.2841

Multiple choice questions:

Exhibit 1

The following information regarding a dependent variable (Y) and an independent variable (X) is provided.

Y	**X**
1	1
2	2
3	3
4	4
5	5

1. Refer to Exhibit 1. The least squares estimate of the Y intercept is
   a. 1
   b. 0
   c. -1
   d. 3

2. Refer to Exhibit 1. The least squares estimate of the slope is
   a. 1
   b. -1
   c. 0
   d. 3

3. Refer to Exhibit 1. The coefficient of correlation is
   a. 0
   b. -1
   c. 0.5
   d. 1

4. Refer to Exhibit 1. The coefficient of determination is
   a. 0
   b. -1
   c. 0.5
   d. 1

5. If all the points of a scatter diagram lie on the least squares regression line, then the coefficient of correlation for these variables based on this data is

   a. 0

   b. 1

   c. either 1 or -1, depending upon whether the relationship is positive or negative

   d. could be any value between -1 and 1

Exhibit 2

For the following data the value of SSE = 18.

Y Dependent Variable	x Independent Variable
25	14
27	16
33	12
27	14

6. Refer to Exhibit 2. The slope of the regression equation is

   a. 1.5

   b. 0.67

   c. 49

   d. -1.5

7. Refer to Exhibit 2. The y intercept is

   a. -1.5

   b. 49

   c. 7

   d. 1.5

8. Refer to Exhibit 2. The total sum of squares (SST) equals

   a. 36

   b. 8

   c. 18

   d. -12

9. Refer to Exhibit 2. The coefficient of determination (r^2) equals

a. 0.78

b. -0.98

c. 0.5

d. -0.5

10. Refer to Exhibit 2. The estimated standard error of the slope equals

a. 1

b. 1.125

c. 1.06

d. 0.5

11. Refer to Exhibit 2. The t statistic for testing the significance of the slope is

a. 1.42

b. 1.96

c. -1.42

d. 0.555

12. Refer to Exhibit 2. The critical t value for testing the significance of the slope at 95% confidence is

a. 1.96

b. 4.3027

c. 2.92

d. 1.645

13. Refer to Exhibit 2. Based on the above estimated regression equation, if the independent variable is 30, then the estimated value of *y* is

a. 4

b. 54

c. 94

d. 30

14. Refer to Exhibit 2. Based on the above estimated regression equation, if the independent variable is 30 and the actual value of *y* is 3, then the residuals equal

a. 1

b. - 1

c. 3

d. can't be found

Exhibit 3

The following information regarding a dependent variable Y and an independent variable X is provided

$\Sigma X = 25$ $\Sigma(Y - \overline{Y})(X - \overline{X}) = -100$

$\Sigma Y = 75$ $\Sigma(X - \overline{X})^2 = 50$

$n = 5$ $\Sigma(Y - \overline{Y})^2 = 1000$

$SSE = 100$

15. Refer to Exhibit 3. The slope of the regression equation is

a. -2

b. 2

c. 0.5

d. -100

16. Refer to Exhibit 3. The total sum of squares (SST) is

a. -156

b. 234

c. 1870

d. 1000

17. Refer to Exhibit 3. The sum of squares due to regression (SSR) is

a. 1000

b. 50

c. 100

d. 900

18. Refer to Exhibit 3. The mean square due to error (MSE) is

a. 100

b. 33.33

c. 1.746

d. 2.120

19. Refer to Exhibit 3. The F statistic for testing the significance of the slope is

a. 900

b. 33.33

c. 27

d. 0.555

20. Refer to Exhibit 3. The critical value of F for testing the significance of the slope at 95% confidence is

a. 17.44

b. 10.13

c. 1.746

d. 2.120

21. Refer to Exhibit 3. The sample coefficient of correlation equals

a. -0.9

b. 0.9

c. 0.95

d. -0.95

22. Refer to Exhibit 3. The t statistic for testing the significance of the slope is

a. -2

b. 2.44

c. -2.44

d. 0.555

23. Refer to Exhibit 3. The critical t value for testing the significance of the slope at 90% confidence is

a. 3.1825

b. 2.3534

c. 1.96

d. 1.6377

24. For a simple regression model, SST = 500 and SSE = 50. The coefficient of determination is ----- and the coefficient of correlation is -----

a. 0.9; -0.95

b. -0.9; -0.95

c. 0.9; -0.95

d. 0.9; -0.95 or +0.95 depending upon whether the relationship is positive or negative.

25. For a simple regression model, the covariance between the dependent and independent variables is 0.50 and the standard deviation of the independent variable is 0.5. The slope of the estimated regression line is

a. 1

b. -2

c. 10

d. 2

Gulf University for Science & Technology
Department of Economics & Finance
ECON 380: Business Statistics
Fall 2011

Dr. Fida Karam
Room N1-115
PHONE: 2530-7351
karam.f@gust.edu.kw

Problem Set VI:
Chapters 11 – 12 - 13

Multiple choice questions:

1. A sample of 20 items provides a sample mean of 15 and a sample variance of 6. Compute a 95% confidence interval estimate for the standard deviation of the population.

$n \qquad \bar{x} \qquad s^2$

a. 3.47 to 12.8
b. 2.88 to 3.88
c. 1.86 to 3.58
d. 5.7 to 6.89

$\frac{(n-1)s^2}{\chi^2_{\alpha/2}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{(1-\alpha/2)}}$

$df = 20-1 = 19$

$95\% = 1-\alpha \Rightarrow \alpha = 0.05 \Rightarrow \alpha/2 = 0.025 \Rightarrow \chi^2_{\alpha/2} = 32.852$

$1-\alpha/2 = 0.975 \Rightarrow \chi^2_{0.975} = 8.907 \rightarrow 19$

Exhibit 1

On the basis of data provided by a salary survey, the variance in annual salaries for seniors in accounting firms is approximately 20 and the variance in annual salaries for managers in accounting firms is approximately 30. Assuming that the salary data were based on samples of 16 seniors and 16 managers, test the hypothesis that the population variance in the salaries for managers is greater than the population variance in salaries for seniors. σ^2

	Managers	Seniors
Sample Size	16	16
Sample Mean	520	540
Sample Variance	30	20

2. Refer to Exhibit 1. The null hypothesis is

a. $S_1^2 > S_2^2$
b. $S_1^2 \leq S_2^2$
c. $\sigma_1^2 > \sigma_2^2$
d. $\sigma_1^2 \leq \sigma_2^2$

σ

$\sigma_1 >$

3. Refer to Exhibit 1. The test statistic is

a. 1.5
b. 0.96

c. 1
d. 4

4. Refer to Exhibit 1. The *p*-value for this test is
a. greater than 0.1
b. less than 0.1
c. between 0.025 and 0.05
d. None of these alternatives is correct.

5. Refer to Exhibit 1. At 99% confidence the null hypothesis
a. should be rejected
b. should not be rejected
c. should be revised
d. None of these alternatives is correct.

Exhibit 2
The filling variance for boxes of breakfast cereal is designed to be 0.25. A sample of 25 boxes of cereal shows a sample variance of 0.4 grams. We need to determine whether the variance in the cereal box fillings is not meeting the design specification.

$n = 25 \quad S^2 = 0.4$

6. Refer to Exhibit 2. The null hypothesis is
a. $S^2 = 0.25$
b. $S^2 \leq 0.25$
c. $\sigma^2 = 0.25$
d. $\sigma^2 \leq 0.25$

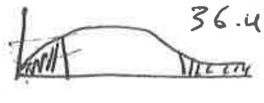
$H_0: \sigma^2 = 0.25$
$H_1: \sigma^2 \neq 0.25$

7. Refer to Exhibit 2. The test statistic is
a. 15.36
b. 38.4
c. 40
d. 24

$$\chi^2 = \frac{(n-1)^2 S^2}{\sigma_0^2} = \frac{(25-1)\,0.4}{0.25} = 38.4$$

$df = 24$

8. Refer to Exhibit 2. The *p*-value for this test is
a. 0.05
b. between 0.025 and 0.05
c. between 0.05 and 0.1
d. 1.96



$df = 25 - 1 = 24$
$36.415 < 38.4 < 39.364$
$0.05 > \text{Area} > 0.025$
$0.1 > P\text{-}V > 0.05$

9. Refer to Exhibit 2. At 95% confidence, the null hypothesis
a. should be rejected
b. should not be rejected
c. should be revised

$95\% = 1 - \alpha$
$\alpha = 0.05$

$P\text{-}V > \alpha \rightarrow$ do not reject H

33

χ^2 (B)

d. None of these alternatives is correct.

Exhibit 3:
During the first 13 weeks of the autumn schedules, the Saturday evening 8:00 p.m. to 9:00 p.m. audience proportions were recorded as: BBC1 & 2: 43%; Sky channels: 34%; and others, 23%. A sample of 400 homes two weeks after a Saturday night schedule revision yielded the following viewing audience data: BBC1 & 2: 164 homes; Sky channels; 172 homes; and others, 64 homes. Test with α = 0.01 to determine whether the viewing audience proportions changed.

10. Refer to Exhibit 3. The expected frequency of BBC1 & 2 is
a. 172
b. 43%
c. 164
d. 64

$H_0: \pi_A = 43\%$ π_B 34% π_C 23%
H_1: H_0 false

400

H_0 $\pi_A = 43$ $\pi_B = 34$ π_C 23%

11. Refer to Exhibit 3. The calculated value for the test statistic equals
a. 0.5444
b. 300
c. 18.42
d. 6.6615

	π_i	f_i	e_i	$f_i - e_i$	$(f_i-e_i)^2$	$(f_i-e_i$
1	0.43	164	0.43(400)=172			
2	0.34	172				
3	0.23	64				
	1	400				

12. Refer to Exhibit 3. The *p*-value is
a. less than .005
b. 0.01
c. between .05 and 0.1
d. greater than 0.1

13. Refer to Exhibit 3. At 95% confidence, the null hypothesis
a. should not be rejected
b. should be rejected
c. was designed wrong
d. None of these alternatives is correct.

$\alpha = 0.05$

p-value < 0.005

14. The chi-square value for a one-tailed (lower tail) hypothesis test at 95% confidence and a sample size of 25 is
a. 13.848
b. 36.415
c. 39.364
d. 12.401

0.05
0.025, 24 = $\chi_{0.95, 24}$

15. The chi-square value for a one-tailed test (upper tail) when the level of significance is 0.1 and the sample size is 15 is

$\chi_{0.1, 14}$

a. 21.064
b. 23.685
c. 7.790
d. 6.571

16. The ANOVA procedure is a statistical approach for determining whether or not
a. the means of two samples are equal
b. the means of two or more samples are equal
c. the means of more than two samples are equal
d. the means of two or more populations are equal

17. The critical value of F for a one-tailed (upper tail) hypothesis test at 90% confidence when there is a sample size of 16 for the sample with the smaller variance, and there is a sample size of 8 for the sample with the larger sample variance is
a. 2.16
b. 2.71
c. 2.63
d. 3.51

18. In a completely randomized design involving four treatments, the following information is provided.

	Treatment 1	Treatment 2	Treatment 3	Treatment 4
Sample Size	45	20	13	19
Sample Mean	30	35	40	50

The overall mean for all treatments is

a. 36.29
b. 38.75
c. 40
d. 24.25

Exhibit 4

In a completely randomized experimental design involving four treatments, 5 observations were recorded for each of the four treatments. The following information is provided.

SSE = 600
SST = 800

19. Refer to Exhibit 4. The sum of squares between treatments (SSTR) is
a. 20
b. 800

c. 600
d. 200

20. Refer to Exhibit 4. The number of degrees of freedom corresponding to between treatments is
a. 16
b. 3
c. 5
d. 4

21. Refer to Exhibit 4. The number of degrees of freedom corresponding to within treatments is
a. 16
b. 59
c. 4
d. 3

22. Refer to Exhibit 4. The mean square between treatments (MSTR) is
a. 3.34
b. 16.67
c. 66.67
d. 12.00

23. Refer to Exhibit 4. The mean square within treatments (MSE) is
a. 50
b. 37.5
c. 200
d. 16.67

24. Refer to Exhibit 4. The test statistic is
a. 1.78
b. 5.0
c. 0.56
d. 15

25. Refer to Exhibit 4. If at 95% confidence we want to determine whether or not the means of the four populations are equal, the *p*-value is
a. between 0.05 to 0.10
b. greater than 0.1
c. between 0.01 to 0.025
d. less than 0.01

Problem 1:

Guitars R. US has three stores located in three different areas. Random samples of the sales of the three stores (in $1000) are shown below.

Store 1	Store 2	Store 3
80	85	79
75	86	85
76	81	88
89	80	
80		

a. Compute the overall mean $\bar{\bar{X}}$.
b. State the null and alternative hypotheses to be tested.
c. Show the complete ANOVA table for this test including the test statistic.
d. The null hypothesis is to be tested at 95% confidence. Determine the critical value for this test. What do you conclude?
e. Determine the p-value and use it for the test.

ANS:

a. *82*

b. H_o: $\mu_1 = \mu_2 = \mu_3$

H_a: *At least one mean is different from the others.*

c.

Source of Variation	*SS*	*df*	*MS*	*F*
Between Groups	*36*	*2*	*18*	*0.8526*
Within Groups	*190*	*9*	*21.11*	
Total	*226*	*11*		

d. *Critical F = 4.26, do not reject Ho and conclude there is no evidence of significant difference.*

e. *p-value > 0.1, therefore do not reject* H_o

Problem 2:

In a completely randomized experimental design, 18 experimental units were used for the first treatment, 10 experimental units for the second treatment, and 15 experimental units for the third treatment. Part of the ANOVA table for this experiment is shown below.

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F
Between Treatments	_____?	_____?	_____?	
				3.0
Error (Within Treatments)	_____?	_____?	6	
Total	_____?	_____?		

a. Fill in **all** the blanks in the above ANOVA table.
b. At 95% confidence, test to see if there is a significant difference among the means.

ANS:
a.

Source of Variation	*Sum of Squares*	*Degrees of Freedom*	*Mean Square*	*F*
Between Treatments	*36*	*2*	*18*	
				3.0
Error (Within Treatments)	*240*	*40*	*6*	
Total	*276*	*42*		

b. *For $F = 3$, the p-value is between 0.05 and 0.1; do not reject H_0 and conclude there is not a significant difference among the means. (Also, test statistic $F = 3 < 3.23$.)*

Gulf University for Science & Technology
Department of Economics & Finance
ECON 380: Business Statistics
Fall 2011

Dr. Fida Karam
Room N1-115
PHONE: 2530-7351
karam.f@gust.edu.kw

Problem Set VI:
Chapter 15

$n = 12$
$p = 2 \ (x_1, x_2)$

Problem 1:

Below you are given a partial computer output based on a sample of 12 observations relating the sales volume of computers (Y, (000s)), unit price (X_1,(€000s)) and advertising expenditures (X_2, (€000s)).

	Coefficient	Standard Error
Constant	17.145	7.865 S_{b_0}
X_1	-0.104	3.282 S_{b_1}
X_2	1.376	0.250 S_{b_2}

a. Use the output shown above and write an equation that can be used to predict the monthly sales of computers.
b. Interpret the coefficients of X_1 and X_2.
c. If the company charges $2,000 for each computer and uses 10 advertising spots, how many computers would you expect them to sell?
d. At $\alpha = 0.05$, test to determine if the price is a significant variable.
e. At $\alpha = 0.05$, test to determine if the number of advertising spots is a significant variable.
f. Construct a 95% confidence interval estimate of β_2.

ANSWER:

a. $\hat{Y} = 17.145 - 0.104X_1 + 1.376X_2$

b. *As unit price increases by $1000, the number of units sold decreases by 0.104 units, given that everything else is kept constant. As advertising expenditures increase by 1, the number of units sold increases by 1.376, given that everything else is kept constant.*

c. *30.697 (round to 31)*

d. H0: $\beta_1 = 0$; H1: $\beta_1 \neq 0$

$t = -0.032$; df =9; p-value > 0.8; do not reject H0; price is not significant (critical t = 2.2622; t > -2.2622, do not reject H0).

e. H0: $\beta_2 = 0$; H1: $\beta_2 \neq 0$

$t = 5.504$; p-value < 0.001; reject H_0; advertising is a significant variable (critical t = 2.2622; t > 2.2622, reject H0).

f. *1.376 ±0.56555 or 0.81045 to 1.94155*

Problem 2:

Below you are given a partial computer output based on a sample of 12 observations relating the sales volume of computers (Y, (000s)), unit price (X_1) and advertising expenditures (X_2, (€000s)).

ANOVA

	DF	*SS*	*MS*	*F*	*Significance F*
Regression	2	655.955	327.977	16.133	
Residual	9	182.962	20.329		
Total	11	838.917			

a. At α = 0.05 level of significance, test to determine if the model is significant. That is, determine if there exists a significant relationship between the independent variables and the dependent variable.

b. Determine the multiple coefficient of determination.

ANSWER:

a. *H0: $\beta_1 = \beta_2 = 0$; H1: H0 false.*

$F = 16.133$; $df_N = 2$ and $df_D = 9$; p-value < 0.01; reject H0; the model is significant (critical F = 4.26; F > 4.26; reject H0)

b. *0.782*

94.6% of the variability on Y is explained by the model

Problem 3:

Multiple regression analysis is used to study how sales at a fast-food outlet (Y, (€000s)) is influenced by the population within one kilometre (X_1, (000s)), the number of drive-up windows available (X_2) and the number of competitors within one kilometre (X_3),. The following results were obtained.

n

ANOVA

	DF	*SS*	MS	F
Regression	3 P	45.9634	15.321	64.280
Residual	11	2.6218	0.2383	
Total	14	48.5848		

	Coefficients	*Standard Error*	
Intercept	0.0136 β_0		
X_1	0.7992 x_1	0.074	b_1 x
X_2	0.2280 x_2	0.190	~~bex~~ S_{b_1}
X_3	-0.5796 x_3	0.920	~~ex~~ S_{b_2}

$\hat{Y} = 0.0136 + 0.7992X_1 + 0.2280X_2 - 0.5796$

a. Compute the ordinary least squares estimates of β_0, β_1, β_2, and β_3, in the model:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$$

3.59 < 64

b. Compute R^2. What can you say about the strength of this relationship? 0.946

c. Carry out a test of whether Y is significantly related to the independent variables. Use a 0.05 level of significance.

d. Carry out a test to see if X_3 and Y are significantly related. Use a 0.05 level of significance.

~~dfn=~~

$t = \frac{-0.5796}{0.920} = -0.63$

6 $\frac{\alpha}{2} = 0.025$

$F = \frac{MSR}{MSE} = 64.28$

$df = p$

~~0.2~~6 < 0.63 < 0.7649

Are

$df_n = P = 3$ $F_{0.05}(3,11)$

0.8 > P_value > 0.5 ~~reject~~ do

$df_n =$

3

e. Construct a 95% confidence interval estimate of β_2. 0.025

ANSWER:

a. $\hat{Y} = 0.0136 + 0.7992X_1 + 0.228X_2 - 0.5796X_3$

b. $R^2 = 0.9460$. *Therefore, 94.6% of the variability in Y is explained by the independent variables*

c. $F = 64.28$; *p-value* < 0.01 *(almost zero); reject* H_o; *the model is significant (critical* $F = 3.59$; $F > 3.59$; *reject H0)*

d. $t = -0.63$; *p-value between 0.5 and 0.8; do not reject* H_o; X_3 *is not a significant variable (critical* $t = 2.201$; $t > -2.201$*)*

e. 0.2280 ± 0.41819 *or -0.19019 to 0.64619*

$\in$ ~~0.025~~,3

0.0826 − 0.3733

0.228

\+ 0.7649 X 0.190 =

Problem 4:

A partial computer output from a regression analysis follows.

	Coefficient	Standard Error
Constant	20.000	5.455
X_1	0.006	0.002
X_2	-0.70	0.200

ANOVA

	DF	*SS*	*MS*	*F*
Regression				
Residual	20	40		
Total		800		

a. Compute the ordinary least squares estimates of β_0, β_1 and β_2 in the model:

$$Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \varepsilon$$

b. Interpret the coefficients of X_2.

c. If X_1 is 10,000 and X_2 is \$50, what is the estimated value of y?

d. What has been the sample size for this regression analysis?

e. At $\alpha = 0.05$, test to determine if X_1 is a significant variable.

f. At $\alpha = 0.05$ level of significance, test to determine if the model is significant.

g. Determine the multiple coefficient of determination.

ANSWER:

a. $\hat{Y} = 20 + 0.006X_1 - 0.7X_2$

b. As X_2 *increases by \$1, Y decreases by 0.7 units, given that* X_1 *is kept constant.*

c. 45

d. 23

e. $t = 3$; $df = 20$; *p-value between 0 .001 and 0.01; reject H0;* X_1 *is significant (critical* $t = 2.086$; $t > 2.086$; *reject H0)* *f.* *F = 190; p-value < 0.01; reject H0; the model is significant (critical F = 3.49; F > 3.49, reject H0)*

g. *0.95*

Problem 5:

A partial computer output from a regression analysis follows. The regression equation is:

$\hat{Y} = 8.103 + 7.602\,X_1 + 3.111\,X_2$

Predictor	**Coef**	**SE Coef**	**t**
Constant	8.103	2.667	8.103 / 2.667 = 3.04
X1	7.602	2.105	3.61
X2	3.111	0.613	5.08

s = 3.335 R-Sq = 92.3%

SST - SSR

Analysis of Variance

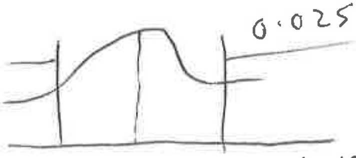
Source	DF	SS	MS	F
Regression	2	1612	806	71.82
Residual	12	134.67	11.225	
Total	14	1746.67		

$R^2 = \frac{SSR}{SST}$ $= SST = \frac{1612}{0.923} = 1746.67$

a. Compute the appropriate *t*-ratios.
b. Compute the entries in the DF, SS, and MS columns.
c. Test for the significance of β_1 and β_2 at $\alpha = 0.05$.
d. Test for the significance of the model at $\alpha = 0.05$.

c) $H_0: B_1 = 0$ $t_1 = 3.61$

$H_1: B_1 \neq 0$ $df = 12$



$0.025, 12$

2.1788

$3.61 > 2.1788$

reject H_0

X_1 significant

ANSWER:

a. Predictor	*Coef*	*SE Coef*	*t*
Constant	*8.103*	*2.667*	*3.04*
X1	*7.602*	*2.105*	*3.61*
X2	*3.111*	*0.613*	*5.08*

s = 3.35 *R-sq = 92.3%*

$H_0: B_2 = 0$

$H_1\ B_2 \neq 0$

$t_2 = 5.08 > 2.1788$

(d) $\alpha = 0.05$

$H_0: \beta_1 = \beta_2 = 0$

$H_1 =$ Ho false

$F = 71.82$

$F_{0.05}(2, 12) = 3.89$

$71.82 > 3.89$

reject Ho

model signific

b. *Analysis of Variance*

SOURCE	*DF*	*SS*	*MS*	*F*
Regression	2	1612	806	71.82
Residual Error	12	134,67	11.2225	
Total	14	1746.67		

c. *Using t table (df = 12), critical t = 2.1788. 3.61 and 5.08 are greater than 2.1788, reject H0, both variables are significant.*

d. *Critical F = 3.89, F > 3.89, reject H0, the model is significant.*

Problem 6:

The college admissions officer developed the following estimated regression equation relating the final college performance GPA to the students SAT mathematics score and secondary education level GPA.

$\hat{y} = -1.41 + 0.0235x_1 + 0.00486x_2$

where x_1 = secondary education level GPA

x_2 = SAT mathematics score

y = final college performance GPA

A portion of the Minitab computer output follows.

Predictor	Coef	SE Coef	t
Constant	-1.4053	0.4848	-2.9
X1	0.023467	0.008666	______
X2	________	0.001077	______

s = 0.1298 R-Sq = ______%

Analysis of Variance

Source	DF	SS	MS	F
Regression	2	1.76209	_____	_____
Residual	_____	_______	_____	
Total	9	1.88000		

a. Complete the missing entries in this output.
b. Compute F and test using $\alpha = 0.05$ to see whether a significant relationship is present.
c. Did the estimated regression equation provide a good fit to the data? Explain.
d. Use the t test and $\alpha = 0.05$ to test H_0: $\beta_1 = 0$ and H_0: $\beta_2 = 0$.

ANSWER:

a. *The regression equation is*

$$Y = -1.41 + .0235\,X1 + .00486\,X2$$

Predictor	*Coef*	*SE Coef*	*T*
Constant	*-1.4053*	*0.4848*	*-2.90*
X1	*0.023467*	*0.008666*	*2.71*
X2	*.00486*	*0.001077*	*4.51*

s = 0.1298 *R-sq = 93.7%*

Analysis of Variance

SOURCE	*DF*	*SS*	*MS*	*F*
Regression	*2*	*1.76209*	*.881*	*52.44*
Residual Error	*7*	*.1179*	*.0168*	
Total	*9*	*1.88000*		

b. *Using F table (2 degrees of freedom numerator and 7 degrees of freedom denominator), p-value is less than .01, reject H0, there is a significant relationship.*

c.

$$R^2 = \frac{SSR}{SST} = .937$$

d. $t_{.025} = 2.3646$ *(df = 7)*

for β_1*: p-value is between .02 and .05; reject* H_0: $\beta_1 = 0$; X_1 *is a significant variable.*

for β_2*: p-value is between .001 and .01; reject* H_0: $\beta_2 = 0$; X_2 *is a significant variable.*

Gulf University for Science & Technology
Department of Economics & Finance
ECON 380: Business Statistics
Fall 2011

Dr. Fida Karam
Room N1-115
PHONE: 2530-7351
karam.f@gust.edu.kw

Problem Set VI:
Chapter 15

Problem 1: $n=12$ $p=2$

Below you are given a partial computer output based on a sample of 12 observations relating the sales volume of computers (Y, (000s)), unit price (X_1,(€000s)) and advertising expenditures (X_2, (€000s)).

	$b_0+b_1+b_2$ Coefficient	$S_{b_1}+S_{b_2}$ Standard Error
$p=2$ Constant	17.145	S_{b_3} 7.865
X_1	-0.104	3.282
X_2	1.376	0.250

a. Use the output shown above and write an equation that can be used to predict the monthly sales of computers.
b. Interpret the coefficients of X_1 and X_2.
c. If the company charges $2,000 for each computer and uses 10 advertising spots, how many computers would you expect them to sell?
d. At $\alpha = 0.05$, test to determine if the price is a significant variable.
e. At $\alpha = 0.05$, test to determine if the number of advertising spots is a significant variable.
f. Construct a 95% confidence interval estimate of β_2.

ANSWER:

a. $\hat{Y}=17.145-0.104X_1+1.376X_2$

b. *As unit price increases by $1000, the number of units sold decreases by 0.104 units, given that everything else is kept constant. As advertising expenditures increase by 1, the number of units sold increases by 1.376, given that everything else is kept constant.*

c. *30.697 (round to 31)*

d. H0: $\beta_1 = 0$; H1: $\beta_1 \neq 0$

$t = -0.032$; $df = 9$; p-value > 0.8; do not reject H0; price is not significant (critical $t = 2.2622$; $t > -2.2622$, do not reject H0).

e. H0: $\beta_2 = 0$; H1: $\beta_2 \neq 0$

$t = 5.504$; p-value < 0.001; reject H_0; advertising is a significant variable (critical $t = 2.2622$; $t > 2.2622$, reject H0).

f. *1.376 ± 0.56555 or 0.81045 to 1.94155*

Problem 2:

Below you are given a partial computer output based on a sample of 12 observations relating the sales volume of computers (Y, (000s)), unit price (X_1) and advertising expenditures (X_2, (€000s)).

ANOVA F table

p / n-p-1

	DF	SS	MS	F	Significance F
Regression	2	655.955	327.9775	16.133	
Residual	9	182.9775	20.394		
Total		838.917			

a. At α = 0.05 level of significance, test to determine if the model is significant. That is, determine if there exists a significant relationship between the independent variables and the dependent variable.

b. Determine the multiple coefficient of determination.

ANSWER:

a. *H0: $\beta_1 = \beta_2 = 0$; H1: H0 false.*

$F = 16.133$; $df_N = 2$ and $df_D = 9$; p-value < 0.01; reject H0; the model is significant (critical F = 4.26; F > 4.26; reject H0)

b. *0.782*

Problem 3:

Multiple regression analysis is used to study how sales at a fast-food outlet (Y, (€000s)) is influenced by the population within one kilometre (X_1, (000s)), the number of drive-up windows available (X_2) and the number of competitors within one kilometre (X_3),. The following results were obtained.

ANOVA

	DF	SS
Regression	3	45.9634
Error Residual	11	2.6218
Total		

	Coefficients	Standard Error
Intercept	0.0136	
X_1	0.7992	0.074
X_2	0.2280	0.190
X_3	-0.5796	0.920

a. Compute the ordinary least squares estimates of β_0, β_1, β_2, and β_3, in the model:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$$

b. Compute R^2. What can you say about the strength of this relationship?

c. Carry out a test of whether Y is significantly related to the independent variables. Use a 0.05 level of significance.

d. Carry out a test to see if X_3 and Y are significantly related. Use a 0.05 level of significance.

e. Construct a 95% confidence interval estimate of β_2.

ANSWER:

a. $\hat{Y} = 0.0136 + 0.7992X_1 + 0.228X_2 - 0.5796X_3$

b. $R^2 = 0.9460$. *Therefore, 94.6% of the variability in Y is explained by the independent variables*

c. $F = 64.28$; *p-value* < 0.01 *(almost zero); reject* H_o*; the model is significant (critical* $F = 3.59$; $F > 3.59$; *reject H0)*

d. $t = -0.63$; *p-value between 0.5 and 0.8; do not reject* H_o*;* X_3 *is not a significant variable (critical* $t = 2.201$; $t > -2.201$*)*

e. 0.2280 ± 0.41819 *or* -0.19019 *to* 0.64619

Problem 4:

A partial computer output from a regression analysis follows.

	Coefficient	Standard Error
Constant	20.000	5.455
X_1	0.006	0.002
X_2	-0.70	0.200

ANOVA

	DF	*SS*	*MS*	*F*
Regression				
Residual	20	40		
Total		800		

a. Compute the ordinary least squares estimates of β_0, β_1 and β_2 in the model:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$$

b. Interpret the coefficients of X_2.

c. If X_1 is 10,000 and X_2 is \$50, what is the estimated value of y?

d. What has been the sample size for this regression analysis?

e. At $\alpha = 0.05$, test to determine if X_1 is a significant variable.

f. At $\alpha = 0.05$ level of significance, test to determine if the model is significant.

g. Determine the multiple coefficient of determination.

ANSWER:

a. $\hat{Y} = 20 + 0.006X_1 - 0.7X_2$

b. *As* X_2 *increases by \$1, Y decreases by 0.7 units, given that* X_1 *is kept constant.*

c. *45*

d. *23*

e. $t = 3$; $df = 20$; *p-value between 0 .001 and 0.01; reject H0;* X_1 *is significant (critical* $t = 2.086$; $t > 2.086$; *reject H0)*

f. *F = 190; p-value < 0.01; reject H0; the model is significant (critical F = 3.49; F > 3.49, reject H0)*

g. *0.95*

Problem 5

A partial computer output from a regression analysis follows. The regression equation is:

$\hat{Y} = 8.103 + 7.602\, X_1 + 3.111\, X_2$

$t_i = \frac{b_i}{s_{b_i}} =$

Predictor	Coef	SE Coef	t
Constant	8.103	2.667	8.103/2.667
X1	b_0 7.602	2.105	7.602/2.105
X2	3.111	0.613	3.111/0.613
s = 3.335	R-Sq = 92.3%		

R^2

Analysis of Variance

Source	DF	SS	MS	F
Regression X_1, X_2	2	1612	806	71.82
Residual df	12		11.225	
Total	14	1746.		

SST

$R^2 = \frac{SSR}{SST} = \frac{1612}{SST} = 0.923$

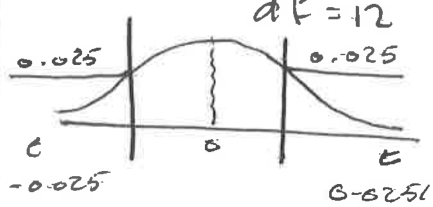
$\frac{1612}{0.923} = 1746.67$

a. Compute the appropriate *t*-ratios.
b. Compute the entries in the DF, SS, and MS columns.
c. Test for the significance of β_1 and β_2 at $\alpha = 0.05$.
d. Test for the significance of the model at $\alpha = 0.05$.

c) $H_0: \beta_1 = 0$ $\quad t_1 = 3.61$

$H_1: \beta_1 \neq 0$

$df = 12$



−2.1788 $\quad$ 2.1788

3.61 > 2.1788 reject H_0 X_1 sign

$H_0: \beta_2 = 0$

$H_1: \beta_2 \neq 0$

$t_2 = 5.08$

5.08

5.08 > 2.1788 reject H_0 X_2 sign

ANSWER:

a.

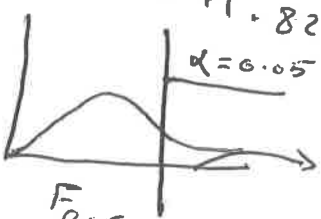
Predictor	*Coef*	*SE Coef*	*t*
Constant	*8.103*	*2.667*	*3.04*
X1	*7.602*	*2.105*	*3.61*
X2	*3.111*	*0.613*	*5.08*
s = 3.35	*R-sq = 92.3%*		

d) $\alpha = 0.05$

$H_0: B_1 = B_2 = 0$

$H_1: H_0$ false

$F = 71.82$



$F_{0.05} = 3.89\,(2,12)$

71.82 > 3.89 reject H_0 model sign

b. Analysis of Variance

SOURCE	*DF*	*SS*	*MS*	*F*
Regression	*2*	*1612*	*806*	*71.82*
Residual Error	*12*	*134,67*	*11.2225*	
Total	*14*	*1746.67*		

c. Using t table (df = 12), critical t = 2.1788. 3.61 and 5.08 are greater than 2.1788, reject H0, both variables are significant.

d. Critical F = 3.89, F > 3.89, reject H0, the model is significant.

Problem 6:

The college admissions officer developed the following estimated regression equation relating the final college performance GPA to the students SAT mathematics score and secondary education level GPA.

$$\hat{y} = -1.41 + 0.0235x_1 + 0.00486x_2$$

where x_1 = secondary education level GPA

x_2 = SAT mathematics score

y = final college performance GPA

A portion of the Minitab computer output follows.

Predictor	Coef	SE Coef	t
Constant	-1.4053	0.4848	-2.9
X1	0.023467	0.008666	2.7
X2	0.00486	0.001077	4.512

s = 0.1298 R-Sq = 93.9 %

Analysis of Variance

Source	DF	SS	MS	F
Regression	2	1.76209	0.880	5.23
Residual	7	0.11791	0.1684	
Total	9	1.88000		

a. Complete the missing entries in this output.
b. Compute F and test using $\alpha = 0.05$ to see whether a significant relationship is present.
c. Did the estimated regression equation provide a good fit to the data? Explain.
d. Use the t test and $\alpha = 0.05$ to test H_0: $\beta_1 = 0$ and H_0: $\beta_2 = 0$.

ANSWER:

a. *The regression equation is*

Y = -1.41 + .0235 X1 + .00486 X2

Predictor	*Coef*	*SE Coef*	*T*
Constant	*-1.4053*	*0.4848*	*-2.90*
X1	*0.023467*	*0.008666*	*2.71*
X2	*.00486*	*0.001077*	*4.51*

s = 0.1298 *R-sq = 93.7%*

Analysis of Variance

SOURCE	*DF*	*SS*	*MS*	*F*
Regression	*2*	*1.76209*	*.881*	*52.44*
Residual Error	*7*	*.1179*	*.0168*	
Total	*9*	*1.88000*		

b. *Using F table (2 degrees of freedom numerator and 7 degrees of freedom denominator), p-value is less than .01, reject H0, there is a significant relationship.*

c.

$$R^2 = \frac{\text{SSR}}{\text{SST}} = .937$$

d. $t_{.025} = 2.3646$ *(df = 7)*

for β_1*: p-value is between .02 and .05; reject* H_0*:* $\beta_1 = 0$*;* X_1 *is a significant variable.*

for β_2*: p-value is between .001 and .01; reject* H_0*:* $\beta_2 = 0$*;* X_2 *is a significant variable.*

Gulf University for Science & Technology
Department of Economics & Finance
ECON 380: Business Statistics
Spring 2013

Dr. Fida Karam
Room N1-115
PHONE: 2530-7351
karam.f@gust.edu.kw

Problem Set VII:
Regression Analysis

1. The regression line $\hat{y} = 3 + 2x$ has been fitted to the data points (4, 8), (2, 5), and (1, 2). The sum of the squared residuals will be:

   a. 7
   b. 15
   c. 8
   d. 22

   $SSE = \Sigma(y_i - \hat{y})^2$

   $\hat{y}$ plug x in the equation

		$y_i - \hat{y}$	$(y_i - \hat{y})^2$
$\hat{y}$	$3+2(4)=11$	-3	9
	$3+2(2)=7$	-2	4
	$3+2(1)=5$	-3	9

   $SSE = 22$

2. The residual is defined as the difference between:

   a. the actual value of y and the estimated value of y
   b. the actual value of x and the estimated value of x
   c. the actual value of y and the estimated value of x
   d. the actual value of x and the estimated value of y

3. In the simple linear regression model, the y-intercept represents the:

   a. change in y per unit change in x.
   b. change in x per unit change in y.
   c. value of y when $x = 0$.
   d. value of x when $y = 0$.

4. In a simple linear regression problem, the following statistics are calculated from a sample of 10 observations: $\sum(x-\bar{x})(y-\bar{y}) = 2250$, $s_x = 10$, $\sum x = 50$, $\sum y = 75$. The least squares estimates of the slope and y-intercept are, respectively,

   a. 1.5 and 0.5
   b. 2.5 and 1.5
   c. 1.5 and 2.5
   d. 2.5 and –5.0

   $n = 10$

   $\Sigma(x-\bar{x})(y-\bar{y}) = 2250$

   $s_x = 10$

   $\Sigma x = 50$

   $\Sigma y = 75$

   $s_x^2 = \frac{\Sigma(x-\bar{x})^2}{n-1}$

   $10^2 = \frac{\Sigma(x-\bar{x})^2}{10-1}$

   $\Sigma(x-\bar{x})^2 = 100(9) = 900$

   $b_1 = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(x-\bar{x})^2} = \frac{2250}{900} = 2.5$ (1)

   $b_0 = \bar{y} - b_1\bar{x} = \frac{75}{10} - 2.5\left(\frac{50}{10}\right) = -5$

5. In a simple linear regression problem, the following sum of squares are produced: $\sum(y_i-\bar{y})^2 = 200$, $\sum(y_i-\hat{y}_i)^2 = 50$, and $\sum(\hat{y}_i-\bar{y})^2 = 150$. The percentage of the variation in y that is explained by the variation in x is:

SST SSE SSR $= R^2$ ← يعني

a. 25%

b. 75%

c. 33%

d. 50%

$R^2 = \frac{SSR}{SST} = \frac{150}{200} = 75\%$

6. Given that the sum of squares for error is 60 and the sum of squares for regression is 140, then the coefficient of determination is:

SSE SSR

a. 0.429

b. 0.300

c. 0.700

d. None of these choices.

$R^2 = \frac{SSR}{SST} = \frac{140}{(60+140)} = 0.7$

7. A simple regression line using 25 observations produced: $\Sigma(X-\bar{X})^2 = 1$, SSR = 118.68 and SSE = 56.32. The standard error of the slope was:

a. 2.11

b. 1.56

c. ~~2.44~~

d. None of these choices.

simple reg line
slope = b_1

$n = 25$
$\Sigma(x-\bar{x}) = 1$
① SSR = 118.68
SSE = 56.32

$MSE = \frac{SSE}{n-2} = \frac{56.32}{25-2} = 2.44$

② $S_{b_1} = \sqrt{\frac{MSE}{\Sigma(x-\bar{x})^2}} = \sqrt{2.44} = 1.56$

8. If the coefficient of correlation is – 0.60, then the coefficient of determination is:

a. – 0.60

b. – 0.36

c. 0.36

d. 0.77

$r_{xy} = -0.6$

$\sqrt{R^2} = \frac{r_{xy}}{\text{Sing } b_1} = \frac{-0.6}{-1} = 0.6$

$R^2 = (0.6)^2 = 0.36$

9. In testing the hypotheses: H_0: $\beta_1 = 0$ vs. H_1: $\beta_1 \neq 0$, the following statistics are available: n = 10, b_0 = - 1.8, b_1 = 2.45, Se (b_1) = 1.20 and $\hat{y}$ = 6. The value of the test statistic is:

a. 2.042

b. 0.306

c. –1.50

d. –0.300

$t = \frac{b_1}{S_{b_1}} = \frac{2.45}{1.2} = 2.042$

$sb_1 = 20$

10. In a simple regression model, if the standard error of estimate $b_1 = 20$, $\Sigma(X - \overline{X})^2 = 4$, and $n = 4$, then the sum of squares for error (SSE) =

$Sb_1 = 20$

$\Sigma(X - \bar{x})^2 = 4$

$n = 4$

$SSE = ?$

$Sb_1 = \sqrt{\frac{MSE}{\Sigma(X - \bar{x})^2}}$

$(sb_1)^2 = \frac{MSE}{\Sigma(X - \bar{x})^2}$

$20^2 = \frac{MSE}{4}$

$MSE = 4(20^2) = 1600 \Rightarrow SSE = 1600$

$\frac{SSE}{n-2}$ $(4-2)$

$= 3200$

a. 400
b. 3,200
c. 4,000
d. 40,000

11. In a multiple regression analysis involving 6 independent variables, the total variation in y is 900 and SSR = 600. What is the value of SSE?

a. 300
b. 1.50
c. 0.67
d. None of these choices.

$P = 6$

$SST = 900$

$SSR = 600$

$SSE = 900 - 600 = 300$ (a)

12. In testing the validity of a multiple regression model in which there are four independent variables, the null hypothesis is:

a. H_0: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 1$
b. H_0: $\beta_0 = \beta_1 = \beta_2 = \beta_3 = \beta_4$
c. H_0: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$
d. H_0: $\beta_1 = \beta_2 = \beta_3 = \beta_4 \neq 0$

Exhibit 1

A multiple regression model involving 4 independent variables and a sample of 15 observations resulted in the following sum of squares: SSR = 165, SSE = 60

13. Refer to Exhibit -1. The coefficient of determination is

$P = 4$ $n = 15$ $SSR = 165$ $SSE = 60$

$R^2 = \frac{SSR}{SST} = \frac{165}{165 + 60} = 0.7335$ (b)

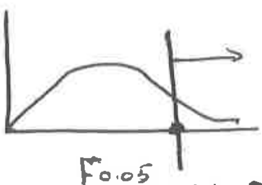
a. 0.3636
b. 0.7333
c. 0.275
d. 0.5

14. Refer to Exhibit - 1. If we want to test for the significance of the model at 95% confidence, the critical F value (from the table) is $\alpha = 0.05$

$df_N = p = 4$

$df_D = n - p - 1 = 15 - 4 - 1 = 10$

a. 3.06
b. 3.48
c. 3.34
d. 3.11



$F_{0.05}(4,10) = 3.48$

15. Refer to Exhibit - 1. The F-test statistic from the information provided is

a. 2.110

b. 3.480

c. 4.710

d. 6.875

$$F = \frac{MSR}{MSE} = \frac{41.25}{6} = 6.875 \quad (d)$$

$$MSR = \frac{SSR}{P} = \frac{165}{4} = 41.25$$

$$MSE = \frac{SSE}{n-P-1} = \frac{60}{10} = 6$$

Exhibit 2

The following information regarding a dependent variable Y and an independent variable X is provided

Simple reg, because one x

$\frac{\Sigma(X-\bar{X})(Y-\bar{Y})}{n-1} = -52$, $\Sigma X = 90$, $\Sigma Y = 340$, $\Sigma(X-\bar{X})^2 = 234$, $\Sigma(Y-\bar{Y})^2 = 1974$, $n = 4$, SSR = 104

23. Refer to Exhibit 2. The total sum of squares (SST) is

a. -156

b. 234

c. 1870

d. 1974

24. Refer to Exhibit 2. The sum of squares due to error (SSE) is

a. -156

b. 234

c. 1870

d. 1974

25. Refer to Exhibit 2. The slope of the regression equation is

a. -0.667

b. 0.667

c. 0.22

d. - 0.22

26. Refer to Exhibit 2. The intercept of the regression equation is b_0

a. 65

b. 100

c. 40

d. - 40

27. Refer to Exhibit 2. The coefficient of correlation of the regression equation is r_{xy}

a. - 0.05

b. 0.05

c. 0.23

d. - 0.23

s_{b_1}

28. Refer to Exhibit 2. The standard error of the slope (b_1) is

a. 4

b. 187

c. 2

d. 935

CH1

Descriptive statistics:

def: population: the set of all elemeants of interest

population parameters

Students	score	$X_i - \mu$	$(x' - \mu)^2$
1	5	-2	4
2	8	1	1
3	6	-1	1
4	10	3	9
5	4	-3	9
6	7	0	0
7	9	2	4
8	8	1	1
9	3	-4	16
10	10	3	9
10	70	0 (always zero)	54

① population mean: $\mu = \frac{\Sigma x_i}{N}$ $\quad N$ = population size

$$\mu = \frac{70}{10} = 7$$

② Median 3 4 5 6 $\underbrace{7\ 8}$ 8 9 10 10 arrange from smalles to biggest select number in the mid

$$\frac{7+8}{2} = 7.5$$

③ - population varance $\sigma^2 = \frac{\Sigma (X_i - \mu)^2}{N}$

$$\sigma^2 = \frac{54}{10} = 5.4$$

4. Standard devation

$$\sigma = \sqrt{5.4} = 2.32$$

(5) Population proportion π what is the proportion of students who scored below 7

$\frac{4}{10} = 0.4 \times 100 = 40\%$ $\pi \rightarrow$ always between zero and

Inferential Statistics (we try to guess)

Sample statiscis. "part of the population"

n= 5

	x_i	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
1	5	-1.6	2.56
2	8	1.4	1.96
3	6	-0.6	0.36
4	10	3.4	11.56
5	4	-2.6	6.76

(1) Sample mean $\bar{X} = \frac{\Sigma X}{n} = \frac{33}{5} =$

(2) Median: 4 5 (6) 8 10 : 6

(3) Sample variance $= S^2$

$$= \frac{\Sigma (X_i - \bar{X})^2}{n-1}$$

$$\frac{23.2}{4} = 5.8$$

sample Standard deveation S

$$\sqrt{S^2} = \sqrt{5.8} = 2.41$$

Sample proportion: P, how many students scored ~~between~~ below 7

$0 \leq P \leq 1$ $\frac{3}{5} = 60\%$

always.

$\bar{X}$ is the point estemation

6.6 is " " estemate μ

S^2 is the point estemater σ^2

5.8 is " " estemat σ^2

P Point estemater π

Probelm Set I

Problem 2:

$\bar{X}$ mean: $\frac{10+20+21+17+16+12}{6} = \frac{96}{6} = 16$

Median: $10 * 12 * \underbrace{16 * 17} * 20 * 21$

$\frac{16+17}{2} = 16.5$

$S^2 = 18.8$ $\quad S = \sqrt{18.8} = 4.335$

$n = 6$

X	$(X-\bar{X})$	$(X-\bar{X})^2$
10	-6	36
20	4	16
21	5	25
17	1	1
16	0	0
12	-4	16

$S^2 = \frac{\Sigma (X_i - \bar{X})^2}{n-1} = \frac{94}{6-1} = 18.8$

$\sqrt{18.8} = 4.3358$

$\bar{X} = \frac{96}{6} = 16$

5 A $\bar{X} = \frac{465}{5} = 93$

B

X	$(X_i - \bar{X})$	$(X_i - \bar{X})^2$
94	1	1
100	7	49
85	-8	64
94	1	1
92	-1	1
		116

$S^2 = \frac{116}{5-1} = 29$

$S \sqrt{29} = 5.39$

(6) (A) $\frac{595}{1008} = 0.59$ (B) $\frac{332}{1008} = 0.33$ (C) $\frac{81}{1008} = 0.08$

~~Rando~~

Random variable: Out come of an experiment

Mean $E(X)$

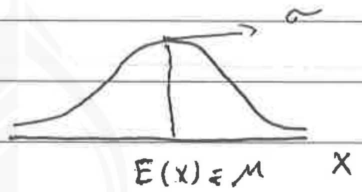
standard deviation σ

probability distribuation

The probability distribuation of random variable describes how probility are distrbuted over the values of random variable

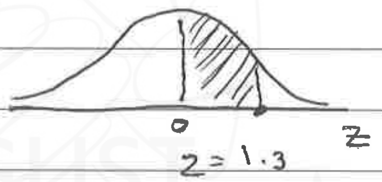
Normal distribuation:

- Symetric
- Bell shaped
- Median = mean



standard normal mean = 0 $\sigma = 1$

$Z \sim N(\overset{\mu}{0}, \overset{\sigma}{1})$



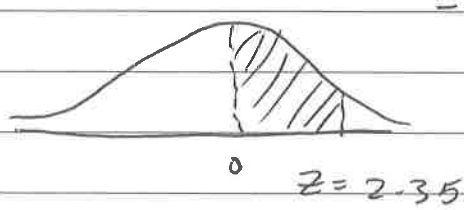
$(0 \leq Z \leq 1.3) = 0.4032 \rightarrow$ table

probleity that Z is between 0 and 1.3

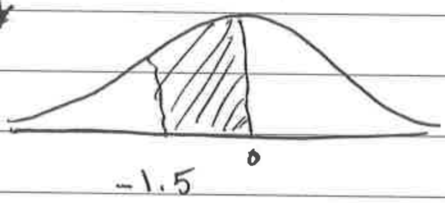
$P(-1.3 \leq Z \leq 0) = P(0 \leq Z \leq 1.3) = 0.4$

$P(0 \leq Z \leq 2.35) = 0.4906$

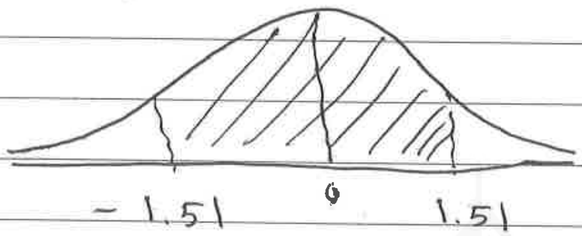
$= 49.06\%$



$P(-1.51 \leq Z \leq 0) = P(0 \leq Z \leq 1.51) = 0.4345$
$= 43.45\,\%$

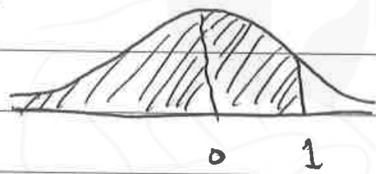


$P(-1.51 \leq Z \leq 1.51) = P(-1.51 \leq Z \leq 0) +$
$(0 \leq Z \leq 1.51)$

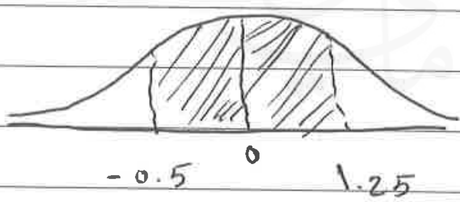


$= 0.4345 + 0.4345 = 86.9\,\%$

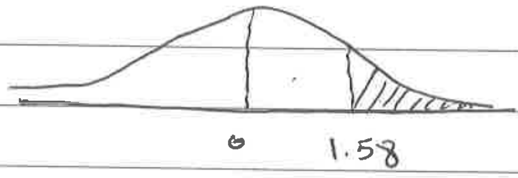
$P(Z \leq 1) = 0.5 + P(0 \leq Z \leq 1) = 0.5 + 0.3413$
$= 0.8413$



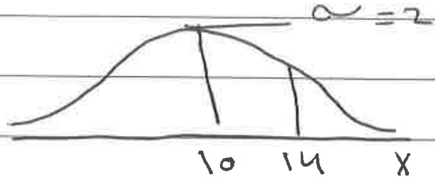
$P(-0.5 \leq Z \leq 1.25) = P(-0.5 \leq Z \leq 0)$
$+ P(0 \leq Z \leq 1.25)$
$= 0.1915 + 0.3944 = 0.5859$



$P(Z \geq 1.5) = 0.5 - P(0 \leq Z \leq 1.50)$
$= 0.5 - 0.4429 = 0.0571$



$X \sim N(\mu = 10, \sigma = 2)$



$P(10 \leq X \leq 14)$

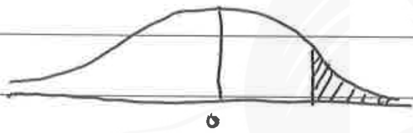
$$Z = \frac{X - \mu}{\sigma}$$

How to transform X to Z

$$P(10 \leq X \leq 14) = P\left(\frac{10-10}{2} \leq \underbrace{\frac{X-\mu}{\sigma}}_{Z} \leq \frac{14-10}{2}\right)$$

$$= (0 \leq Z \leq 2) = 0.4772$$

$$P(X \geq 14) = P\left(\frac{X-\mu}{\sigma} \geq \frac{14-10}{2}\right) = P\ P(Z \geq 2)$$



$0.5 - (0 \leq Z \leq 2)$

$0.5 - .4772 = 0.0228$

Example:

$P(0 \leq Z \leq Z_1) = 0.4394$

$Z = 1.55$

$P(0 \leq Z \leq Z_1) = 0.2055 \simeq 0.2054$

$Z = 0.54$

$P(0 \leq Z \leq Z_1) = 0.44$

not in the table

between 0.4394 0.44 0.4406

1.55 1.56

take the average. $\frac{1.55 + 1.56}{2} = 1.555$

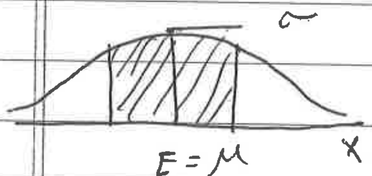
X random variable

- $E(X) = \mu$
- St. Devation σ
- prob Distribuation

Normal distribuation graph

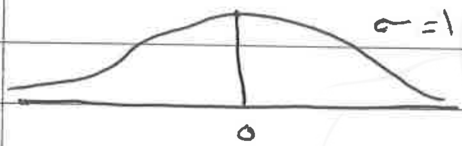
$\mu = 2$

$\sigma = 0.5$



$P(0 \leq X \leq 12)$

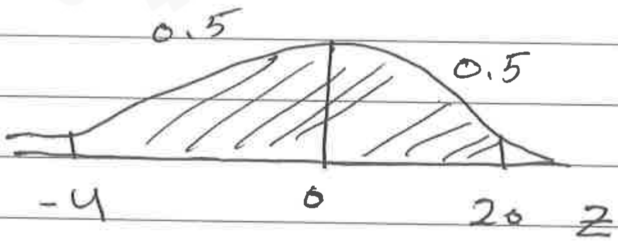
St. normal distribuation graph



$$Z = \frac{X - \mu}{\sigma}$$

$$P(0 \leq X \leq 12) = P\left(\frac{0-2}{0.5} \leq \frac{X-\mu}{\sigma} \leq \frac{12-2}{0.5}\right)$$

$$= (-4 \leq Z \leq 20) = 1$$



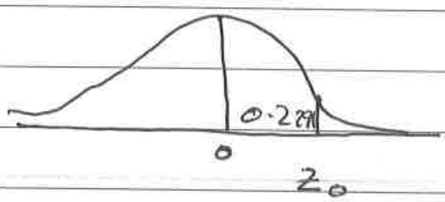
problem set 1 , problem 9

a) $P(0 \leq Z \leq Z_0) = 0.475$

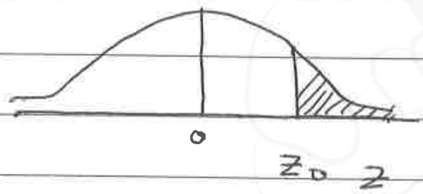
take it directly from the table
if found between two numbers take
the average.

b) $P(0 \leq Z \leq Z_0) = 0.2291$

$Z_0 = 0.61$



c) $P(Z \geq Z_0) = 0.1314$



$P(0 \leq Z \leq Z_0) = 0.5 - 0.1314$

$= 0.3686$

Now go to the
table $= 1.12$
positive only.

d) $P(Z \leq Z_0) = 0.67$

$P(0 \leq Z \leq Z_0) = 0.67 - 0.5 = 0.17$

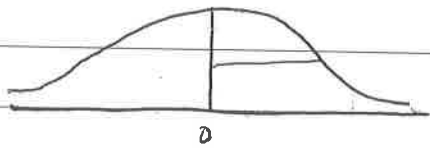
$Z_0 = 0.44$

problem 10).

A) $X \sim N \rightarrow$ follows normal distribuation

$$X \sim N(\overset{\mu}{28000}, \overset{\sigma}{4000})$$

a) $P(X \geq 28000)\ P\left(\frac{X-\mu}{\sigma} \geq \frac{28000-28000}{4000}\right)$



$P(Z \geq 0) = 0.5$

b) $(X \leq 14000) = P\left(\frac{X-\mu}{\sigma} \leq \frac{14000-28000}{4000}\right)$

$= P(Z \leq -3.5) = 0.5 - 0.5 = 0$

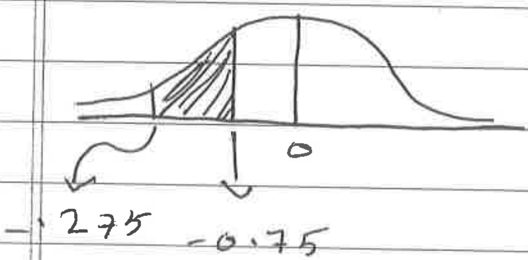
c) $P(17000 \leq X \leq 25000)$

$$P\left(\frac{17000-28000}{4000} \leq \frac{X-\mu}{\sigma} \leq \frac{25000-28000}{4000}\right)$$

$= P(-2.75 \leq Z \leq -0.75).$

$= P(-2.75 \leq Z \leq 0) - P(0.75 \leq Z \leq 0)$

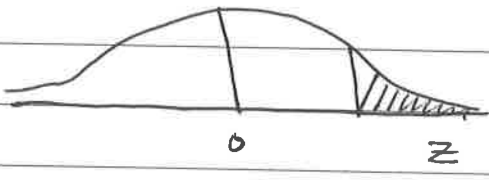
$= 0.2236$



D) X_1 $P(X \geq X_1) = 5\% = 0.5$ Test

$$P\left(\frac{X-\mu}{\sigma} \geq \frac{X_1 - \overset{28000}{\mu}}{\sigma_{4000}}\right) = 0.5$$

$$P(Z \geq Z_1) = 0.05$$



$P(0 \leq Z \leq Z_1) = 0.5 - 0.05$

from the table

0.4495	0.45	0.4505
1.64		1.65

$$Z = \frac{1.64 + 1.65}{2} = 1.645$$

$$Z_1 = \frac{X_1 - 28000}{4000} = 1.645$$

$$X_1 - 28000 = 1.645\,(4000)$$

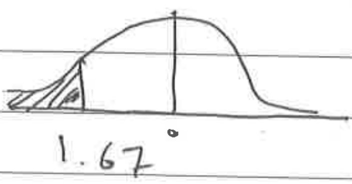
$$X_1 = 1.645\,(4000) + 28000 = 34580 \text{ people}$$

11.) $X \sim N(\mu = 45 \quad \sigma = 6)$ A $\rightarrow$ B

a) $P(X \leq 35) = P\left(\frac{X-\mu}{\sigma} \leq \frac{35-45}{6}\right) = P(Z \leq -1.67)$

$$0.5 - P(-1.67 \leq Z \leq 0)$$

$$0.5 - 0.4525 = 0.0475$$

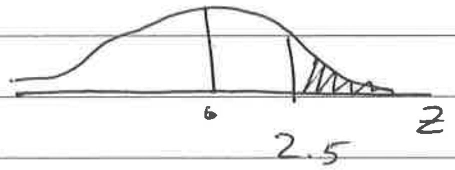


b) $P(X \geq 60) = P\left(\frac{x-\mu}{\sigma} \geq \frac{60-45}{\sigma}\right)$

$P(Z \geq 2.5)$

$0.5 - P(0 \leq Z \leq 2.55)$

$= 0.5 - 0.4938 = 0.0062$



c) $P(X \leq X_1) = 0.7$

$P\left(\frac{X-\mu}{\sigma} \leq \frac{X_1-45}{6}\right) = 0.7$

$P(Z \leq Z_1) = 0.7$

$P(0 \leq Z \leq Z_1) = 0.7 - 0.5 = 0.2$

close to 0.1985

$Z_1 = 0.52$

$Z = \frac{X_1-45}{6} = 0.52$

$X_1 - 45 = 6(0.52) = 6(0.52) + 45 = 48.12$

Chapter 7

Sampling distirbuation of $\bar{X}$

$\bar{X}$ random

mean $E(\bar{X}) = \mu$

St. Devation $\sigma_{\bar{X}}$

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

Pb distribuation

$\bar{X}$ follows normal distribuation

1- if population is normal

2- $n \geq 30$ Centrel limit theorm.

$\bar{X}$ random variable

- $E(\bar{X}) = \mu$ $\bar{X}$ non biase estimate of μ

St Devation of $\bar{X}$ S.t error of the mean

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

According to the central limit therom we can approximate the pb distribuation

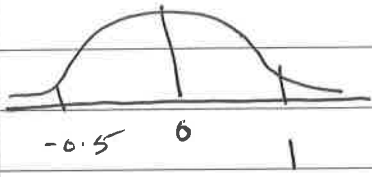
$\mu = 50$ $\sigma_{\bar{X}} = 20$

$P(40 \leq \bar{X} \leq 70)$ $Z = \dfrac{\bar{X} - \mu}{\sigma_{\bar{X}}}$

$$P\left(\frac{40-50}{20} \leq \frac{\bar{X}-\mu}{\sigma_{\bar{X}}} \leq \frac{70-50}{20}\right)$$

$P(-0.5 \leq Z \leq 1)$

$$= P(-0.5 \leq Z \leq 0) + P(0 \leq Z \leq 1)$$



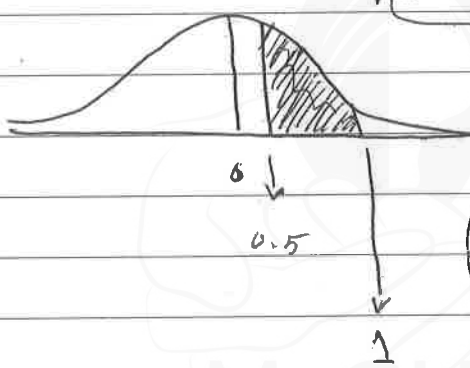
$$= 0.1915 + 0.3413 = 0.5328$$

$$P(60 \leq \bar{X} \leq 70)$$

$$= P\left(\frac{60-50}{20} \leq \frac{\bar{X} - \mu}{\sigma_{\bar{X}}} \leq \frac{70-50}{20}\right)$$

$$= P(0.5 \leq Z \leq 1)$$

$$= 0.3413 - 0.1915 = 0.1498$$



find 0.5 and subtract it from 1

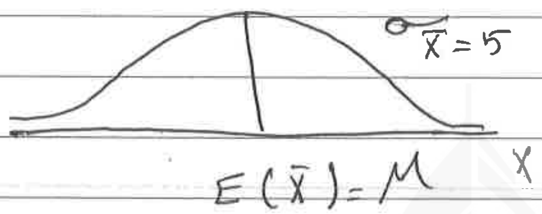
Book problems:

13. $\mu = 200$ $\sigma = 50$ $n = 100$

a) $E(\bar{X})$ $\mu = 200$

b) $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{50}{\sqrt{100}} = 5$

c) $n \geqslant 30$ normal $\rightarrow$ bigger than 30 means normal distribuattion



$E(\bar{X}) = \mu$
$= 200$

D) The sampling distribuation of $\bar{x}$ shows how probilityes are distrbuted over $\bar{X}$.

14. a) $P(-5 \leq \bar{x} - \mu \leq 5) = P\left(\frac{-5}{5} \leq \frac{\bar{X} - \mu}{\sigma_{\bar{X}}} \leq \frac{5}{5}\right)$

$= P(-1 \leq Z \leq 1) = 2P(0 \leq Z \leq 1)$

$Z = \frac{\bar{X} - \mu}{\sigma_{\bar{x}}}$

$= 2(0.3413) = 0.6826$

b) $(-10 \leq \bar{X} - \mu \leq 10) = P\left(\frac{-10}{5} \leq \frac{\bar{X} - \mu}{\sigma_{\bar{x}}} \leq \frac{10}{5}\right)$

$= P(-2 \leq Z \leq 2) = 2P(0 \leq Z \leq 2)$

$= 2(0.4772) = 0.9544.$

19. $\mu = 106.4$ $\quad \sigma = 4.5$ $\quad n = 30$ normal

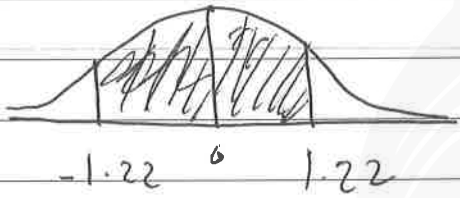
a) $P(-1 \leq \bar{X} - \mu \leq +1) = P\left(\frac{-1}{0.822} \leq \frac{\bar{X} - \mu}{\sigma_{\bar{X}}} \leq \frac{+1}{0.822}\right)$

$Z = \dfrac{\bar{X} - \mu}{\sigma_{\bar{X}}}$ $\quad = P(-1.22 \leq Z \leq 1.22)$

$\sigma_{\bar{X}} = \dfrac{\sigma}{\sqrt{n}} = \dfrac{4.5}{\sqrt{30}} = 0.822$

$= 2P(0 \leq Z \leq 1.22)$

$= 2(0.3888) = 0.7776$



b) $n = 50$ $\quad \sigma_{\bar{X}} = \dfrac{4.5}{\sqrt{50}} = 0.636$

$P(-1 \leq \bar{X} - \mu \leq 1) = P\left(\frac{-1}{0.636} \leq \frac{\bar{X} - \mu}{\sigma_{\bar{X}}} \leq \frac{1}{0.636}\right)$

$= P(-1.57 \leq Z \leq 1.57)$

$2P(0 \leq Z \leq 1.57) = 2(0.4418) = 0.8836$

c) $P(-1 \leq \bar{X} - \mu \leq 1) = P\left(\frac{-1}{0.45} \leq Z \leq \frac{1}{0.45}\right)$

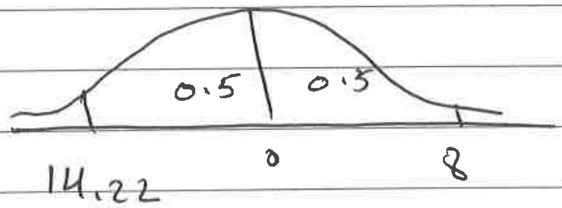
$= P(-2.22 \leq Z \leq 2.22) = 2P(0 \leq Z \leq 2.22)$

$= 2(0.4868) = 0.9736.$

Extra

what is the probibilty that the mean price for a sample of 100 Stations is between 100 and 110

d) $p(100 \leq \bar{x} \leq 110) =$



$$p\left(\frac{100-106.4}{0.45} \leq \frac{\bar{x}-\mu}{\sigma_{\bar{x}}} \leq \frac{110-106.4}{0.45}\right)$$

$$= p(-14.22 \leq z \leq 8) = 1$$

21. A) $\sigma_{\bar{x}} = 20$ $\quad\quad \sigma = 500$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

$$20 = \frac{500}{\sqrt{n}}$$

$$\sqrt{n} = \frac{500}{20} = 25$$

$$(\sqrt{n})^2 = 25^2$$

$$n = 625$$

B) $p(-25 \leq \bar{X}-\mu \leq +25)$

$$= p\left(\frac{-25}{20} \leq \frac{\bar{X}-\mu}{\sigma_{\bar{X}}} \leq \frac{25}{20}\right)$$

$$= p(-1.25 \leq z \leq 1.25)$$

$$2p(0 \leq z \leq 1.25)$$

$$= 2(0.3944)$$

$$= 0.7888$$

Chapter 8

PT estimation Estimate

$\mu = \bar{x} \pm M.E \longrightarrow$ margin of error

$n = 100 \quad \bar{x} \quad \sigma = 12$

$\bar{x}$ random variable

$E(\bar{x}) = \frac{\sigma}{\sqrt{n}} = \frac{12}{\sqrt{100}} = 1.2$

Confidence Internal for μ

$$\mu = \bar{x} \pm \underbrace{Z_{\frac{\alpha}{2}} \sigma_{\bar{x}}}_{\text{margin of error}} \longrightarrow \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

confidence level 95% } → same thing.
confidence coefficient 0.95

find 95% CI
for μ when

$\bar{x} = 72 \quad \sigma = 12 \quad n = 100$

$\mu = \bar{x} \pm Z_{\frac{\alpha}{2}} \sigma_{\bar{x}}$

$= 72 \pm 1.96 \frac{12}{\sqrt{100}}$

$= 72 \pm 2.352$

$72 - 2.352$ or $72 + 2.352$

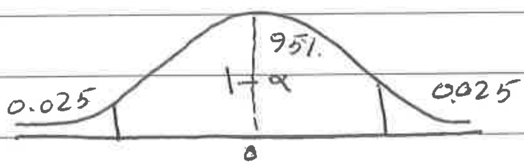
$69.65 \qquad 74.35$

$\mu \in [69.65 ; 74.35]$

$1 - \alpha = 0.95$

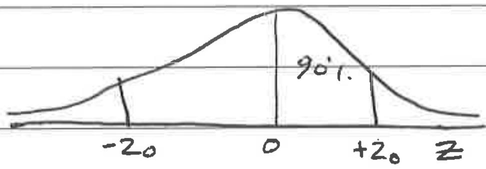
$\alpha = 1 - 0.95 = 0.05$

$\frac{\alpha}{2} = \frac{0.05}{2} = 0.025$



Find a 90% CI for μ

Find a 90% CI for μ



$P(0 \le z \le z_0) = \frac{0.90}{2} = 0.45$

$0.4495 < 0.45 < 0.4505$

1.64 1.65

$z_0 = \frac{1.64 + 1.65}{2} = 1.64$

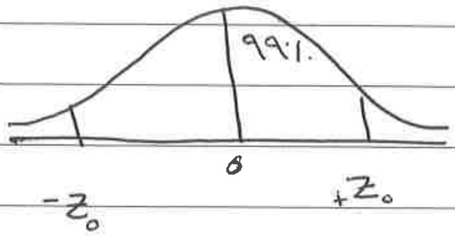
$\mu = 72 \pm 1.645 \frac{12}{\sqrt{100}}$

$= 72 \pm 1.97 \longrightarrow M.E$

or 70.03 to 73.97

$\mu \in [70.03 ; 73.97]$

Find a 99% CI for μ



$$p(0 \leq z \leq z_0) = \frac{0.99}{2} = 0.495$$

$$\underset{2.57}{0.4949} < 0.495 < \underset{2.58}{0.4951}$$

$$z_0 = \frac{2.57 + 2.58}{2} = 2.575$$

$$\mu = 72 \pm 2.575 \frac{12}{\sqrt{100}} = 72 \pm 3.09$$

68.93 to 75.09

$$\mu \in [68.93 ; 75.09]$$

① $n = 40$ $\bar{X} = 25$ $\sigma = 5$ Book problem ① Chapter ⑧

a) $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{5}{\sqrt{40}} = 0.79$

b) $\mu = \bar{X} \pm \underbrace{Z_{\frac{\alpha}{2}} \sigma_{\bar{X}}}_{\to M.E}$

$95\% \to Z_{\frac{\alpha}{2}} = 1.96$

$95\% \longrightarrow Z_{\frac{\alpha}{2}} = 1.96$

$M.E = Z_{\frac{\alpha}{2}} \sigma_{\bar{X}}$

$= 1.96\,(0.79)$
$= 1.55$

③ $n = 60$ $\bar{X} = 80$ $\sigma = 15$

$\mu = \bar{X} \pm Z_{\frac{\alpha}{2}} \sigma_{\bar{X}}$

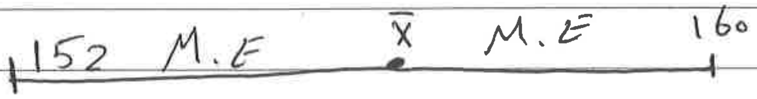
a) $95\% \to Z_{\frac{\alpha}{2}} = 1.96$

$\mu = 80 \pm 3.8$
or 76.2 to 83.8
$\mu \in [76.2 ; 83.8]$

b) $n = 120$
$\mu = 80 \pm 1.96 \frac{15}{\sqrt{120}}$
$= 80 \pm 2.68$
~~$\mu \in [\,82.68 ;$~~
$\mu \in [77.32 ; 82.68]$

(4) $M \in [152 ; 160]$ 95%

$\sigma = 15$ $n?$



$M.E = \dfrac{160 - 152}{2} = 4$

$M.E = Z_{\frac{\alpha}{2}} \cancel{\sigma_{\bar{x}}}$ $\sigma_{\bar{x}} = Z_{\frac{\alpha}{2}} \dfrac{\sigma}{\sqrt{n}}$

$4 = 1.96 \dfrac{15}{\sqrt{n}}$

$4\sqrt{n} = 1.96(15)$

$\sqrt{n} = \dfrac{1.96(15)}{4} = 7.35$

$n = (7.35)^2 = 54$

(6) 95%

$\bar{x} = 11500$ $n = 60$ $\sigma = 4000$

$M.E = Z_{\frac{\alpha}{2}} \dfrac{\sigma}{\sqrt{n}}$

$= 1.96 \dfrac{4000}{\sqrt{60}} = 1012$

if σ Known

CI for μ
Confidence level $(1-\alpha)\%$
Confidence cofficent $1-\alpha$

95% $Z_{\frac{\alpha}{2}} = 1.96$

99% $Z_{\frac{\alpha}{2}} = 2.575$ ~~ljel~~

90% $Z_{\frac{\alpha}{2}} = 1.645$

$\mu = \bar{X} \pm M.E$

$\mu = \bar{X} \pm Z_{\frac{\alpha}{2}} \sigma_{\bar{X}}$

$= \bar{X} \pm Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$ Z distribuation

σ unKnown

Degrees of freedom are the number of intebendt observations that you need to calclate statistices.

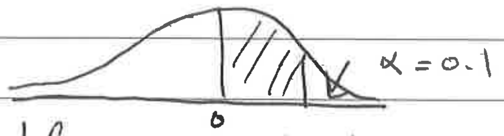
$\mu = \bar{X} \pm t_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}}$ $\frac{S}{\sqrt{n}} = S_{\bar{X}}$

estimated st. error of the mean

degrees of freedom
↓
$df = n-1$

Example $df = 10$ $\alpha = 0.1$

$t_{0.1, 10} = 1.3722 \rightarrow$ table



$df = 10$ $t = 13722$

$t_{0.05, 27} = 1.7033$

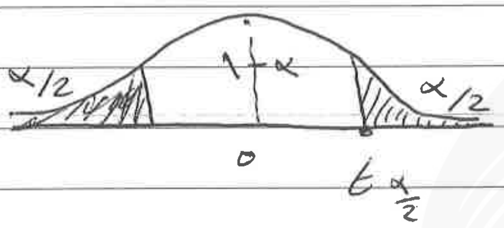
$t_{0.025, 18} = 2.1009$

$$S = \sqrt{S^2}$$

$$S^2 = \frac{\Sigma (X_i - \bar{X})^2}{n-1}$$

A reporter for student news paper is writing an articale on the cost of, off campus housing a sample of 16 apartments within 1 K.m of campus resullted in a sample mean of 650 Euros per month and a sample standard devation of 55 Euro provide 95% confidence intervel estimate of the mean rent per month for population of appartments within 1Km of campus.

$n = 16 \quad \bar{X} = 650 \quad S = 55 \quad df = 16 - 1 = 15$



$1 - \alpha = 95\%$

$\alpha = 1 - 0.95 = 0.05$

$\alpha/2 = \frac{0.05}{2} = 0.025$

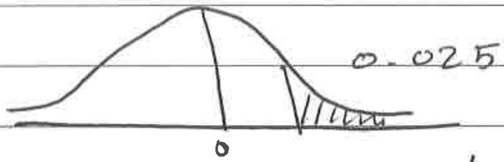
$t_{0.025, 15} = 2.131$

$\mu = 650 \pm 2.131 \frac{55}{\sqrt{16}}$

$= 650 \pm 29.3$

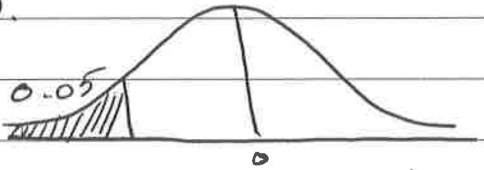
$\mu \in [620.70 ; 679.30]$

9). A)



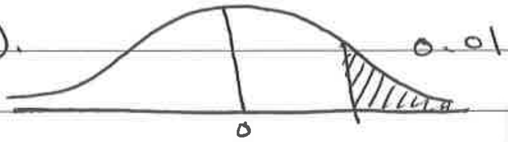
$t_{0.025, 12} = 2.1788$

B).



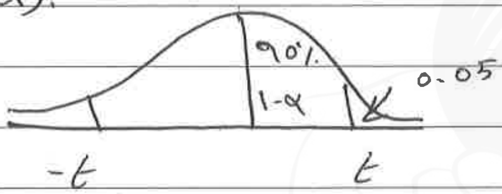
$t_{0.05, 50} = -1.6449$

C).



$t_{0.01, 30} = 2.4573$

d).

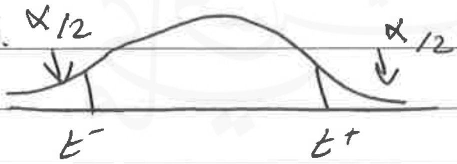


$1 - \alpha = 90\% \Rightarrow \alpha = 1 - 0.9 = 0.1$

$\frac{\alpha}{2} = 0.1/2 = 0.05$

$t_{0.05, 25} = \pm 1.7081$

e).



$1 - \alpha = 0.95$

$\alpha = 0.05$

$\frac{\alpha}{2} = 0.025$

$t_{0.025, 45} = \pm 1.96$

11.) $n = 54$ $\bar{x} = 22.5$ $S = 4.4$

a) 90.1. $\mu = \bar{x} \pm t_{\frac{\alpha}{2}} \frac{S}{\sqrt{n}}$

$df = 54 - 1 = 53$

$1 - \alpha = 0.90$; $\alpha = 0.1$; $\alpha/2 = 0.05$

$t = 0.05, 53 = 1.6449$

$\mu = 22.5 \pm 1.6449 \frac{4.4}{\sqrt{54}}$

$= 22.5 \pm 0.984$

$\mu \in [21.51, 23,48]$

b) $1 - \alpha = 0.95$; $\alpha = 0.05$; $\alpha/2 = 0.025$

$t_{0.025}, 33 = 1.96$

$\mu = 22.5 \pm 1.96 \frac{4.4}{\sqrt{54}}$

$= 22.5 \pm 1.17$

$\mu \in [21.32 ; 23.67]$

c). $99\% = 1-\alpha \Rightarrow \alpha = 0.01 \Rightarrow \alpha/2 = 0.005$

$t_{0.005, 53} = 2.5758$

$\mu = 22.5 \pm 2.5758 \dfrac{4.4}{\sqrt{54}}$

$= 22.5 \pm 1.54$

$\mu \in [20.95 ; 24.04]$

15). $n = 40$ $\bar{X} = 40{,}000$ $S = 15300$

A) $M.E = t_{\frac{\alpha}{2}} \dfrac{S}{\sqrt{n}}$

$95\% = 1-\alpha \Rightarrow \alpha = 0.05 \Rightarrow \alpha/2 = 0.025$

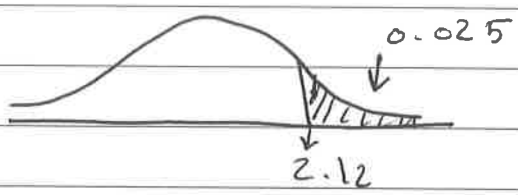
$df = 40 - 1 = 39$

$t_{0.025, 39} = 1.96 \dfrac{15{,}300}{\sqrt{40}} = 2421.10$

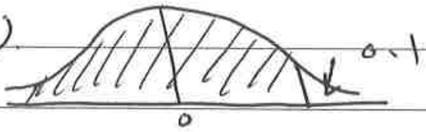
b) $\mu = 40{,}000 \pm 2{,}421.16$

$\mu \in [37578.9 ; 42421.1$

8) A) 0.025

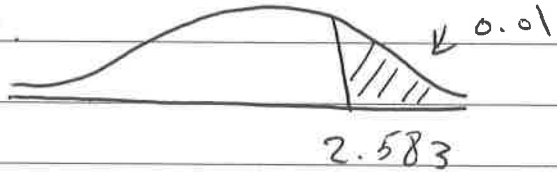


B)



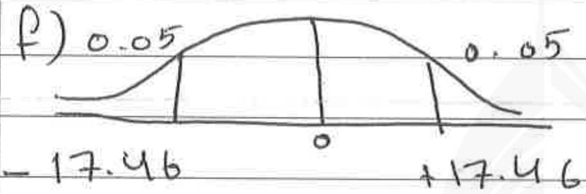
$1-0.1=90$

c)



$1-(0.025+0.025)$
$=95\%$

f)

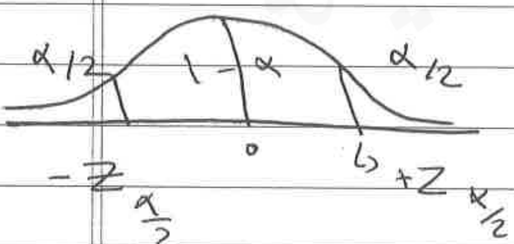


$1-(0.05+0.05)$
$=0.90$

C.I for μ

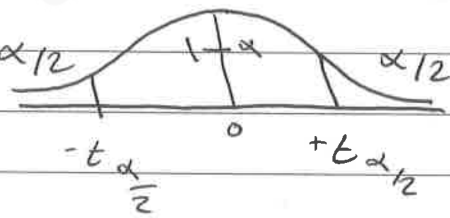
σ known

$\mu = \bar{X} \pm Z_{\alpha/2}\, \sigma_{\bar{x}}$



σ unknown

$\mu = \bar{X} \pm t_{\frac{\alpha}{2}}\, S_{\bar{x}}$



$95\% = Z_{\frac{\alpha}{2}} = 1.96$

$90\% = Z_{\frac{\alpha}{2}} = 1.645$

$99\% = Z_{\frac{\alpha}{2}} = 2.575$

$M.E = 95\%$ $\quad n = ?$ Beeecause n is unknown use Z

$$M.E = Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

$$\sqrt{n} = \frac{Z_{\frac{\alpha}{2}} \sigma^*}{M.E}$$ $\longrightarrow$ planing value of σ

$M.E = 2$ $\quad 95\%$ $\quad n?$ $\quad \sigma = 9.65$

$$n = \frac{Z^2_{\frac{\alpha}{2}} \sigma^2}{M.E^2} = \frac{1.96^2 \times 9.65^2}{2^2}$$

$= 89.43$ round up, (always round up!)

$\simeq 90$ $\quad$ ~~$\sigma^* \rightarrow$ plani~~

if σ^* not avaible

$$\sigma^* = \frac{\text{range}}{4}$$

range = Biggest value - Smallest value.

8.3

19.) range $= 36$

A) $\sigma^* = \frac{36}{4} = 9$

b) M.E 3 95%

round up

$$n = \left(\frac{Z_{\frac{\alpha}{2}} \sigma^*}{M.E} \right)^2 = \left(\frac{1.96 \times 9}{3} \right)^2 = \cancel{34.57} = 35$$

c). $$n = \left(\frac{1.96 (9)}{2} \right)^2 = \cancel{77.79} = 78$$

22.) range $= 100 - 1 = 99$

$\sigma^* = \frac{99}{4} = 24.75$

$$n = \left(\frac{Z_{\frac{\alpha}{2}} \sigma^*}{M.E} \right) = \left(\frac{1.96 (24.75)}{3} \right)^2$$

$= \cancel{261.4} = 262$ round up always.

CI fo π

$\pi = P \pm z_{\frac{\alpha}{2}}$ $Sp \Rightarrow$ ~~s~~ estimated st. devation of the population

$Sp = \sqrt{\frac{P(1-P}{n}}$

$\pi = P \pm z_{\frac{\alpha}{2}} \sqrt{\frac{P(1-P)}{n}}$

$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}}$

25).

A) $P = \frac{100}{400} = 0.25$

b) $sp \sqrt{\frac{P(1-P)}{n}} = \sqrt{\frac{0.25(0.75)}{400}}$

$= 0.0217$

c) $\pi = P \pm z_{\frac{\alpha}{2}} SP$

$= 0.25 \pm 1.96(0.0217)$

$= 0.25 \pm 0.0424$

$\pi \in [0.2076 ; 0.2924]$

$0 \leq \pi \leq 1$ always.

M.E $\quad n = ?$

$$M.E = Z_{\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}}$$

$$\sqrt{n} = \frac{Z_{\frac{\alpha}{2}} \sqrt{p(1-p}}{M.E}$$

$$n = \frac{Z^2_{\frac{\alpha}{2}} \times p^*(1-p^*)}{M.E^2}$$

planning value for p^* or $p^* = 0.5$

28) $p^* = 0.5 \qquad M.E = 0.03$

$$n = \frac{Z^2_{\frac{\alpha}{2}} \, p^*(1-p^*)}{M.E^2}$$

$$= \frac{1.96^2 (0.5)(1-0.5)}{0.03^2} = 106\cancel{7}.11$$

$= 1068$

29) a) $n = 611$

$$p = \frac{281}{611} = 0.46$$

b) $$M.E = Z_{\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}}$$

$$= 1.64 \sqrt{\frac{0.46(0.54)}{611}}$$

$$= 0.033$$

c) $\pi = 0.46 \pm 0.033$

$$\pi \in [0.4268 ; 0.493]$$

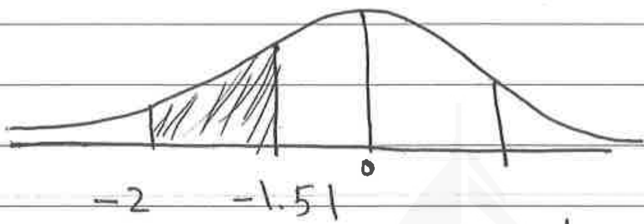
problem Set 2

~~7) p (-Z~~

7). $P(-2 \leq Z \leq -1.51) =$

$P(-2 \leq Z \leq 0) - P(-1.51 \leq Z \leq 0)$

$= 0.4772 - 0.4345 = 0.0427$



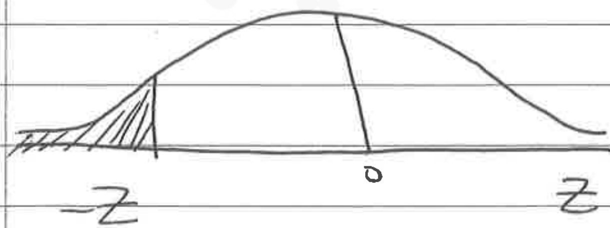
get -2 and -1.51 and subtract them.

9.) $X \sim N(\overset{\mu}{19}, \overset{\sigma}{7})$

$P(X < 5) = P\left(\frac{X-\mu}{\sigma} < \frac{5-9}{7}\right)$

$P(Z < -2)$

$0.5 - 0.4772$
$= 0.0228$



Chapter 9: Hypothesis testing

$H_0: \mu \geq \mu_0$	$H_0: \mu \leq \mu_0$	$H_0: \mu = \mu_0$
$H_1: \mu < \mu_0$	$H_1 \mu > \mu_0$	$H_1: \mu \neq \mu_0$
lower tail test	upper tail test	two tailed tes

- Research Hypothesis $H_0 = \mu \leq 50$
  $H_1: \mu > 50$
- Claim $H_0: \mu \geq 1$
  $H_1: \mu < 1$

  equality H_0

  $H_0 \ \mu = 1$

  $H_1 \ \mu \neq 1$

page 287

① $H_0: \mu \leq 600$
$H_1: \mu > 600$

② $H_0: \mu \leq 14$
$H_1: \mu > 14$

③ equilty
$H_0 \ \mu = 0.75$
$H_1 \ \mu \neq 0.75$

④ $H_0: \mu \geq 320$
$H_1: \mu < 320$

The level of signifegance α is the probabilty of making a type one error (rejecting Ho when Ho is true)

Type II erro ~~accepting~~ rejecting H_1 when H_1 is true.

Example 1:

suppose the lable on a large bottle of cola states that the bottle contains (3) liters of cola Europen legislation Acknowelgeds that the botteling processe can't garntee exactley (3) liters of cola in each bottle however if the population mean filling valume is at least (3) liters the right of consumers will be protected suppouse a sample of 36 bottles is selected and the sample mean computed an estimate of population mean is 2.92 liters the population standard deration is 0.18 and the level of signfignese is 0.1

① $H_0: \mu \geq 3$

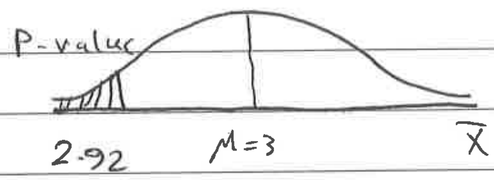
$H_1: \mu < 3$

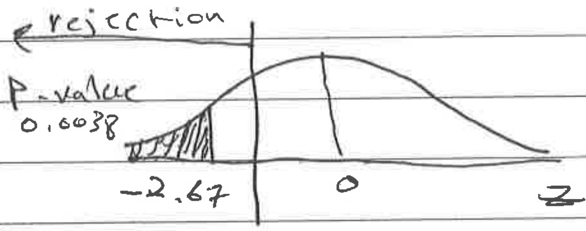
$n=36 \quad \bar{X}=2.92 \quad \sigma=0.18$

$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{0.18}{\sqrt{36}}$

② test statistic

$$Z = \frac{\bar{X}-\mu_0}{\sigma_{\bar{X}}} = \frac{2.92-3}{0.18/\sqrt{36}} = -2.67$$





④ P-value apporach
P-value is the probability
that support
(or lack of support) to H_0

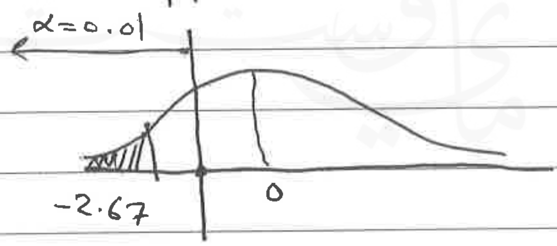
$$P\text{-value} = 0.5 - P(2.67 \leq Z \leq 0)$$
$$= 0.5 - 0.4962$$
$$= 0.0038 < \alpha \quad \text{reject } H_0$$

rule

P-value $\leq \alpha$ reject H_0

P-value $> \alpha$ Do not reject H_0

⑤ Critical value approach



Critical value $-Z_\alpha = -2.32$

$$P(Z_\alpha \leq Z \leq 0) = 0.5 - 0.01 = 0.49$$

$$-Z\alpha = -2.33$$

rule

$Z > -Z\alpha \rightarrow$ do not reject H_0

$Z \leq -Z\alpha \rightarrow$ accept H_0

$-2.67 < -2.33$ reject H_0.

Setps

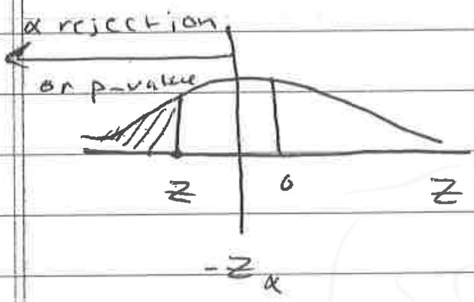
① $H_0: \mu \geq \mu_0$

$H_1: \mu_1 < \mu_0$

② α

③ tes statistic $Z = \frac{\bar{X} - \mu}{\sigma_{\bar{x}}}$

④ P-value / critical approch



⑤ Conlution rejet H_0

if p-value $\leq \alpha$ (P-value approach)

$Z \leq -Z\alpha$ (critical value approch).

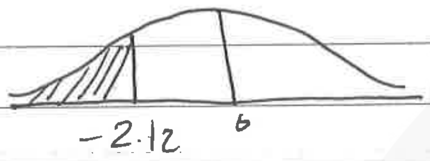
※ 8 page 303

$H_0 : \mu \geq 20$
$H_1 : \mu < 20$

$n = 50$ $\bar{X} = 19.4$ $\sigma = 2$

a) $Z = \dfrac{\bar{X} - \mu_0}{\sigma_{\bar{X}}} = \dfrac{19.4 - 20}{2/\sqrt{50}} = -2.12$

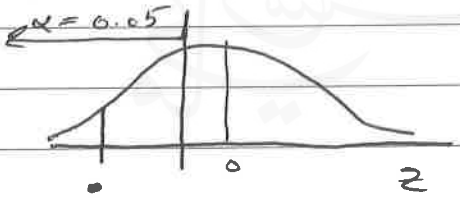
b)



$p\text{-value} = 0.5 - P(-2.12 \leq Z \leq 0)$
$= 0.5 - 0.483 = 0.017$

c) $\alpha = 0.05$
$0.017 < 0.05$
$p\text{-value} < \alpha$ reject H_0

d



$-2.12 - Z_\alpha$

$P(-Z_\alpha \leq Z < 0) = 0.5 - 0.05 = 0.45$

$0.4495 < 0.45 < 0.4505$

$Z_\alpha = \dfrac{1.64 + 1.65}{2} = 1.645$, $-2.12 < -1.645$

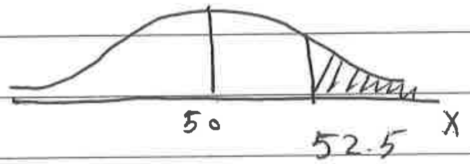
reject H_0

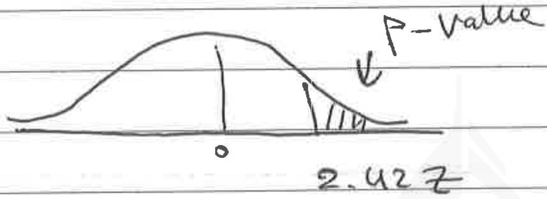
10) b) $H_0: \mu \leq 50$

$H_1: \mu > 50$

$n = 60 \quad \bar{X} = 51 \quad \sigma = 8 \quad \alpha = 0.01$

a) $Z = \dfrac{\bar{X} - \mu_0}{\sigma_{\bar{X}}} = \dfrac{52.5 - 50}{8/\sqrt{60}} = 2.42$

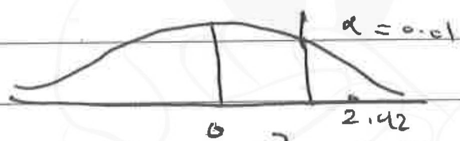




$P\text{-value} = 0.5 - P(0 \leq Z \leq 2.42)$

$= 0.5 - 0.4922 = 0.088$

$0.0088 < 0.01 \Rightarrow$ reject H_0



$Z_\alpha = 2.33$

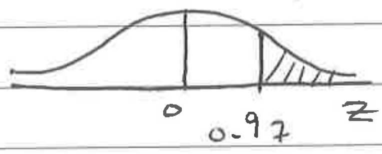
$P(0 \leq Z \leq Z_\alpha) = 0.5 - 0.01 = 0.49$

$Z_\alpha = 2.33$

$2.42 > 2.33 \Rightarrow$ reject H_0

$\bar{X} = 51$

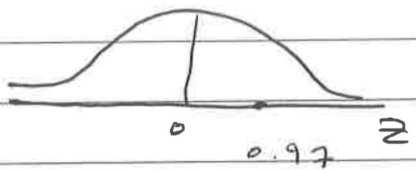
$Z = \dfrac{51 - 50}{8/\sqrt{60}} = 0.97$



$P\text{-value} = 0.5 - 0.334 = 0.166$

$0.166 > 0.01$

do not reject H_0



$z_\alpha = 2.33$

$0.97 < 2.33$ do not rejet H_0

week ③

Setps

① define iris:

lower tail test	upper tail test	two tailed test
$H_0: \mu \geq \mu_0$	$H_0: \mu \leq \mu_0$	$H_0: \mu = \mu_0$
$H_1: \mu < \mu_0$	$H_1: \mu > \mu_0$	$H_1: \mu \neq \mu_0$

② find α

③ test statistac $z = \dfrac{\bar{x} - \mu_0}{\sigma_{\bar{x}}} = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$

④ p-value

⑤

Critical value

Rejetion rule p-value $\leq \alpha \rightarrow$ reject H_0

Example:

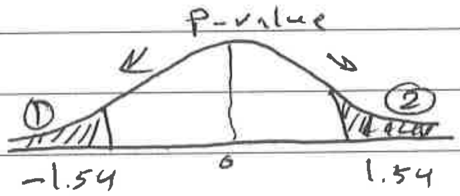
page 303

※11. $H_0 : \mu = 39.2$

$H_1 : \mu \neq 39.2$

b) $n = 112$ $\bar{X} = 38.5$ $\sigma = 4.8$

$$z = \frac{38.5 - 39.2}{4.8/\sqrt{112}} = -1.54$$



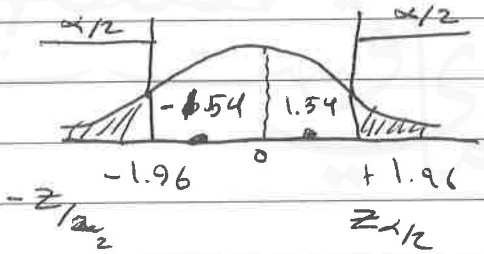
Area ① $= 0.5 - p(-1.54 \leq z \leq 0)$

$= 0.5 - 0.4382$

p-value $= 2(0.5 - 0.4382) = 0.1236$

c) $\alpha = 0.05$ $0.1236 > \alpha$ don't reject H_0

d) $\alpha = 0.05$



$p(0 \leq z \leq +z_{\alpha/2}) = 0.5 - 0.025 = 0.475$

$-1.54 > -1.96$ $z_\alpha = 1.96$

don't reject H_0

page 304

12) lower tailed test

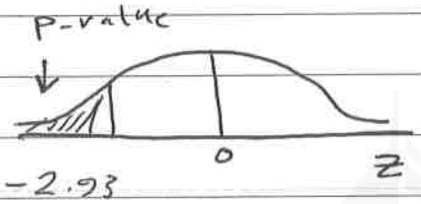
$H_0: \mu \geq 181900$

$H_1: \mu < 181900$

b) $n = 40$ $\quad \bar{X} = 166400$ $\quad \sigma = 33500$

$$Z = \frac{166400 - 181900}{33500/\sqrt{40}} = -2.93$$

c)



p-value $= 0.5 - p(-2.76 \leq z \leq 0)$

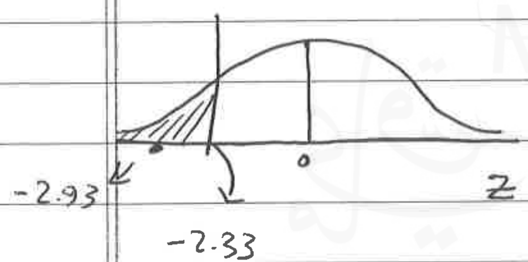
$= 0.5 - 0.4983$

$= 0.0017$

$\alpha = 0.01$

$0.0017 < \alpha$ rejet H_0

d) critical value



$p(-z \leq z \leq 0) = 0.5 - 0.01 = 0.49$

$-z = -2.33$

$-2.93 < -2.33$ rejet H_0

13) upper tail test

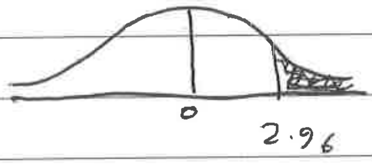
$H_0: \mu \leq \mu_0$

$H_1: \mu > \mu_0$

b) $n = 35$ $\quad \bar{x} = 17$ $\quad \sigma = 4$

$$z = \frac{17 - 15}{4/\sqrt{35}} = 2.96$$

c)

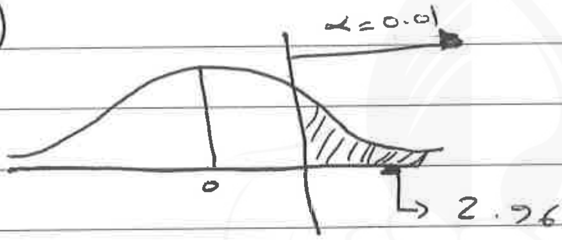


p-value $= 0.5 - (0 \leq z \leq 2.96)$

$= 0.5 - 0.4985$

$= 0.0015 < \alpha$ reject H_0

d)



$z_{0.01}$

$= 2.33$

$p(0 \leq z \leq z_{0.01}) = 0.5 - 0.01 = 0.49$

$z_{0.01} = 2.33$

$2.33 < 2.96$

reject H_0

14)

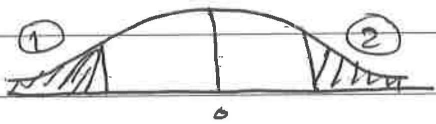
$H_0: \mu = 8$

$H_1: \mu \neq 8$

b) $n = 120$ $\bar{X} = 8.5$ $\sigma = 3.2$

$$Z = \frac{8.5 - 8}{3.2/\sqrt{120}} = 1.71$$

p-value

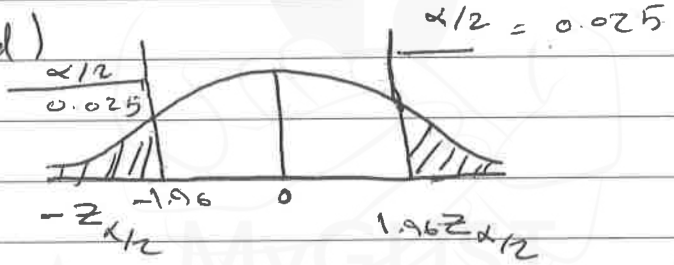


$2(0.5 - 0.4564) = 0.0872$

c) $\alpha = 0.05$

$0.0872 > \alpha$ do not rejekt

d)



$1.71 < 1.96$ do not reject H_0

$H_0: \mu \leq 12$

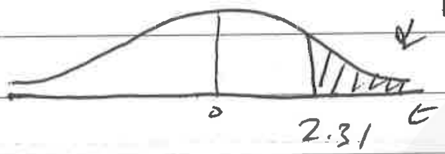
$H_1: \mu > 12$

t

$n = 25 \quad \bar{X} = 14 \quad S = 4.32$

~~n~~ a) $t = \dfrac{\bar{X} - \mu_0}{s/\sqrt{n}} = \dfrac{14 - 12}{4.32/\sqrt{35}} = 2.31$

b) $df = n-1 = 25-1 = 24$

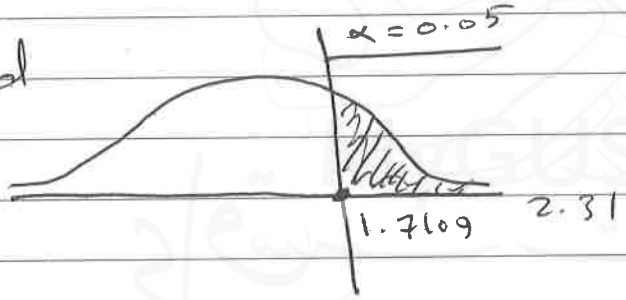


$2.0639 < 2.31 < 2.4922$

$0.025 > \text{p-value} > 0.01$

c) $\alpha = 0.05$ p value $< \alpha$ rejet H_0

d



$df = 24$

$t_{0.05, 24} = 1.7109$

$2.31 > 1.709$

rejet H_0

page 309

※ 18

$H_0 : \mu = 18$

$H_1 : \mu \neq 18$

$n = 48 \quad \bar{X} = 17 \quad S = 4.5$

a) $t = \dfrac{17-8}{4.5/\sqrt{48}} = -1.54$

b) $df = 48-1 = 47 \rightarrow \infty$ infinity row

$1.2816 < 1.54 < 1.6449$

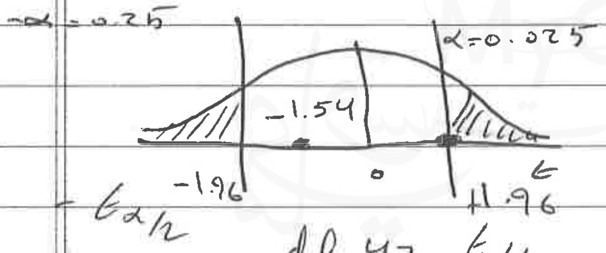
$0.1 > \text{Area} > 0.05$

$0.2 > \text{p-value} > 0.1$

c) $\alpha = 0.05$

p-value $> \alpha$ do not reject H_0

critical test



$df = 47 \quad t_{\alpha/2}$

$t_{0.025, 47} = 1.96$

$-1.54 > -1.96$ do not reject H

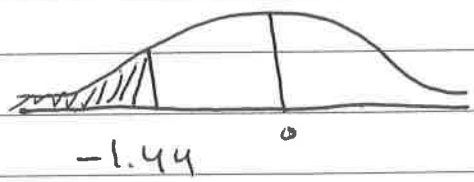
20) $H_0: \mu \geq 300$

$H_1: \mu < 300$

$n = 30 \quad \bar{X} = 299.5 \quad S = 1.9$

b) $t = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} = \frac{299.5 - 300}{1.9/\sqrt{30}} = -1.44$

$df = 30 - 1 = 29$

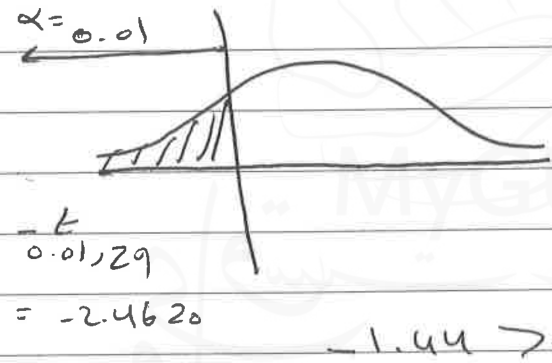


$1.3114 < 1.44 < 1.6991$

$0.1 > p\text{-value} > 0.05$

c) $\alpha = 0.01$

p-value $> \alpha$ do not reject H_0



$-1.44 > -2.4620$

do not reject H_0

21) $H_0: \mu = 600{,}000$

$H_1: \mu \neq 600{,}000$

$n = 40$ $\bar{X} = 612000$ $S = 65000$

b) $t = \dfrac{612000 - 600{,}000}{6500000/\sqrt{46}} = 1.17$

$df = 40 - 1 = 39$

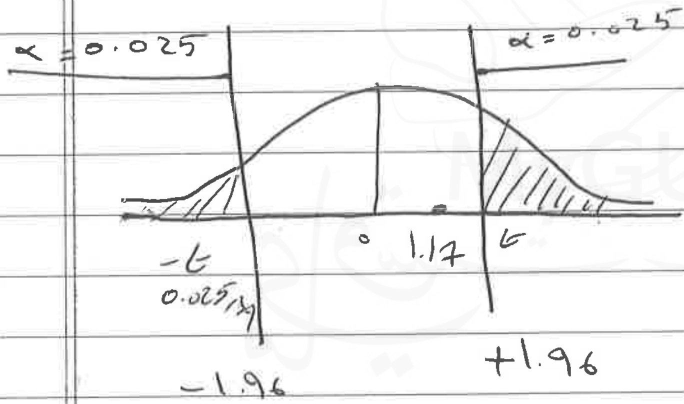
$0.67452 < 1.17 < 1.2816$

$0.25 > \text{Area} > 0.1$

$0.5 > \text{p-value} > 0.2$

c) $\alpha = 0.05$

p-value $> \alpha$ do not reject H_0



$1.17 < 1.96 \Rightarrow$ don't reject H_0

22). $H_0: \mu \leq 4671264$

$H_1: \mu > 4671264$

b) $n = 30$ $\quad \bar{X} = 5802333$ $\quad S = 2460810$

$$t = \frac{5802333 - 4671264}{2460810/\sqrt{30}} = 2.52$$

p-value

0 2.52

$df = 30 - 1 = 29$

$2.462 < 2.52 < 2.764$

$0.01 > p\text{-value} > 0.005$

p-value $< \alpha$ reject H_0

c) $\alpha = 0.05$

problem set III

End of test

one

Chapter 13:

Analysis of variance (ANOVA).

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 \ldots\ldots \mu_n$

H_1: H_0 false

a company manufactures printers at plants located in China USA and France to messure how much employees at these plants know about total quality managment a random sample of six employees was selected from each plant and given a quality awarness examanation

test to see if the average exam scores defer between the three populations

	USA	France	China
n	6	6	6
$\bar{X}$	79	74	66
S^2	34	20	32

$K = 3$ number of pop (treatment)

examination score $\Rightarrow$ response variable

$H_0: \mu_1 = \mu_2 = \mu_3$

H_1: H_0 false

Assumption; 1- Response variable should be normally distrbuat

2- all pop have the same σ^2

3- The obs are independent.

Between treatments estimate of σ^2 = Mean Squere to treatment.

$$= MSTR = \sum_{j=1}^{K} \frac{n_j(\bar{x}_j - \bar{\bar{x}})^2}{K-1}$$

$\bar{\bar{X}}$ over all ~~mean~~ sample mean

$$= \sum_{j=1}^{K} \frac{\sum_{i} x_{ij}}{n_T}$$ ~~$n_1 x_1 + n_2 + n_3$~~

$$\bar{\bar{X}} = \frac{n_1\bar{X}_1 + n_2\bar{X}_2 + n_3\bar{X}_3}{n_1 + n_2 + n_3} = \frac{6(79) + 6(74) + 6(66)}{6+6+6}$$

$$= 73$$

$SSTR = 6(79-73)^2 + 6(74-73)^2 + 6(66-73)^2 = 516$

$$MSTR = \frac{516}{3-1} = 258$$

Pooled or withen treatments estimate of σ^2

= Mean square due in Error

$$= MSE = \frac{\sum_{j=1}^{K} (n_j - 1) s_j^2}{n_T - K}$$

Sum of square duein erro = SSE

$SSE = (6-1)\,34 + (6-1)\,20 + (6-1)\,32 = 430$

$$MSE = \frac{SSE}{n_T - K} = \frac{430}{6+6+6-3} = 28.67$$

$$F = \frac{MSTR}{MSE} = \frac{258}{28.67} = 9$$ upper tale

$df_n = K-1 = 3-1 = 2$

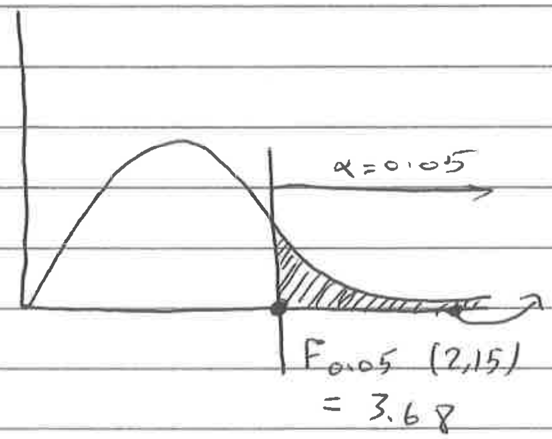
$df_d = n_T - K = 6+6+6-3 = 15$

$9 > 6.36$

P-value < 0.01

$\alpha = 0.05$ P-value < α

reject H_0



$9 > 3.68$

reject H_0

ANOVA table

	df	SS	MS	F
Treatments	2 $K-1$	516 SSTR	258 MSTR	9 $\frac{MSTR}{MSE}$
Error	+ 15 n_T-K	450 + SSE	28.67 MSE	
	17 n_T-1	SST 946 ← sum of squares		

~~number ① page~~ 441

$MSTR = \frac{SSTR}{K-1}$ $\quad SSTR = \sum_{j=1}^{K} n_j(\bar{x}_j - \bar{\bar{x}})^2$

$\bar{\bar{x}}$ overall sample mean $= \sum_{j=1}^{K} \frac{n_j \bar{x}_j}{n_T}$

$MSE = \frac{SSE}{n_T - K}$ $\quad SSE = \sum_{j=1}^{K} (n_j - 1) s_j^2$

$F = \frac{MSTR}{MSE}$ $\quad df_N = K-1$ $\quad df_N = n_T - K$

problem 1 - p441

$K = 3$

	S_1	S_2	S_3
n	5	5	5
$\bar{x}$	30	45	36
s^2	6	4	6.5

$MSTR \; \frac{SSTR}{K-1} = \frac{5(30) + 5(45) + 5(36)}{5+5+5} = 37$

$SSTR = 5(30-37)^2 + 5(45-37)^2 + 5(36-37)^2 = 570$

$MSTR = \frac{570}{3-1} = 285$

b) $MSE = \frac{SSE}{n_T - K}$

$SSE = (5-1)6 + (5-1)4 + (5-1)6.5 = 66$

$MSE = \frac{66}{5+5+5-3} = 5.5$

c) $H_0: \mu_1 = \mu_2 = \mu_3$

H_1: H_0 false $\qquad \alpha = 0.05$

$F = \frac{MSTR}{MSE} = \frac{285}{5.5} = 51.82$

$df_N = 3-1 = 2 \qquad df_d = 5+5+5-3 = 12$

$51.82 > 6.93$

P-value < 0.01

P-value $< \alpha \rightarrow$ reject H_0

d)

	df	SS	MS	F
Treatment	2	570	285	51.82
Exam	12	66	5.5	
	14	636		

ركزنا عليها page 442

problem ③

$K = 3$

	S_1	S_2	S_3
n	4	6	5
$\bar{x}$	100	85	79
S^2	35.35	35.6	43.5

a) $MSTR = \dfrac{SSTR}{K-1}$ $\quad \bar{\bar{x}} = \dfrac{4(100) + 6(85) + 5(79)}{4+6+5} = 87$

$SSTR = 4(100-87)^2 + 6(85-87)^2 + 5(79-87)^2 = 1020$

$MSTR = \dfrac{1020}{3-1} = 510$

b) $MSE = \dfrac{SSE}{n_T - K}$

$SSE = (4-1)\,35.33 + (6-1)\,35.6 + (5-1)\,43.5$

$= 458$

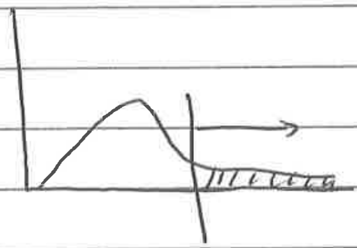
$MSE = \dfrac{458}{4+6+5-3} = 38.17$

c) $H_0: \mu_1 = \mu_2 = \mu_3$

H_1: H_0 false $\qquad \alpha = 0.05$

$F = \dfrac{MSTR}{MSE} = \dfrac{510}{38.17} = 13.36$

$df_N = 3-1 \qquad df_d = 4+6+5-3 = 12$



$F_{0.05}(2, 12) = 3.89$

$13.36 > 3.89$

reject H_0

D)

table				F
treatment	2	1020	510	13.36
error	12	958	38.17	
	14	1970	[illegible]	

problem 4)

$K=4$ $\quad n_T = 16+16+16+16$

~~Treatment~~	df	SS	MS	F
Treatment	$K-1=4-1=3$	1200	400	80
Error	60	300	5	
		1500		

MSTR

$$MSTR = \frac{SSTR}{K-1}$$

$$SSTR = MSTR(K-1)$$
$$=400(3)$$
$$=1200$$

$$SSE = SST - SSTR$$
$$= 1500 - 1200$$
$$=300$$

$$df_D = n_T - K$$
$$=16+16+16+16-4=60$$

$$MSE = \frac{SSE}{n_T-K} = \frac{300}{60} = 5$$

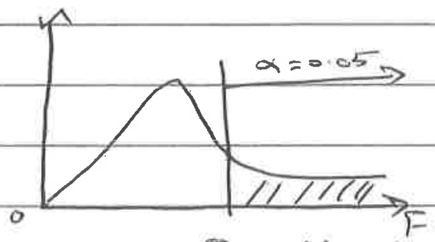
$$F = \frac{MSTR}{MSE} = \frac{400}{5} = 80$$

$\alpha = 0.05$

$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$

$H_1: H_0$ false

$F=80 \quad df_N = 3 \quad df_D = 60$



$F_{0.05}(3,60)$

$=2.76$

$80 > 2.76$

reject H_0

problem 5

$K = 3$ $\qquad\qquad$ $n_T = 25 + 25 + 25 = 75$

	df	SS	MS	F
Treatments	$K-1 = 3-1 = 2$	120	$120/2 = 60$	$\frac{60}{3} = 20$
Error	$n_T - K = 75-3 = 72$	216	$216/72 = 3$	
	74	336		

$\alpha = 0.05$

$H_0: \mu_1 = \mu_2 = \mu_3$

H_1: H_0 false

$df_N = 2$ $\qquad$ $df_D = 72 \approx 60$ $\qquad$ اخذ الاقرب 60

20 74.98

p-value < 0.01

p-value < α

reject H_0

CH13 16-18 11-12-13

Problem set

~~10~~ 16 Asw: D

18) $\bar{\bar{x}} = \frac{45(30)+20(35)+13(40)+19(50)}{45+20+13+19} = 36.29$

EX 41) $K=4$ $n_T = 5+5+5+5 = 20$

	df	SS	MS	F
Treatments	$K-1=4-1=3$	200	66.67	1.77
Error	$n_T-K=20-4=16$	+ 600	37.5	
	19	800		

$SSTR = SST - SSE$

$= 800 - 600 = 200$

$MSTR = \frac{600}{16} = 37.5$

$F = \frac{66.67}{37.5} = 1.77$

25 $\alpha = 0.05$ $F = 1.78 < 2.46$

$df_N = 3$ p-value > 0.1

$df_D = 16$ p-value $> \alpha$. do not reject H_0

Problem ②

$K = 3$ $\qquad$ $n_T = 18 + 10 + 15 = 43$

	df	SS	MS	F
Treatment	$K-1 = 3-1 = 2$	36	18	3
Error	$n_T - K = 43-3 = 40$	240	6	
	42	276		

$$F = \frac{MSTR}{MSE} \Rightarrow MSTR = F(MSE)$$
$$= 3 \times 6 = 18$$

$$SSTR = MSTR\,(K-1) = 18 \times 2 = 36$$
$$SSE = MSE\,(n_T - K) = 6(40) = 240$$

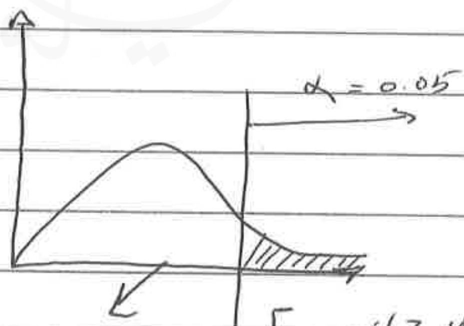
$\alpha = 0.05$

$H_0: \mu_1 = \mu_2 = \mu_3$

$H_1: H_0$ false

$F = 3$

$df_N = 2 \qquad df_D = 40$



$F_{0.05}(2,40)\ F$

$F < 3.23 \qquad = 3.23$

do not reject H_0

Chapter 14

Sempile ~~re~~ linear regresion model.

2 variamples.

Sales y = explain variable, dependt variable, Predicted variable

Adver X = explanatory variable

independent variable

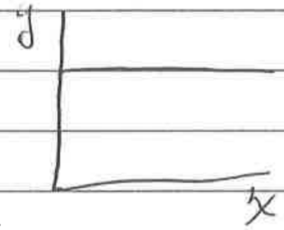
Predictory

$y = mx + b \rightarrow$ y intersept

slope

$x = 0$

$y = b$



regression coffecint

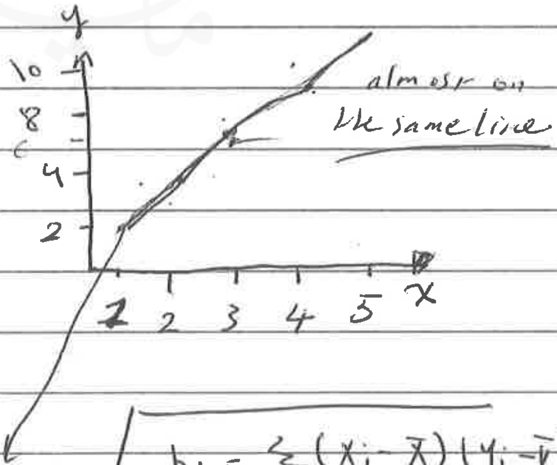
$\cancel{y = \beta_0 + \beta_1 x}$ / $y = \beta_0 + \beta_x + \varepsilon$

error term

Adi. exp		sales volume
x		y
1	given	2
2		4
3		~~6~~ 7
4		9
5		10

$n = 5$

Scatter diagram



$y_i = b_0 + b_i x_i + e_i$

$\hat{y}$ estimated regression equation

y obsefed values of y

$\hat{y}$ estimated or predicted value of y

estimated regression line.

$$b_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

$$b_0 = \bar{y} - b_1 \bar{x}$$

find the estimated regression equation
find the least squares estimated regression equation
find ,, ,, ,, —— ,, line

$y = b_0 + b_1 x$

Ad. Ex	Sales
x	y
1	2
2	4
3	7
4	9
~~10~~5	10

	Ad. Ex	Sales	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i-\bar{x})(y_i-\bar{y})$	$(x_i - \bar{x})^2$
	x	y				
	1	2	-2	-4.4	8.8	4
	2	4	-1	-2.4	2.4	1
	3	7	0	0.6	0	0
	4	9	1	2.6	2.6	1
	5	10	2	3.6	7.2	4
Total	15	32	0		21	10

$\bar{x} = \frac{15}{5} = 3$

$\bar{y} = \frac{32}{5} = 6.4$

$b_1 = \frac{21}{10} = 2.1$ slope

$b_0 = 6.4 - 2.1(3) = 0.1$ y intersept

$\hat{y} = 0.1 + 2.1x$

predict sales volume for a ~~value~~ value of $x = 6$

$$\hat{y} = 0.1 + 2.1(6) = 12.7$$

predict value of y when $x = 3$

$$\hat{y} = 0.1 + 2.1(3) = 6.4$$

$$e_i = \hat{y}_i - \hat{y}_i = 7 - 6.4 = 0.6$$

error term, residual

Summery

$$Y = \beta_0 + \beta_1 x + \varepsilon \hookrightarrow \text{error term}$$

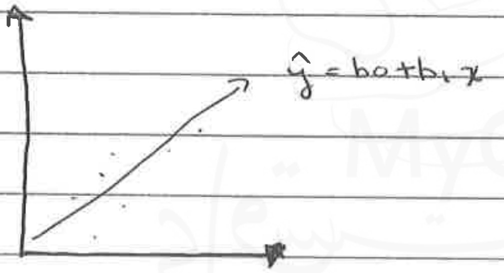
regression coffecient

n

$$y_i = b_0 + b_1 x_i + e_i \rightarrow \text{residual}$$

estimated regression coffecint

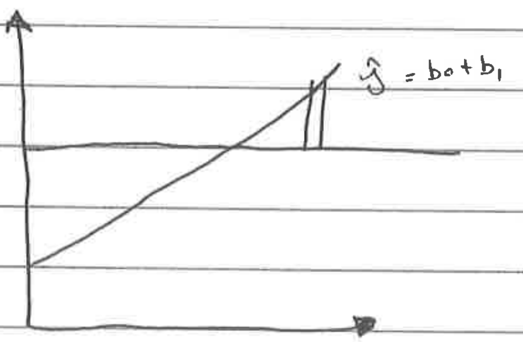
$\hat{y}_i = b_0 + b_1 x_i$



use OLS $\min \sum e_i^2 = \min \sum (y_i - \hat{y})^2$

$$b_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

$$b_0 = \bar{y} - b_1 \bar{x}$$



find the coffecint of determination [how well ~~do~~ does the estimated regression equation fit the data

what is the percentage of the variablity in y that is explained by ~~the~~ the relationship with x] same thing

[مهمة من علي]

Adv. exp — sales value

x	y	$(y_i-\bar{y})^2$	$(y_i-\bar{y})$	$\hat{y}$	$\hat{y}-\bar{y}$	$(\hat{y}-\bar{y})^2$
1	2	19.36	-4.4	2.2	-4.2	17.64
2	4	5.36	-2.4	4.3	-2.1	4.41
3	7	0.36	0.6	6.4	0	0
4	9	6.76	2.6	8.5	2.1	4.41
5	10	12.96	3.6	10.6	4.2	17.64
		SST = 45.2				SSR = 44.1

$0 \le R^2 \le 1$

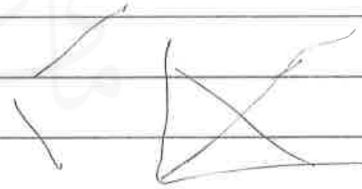
$R^2 = \frac{\cancel{44}\,44.1}{45.2} = 0.98 \to 1$

good fit

98% of the regression ~~line~~ in y is explained by the regerssion

---

sample ~~coffeeion~~ correlation coefficient



relateion $\to$ $r_{xy} = (\text{Sip of } b_1)\sqrt{R^2}$

$= +\sqrt{0.98}$

$= +0.99$

$r_{xy} \to 1$ strong liner positive relationship between x and y

$r_{xy} \to \pm 1$ perefect

$r_{xy} \to -1$ negative relationship

$r_{xy} \to 0$ no linear relation

مراجعة

$0 \leq R^2 \leq 1$

$R^2 = \frac{SSR}{SST} = \frac{\Sigma(\hat{y}_i - \bar{y})^2}{\Sigma(y_i - \bar{y})^2}$ $\longrightarrow 1$ good fit, $\longrightarrow 0$ bad fit

$r_{xy} = (\text{sign } b_1)\sqrt{R^2} = +\sqrt{0.98} = +0.99$ $-1 \leq r_{xy} \leq 1$

$r_{xy} \to +1$ strong positive lineare relation

$r_{xy} \to -1$ " negative " "

$r_{xy} \to$ no linear relation.

b_0 and b_1 random variables

b_1 random variable

- Expected $(b_1) = \beta_1$
- Stand. Dev of b_1 $\sigma_{b_1} = \frac{\sigma}{\sqrt{\Sigma(x_i - \bar{x})^2}}$ S_e or $S_b = \sqrt{\frac{MSE}{\Sigma(x_i - \bar{x})^2}}$

with $= \frac{SSE}{n-2}$

$SSE = \Sigma(y_i - \hat{y}_i)^2$

$H_0: \beta_1 = 0$ $\alpha = 0.05$

$H_1: \beta_0 \neq 0$ $t = \frac{b_1 - \beta_{10}}{S_{b_1}}$ $df = n-2$

2 hidden test

$SSE = SST - SSR = 45.2 - 44.1 = 1.1$

$MSE = \frac{1.1}{5-2} = 0.37$

$S_{b_1} = \sqrt{\frac{0.37}{10}} = 0.19$

$t = \frac{2.1 - 0}{0.19} = 11.05$

$df = n-2$

$df = 5 - 2 = 3$

$5.8409 < 11.05 < 12.924$

$0.005 > \text{Area} > 0.005$

$0.01 > \text{P-value} > 0.001$

p-value $< \alpha$ reject H_0

X significant variable.

CI for β_1

$\beta_1 = b_1 \pm t_{\alpha/2} \, S b_1$ $\quad df = n-1$

$= 2.1 \pm 3.1825 \, (0.19$

$1-\alpha = 0.95$ $\quad = 2.1 \pm 0.605$

$\alpha = 0.05$ $\quad \beta_1 \in [1.495 ; 2.705]$

$\left.\begin{array}{l}\alpha/2 = 0.025 \\ df = 3\end{array}\right\} t_{\frac{\alpha}{2}} = 3.1825$

$H_0 : \beta_1 = 0$

$H_1 : \beta_1 \neq 0$

$F = \frac{MSR}{MSE}$

$df_N = 1$

$df_D = n-2$

$MSR = \frac{SSR}{1} = \frac{44.1}{1} = 44.1$

$MSE = \frac{SSE}{n-2} = 0.37$

$F = \frac{44.1}{0.37} = 119.18$

$df_N = 1$

$df_D = 5-2 = 3$

$119.18 > 34.12$

p-value < 0.01

p-value $< \alpha$

reject H_0

Model is Signific

Anova

	df	SS	MS	F
Regression	df_N 1	SSR $= 44.1$	MSR $= 44.1$	$\frac{MSR}{MSE}$
Error	$df_D = 3$	SSE 1.1	MSE $= 0.37$	$= 119.18$

$\cancel{CH}$ $\mathcal{\cancel{L}}$ خ 1 = 1.5.1.2

CH14

[illegible]

[illegible]

problem set 14

1

9333,210

9

1111126

7
5

critical

2

10

$$y = \underbrace{b_0 + b_1 x}_{\hat{y}} + e_i$$

OLS method $\Sigma e_i^2 = \min \Sigma (y - \hat{y})^2$ $\leftarrow$

slope $\Rightarrow$ $b_1 = \dfrac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{\Sigma (x_i - \bar{x})^2}$

independt $\leftarrow$ $b_0 = \bar{y} - b_1 \bar{x}$

$R^2 = \dfrac{SSR}{SST} = \dfrac{\Sigma (\hat{y} - \bar{y})^2}{\Sigma (y - \bar{y})^2}$ $\quad 0 \leq R^2 \leq 1$

$r_{xy} = (\text{Sign } b_1) \sqrt{R^2}$ $\quad -1 \leq r_{xy} \leq +1$

$H_0: \beta_1 = 0$

$H_1: \beta \neq 0$

$t = \dfrac{b_1}{s_{b_1}}$ $\quad s_{b_1} = \sqrt{\dfrac{MSE}{\Sigma (x_i - \bar{x})^2}}$

$df = n - 2$ $\quad MSE = \dfrac{SSE}{n-2}$

$F = \dfrac{MSR}{MSE}$ $\quad MSR = \dfrac{SSR}{1}$

upper tail

$df_n = 1$

$df_d = n - 2$

Problem set 14 Exhibit 1

y	x	$x_i-\bar{x}$	$y_i-\bar{y}$	$(x_i-\bar{x})(y_i-\bar{y})$	$(x_i-\bar{x})^2$
1	1	-2	-2	4	4
2	2	-1	-1	1	1
3	3	0	0	0	0
4	4	1	1	1	1
5	5	2	2	4	4
		0	0	10	10

$b_0 = \bar{y} - b_1\bar{x} = 3-1(3) = 0$

$\underline{b_1} = \frac{\Sigma(x_i-\bar{x})\;(y_i-\bar{y})}{\Sigma(x_i-\bar{x})^2} = \frac{10}{10} = 1$

$\bar{y} = \frac{\Sigma y}{n} = \frac{15}{5} = 3$

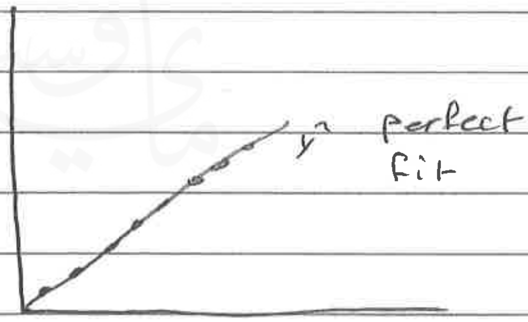
$\hat{y} = b_0 + b_1 x$

$= x$

$\bar{x} = \frac{\Sigma x}{n} = \frac{15}{5}\ 3$

$R^2 = 1$

$r_{xy} = +\sqrt{R^2} = 1$



5) $r_{xy} = \text{Sign } b_1 \sqrt{R^2}$

$= \pm\sqrt{1}$

Exhibit ②

y	x	$x_i - \bar{x}$	$y_i - \bar{y}$	$(x_i-\bar{x})(y_i-\bar{y})$	$(x_i-\bar{x})^2$	$(y_i-\bar{y})^2$
25	14	0	-3	0	0	9
27	16	2	-1	$-\cancel{8}2$	4	1
33	12	-2	5	-10	4	25
27	14	0	-1	0	0	1
112	56	0	0	$\cancel{-4}$ -12	8	

$\bar{y} = \frac{\sum y_i}{n} = \frac{112}{4} = 28$

$\bar{x} = \frac{\sum x_i}{n} = \frac{56}{4} = 14$ $\qquad \hat{y} = 49 - 1.5x$

$\hat{y} = 49 - 1.5x$

$b_1 = \frac{-12}{8} = -1.5$

$b_0 = 28 - (-1.5)(14) = 49$

8) $SST = \sum(y_i - \bar{y})^2 = 36$

9) $R^2 = \frac{SSR}{SST} = \frac{SST - SSE}{SST} = \frac{36-18}{36} = 0.5$

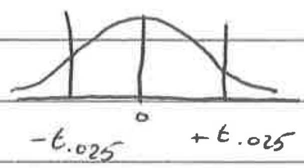
10) $S_{b_1} = \sqrt{\frac{MSE}{\sum(x_i-\bar{x})^2}} = \sqrt{\frac{18/(4-2)}{8}} = 1.06$

11) $H_0: \beta_1 = 0$ $\qquad t = \frac{b_1}{S_{b_1}} = \frac{-1.5}{1.06} = -1.42$

$H_1: \beta_1 \neq 0$

12) $\alpha = 0.05$ $\qquad \left.\begin{array}{l}\alpha/2 = 0.025 \\ df = n-2 = 2\end{array}\right\} t_{\alpha/2} = 4.3027$

table



13) $\hat{y} = 49 - 1.5(30) = 4$

14) $y = 3$ $\qquad y - \hat{y} = 3 - 4 = -1$

Exhibit 3

19

$$\boxed{MSE = \frac{SSE}{n-2}}$$

Chapter 15 second class.

$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 \ldots + \beta_p x_p$

$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \ldots b_p x_p$.

test for individual significance

(1 variable in sign)

$H_0: \beta_i = 0 \qquad i = 1 \ldots p$

$H_1: \beta_0 \neq 0$

$t_i = \frac{b_i}{S_{b_i}} \quad df = n-p-1$ (2 tailes)

test for overall significance

(model in sign)

$F = \frac{MSR}{MSE}$

$MSR = \frac{SSR}{p} \qquad MSE = \frac{SSE}{n-p-1} \qquad df_n = p$

$df_D = n-p-1$

$\boxed{\beta_i = b_i \pm t_{\alpha/2} S_{b_i} \quad df = n-p-1}$ upper tail.

problem 1 problem set 15

1 $n = 12 \qquad p = 2$

a) $\hat{y} = b_0 + b_1 x_1 + b_2 x_2$

$= 17.145 - 0.104 X_1 + 1.376 X_2$

b) As unitry price increases by 1 \$, Sales volume will decress by 0.14 given that ~~x_2~~ everything else is kept constant.

c) $X_1 = 2000 \qquad X_2 = 10$

$\hat{y} = 17.145 - 0.104(2000) + 1.376(10) = 30.697$

d) $\alpha = 0.05$

$H_0: \beta_1 = 0 \qquad t_1 = \frac{b_1}{S_{b_1}} = \frac{0.104}{3.282} = -0.032$

$H_a: \beta_1 \neq 0$

$df = n-p-1 = 12-2-1 = 9$

$0.032 < 0.261$

Area > 0.4

P-V > 0.8

P-V $> \alpha$ do not reject

H_0 X_1 not signif

e)

$H_0: \beta_2 = 0$

$H_1: \beta_2 \neq 0$

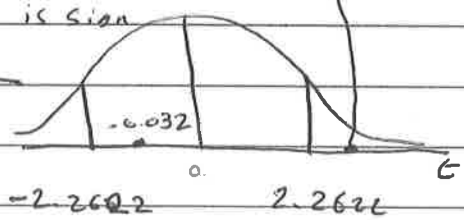
$$t_2 = \frac{b_2}{Sb_2} = \frac{1.376}{0.250} = 5.504$$

$5.504 > 2.2622$

reject H_0

X_2 is sign

$-0.327 - 2.2622$



f) $1-\alpha = 95\%$

$\alpha = 0.05 \Rightarrow \alpha/2 = 0.025 \quad df = 9$

$$\beta_2 = b_2 \pm t_{\alpha/2} \quad Sb_2$$

$= 1.376 \pm 2.2622 \ (0.250)$

$= 1.376 \pm \cancel{2.2622}\ 0.56555$

$\beta_2 \in [0.81045 ; 1.94155]$

problem ②

$MSR = \frac{655.955}{2} = 327.9775$ (look at table)

$SSE = 838.917 - 655.955 = 182.962$

~~MSR~~

$MSE = \frac{182.962}{9} = 20.329111$

$F = \frac{MSR}{MSE} = \frac{327.9775}{20.32911} = 16.133$

$H_0: \beta_1 = \beta_2 = 0$

$H_1: H_0$ false

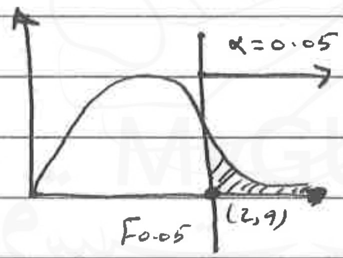
$F = 16.133 \quad df_n = 2 \quad df_d = 9 \quad \alpha = 0.05$

$16.133 > 8.02$

p-value < 0.01

$p\text{-}v < \alpha \rightarrow$ reject H_0

model is sign



$F_{0.05} = 4.26$

$16.133 > 4.26$

reject H_0

b) $R^2 = \frac{SSR}{SST} = \frac{655.955}{838.917} = 0.782$

problem ③ $p=3$

① $n=$ $p=3$

a) $\hat{y} = 0.0136 + 0.7992\,x_1 + 0.2280\,x_2 - 0.5796\,x_3$

b) $R^2 = \frac{SSR}{SST} = \frac{45.9634}{45.9634 + 2.6218} = 0.946 \Rightarrow$ good fit

$SST = SSR + SSE$

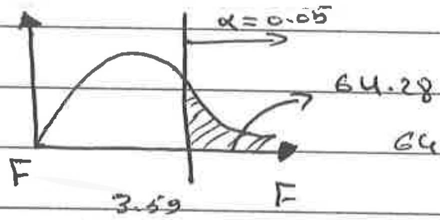
0.946 of the variability in y is explained by the model

c) $H_0: \beta_1 = \beta_2 = \beta_3 = 0$

$H_1: H_0$ false

$$F = \frac{MSR}{MSE} = \frac{45.9634/3}{2.6218/11} = 64.28$$

$\alpha = 0.05$
$df_N = 3$
$df_D = 11$
$\Big\}$ $F_{0.05} = 3.59$



$64.28 > 3.59$

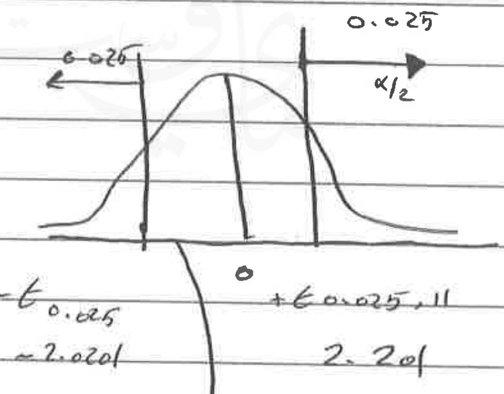
reject H_0, model is sign

d) $\alpha = 0.05$ $H_0: \beta_3 = 0$

$H_1: \beta_3 \neq 0$

$t_3 = \frac{b_3}{S_{b_3}} = \frac{-0.5796}{0.92} = -0.63$

$df = n - p - 1 = 11$



$-0.63 > -2.201$

do not reject H_0

x_3 not sign

e) $\beta_2 = b_2 \pm t_{\alpha/2}\, S_{b_2}$

$1-\alpha = 95\% \Rightarrow \alpha/2 = 0.025$ $df = 11$

$t_{\alpha/2} = 2.201$

$\beta_2 = 0.2280 \pm 2.201\,(0.19)$

$\beta_2 \in [-0.19019,\ 0.64619]$

problem 4 $\quad n =$ $\qquad p = 2$

a) $\hat{y} = 20 + 0.006\, x_1 - 0.7\, x_2$

b) مكتوب

c) $x_1 = 10000 \quad x_2 = 50$

$\hat{y} = 20 + 0.006(10{,}000) - 0.7(50) = 45$

d) $df_E = 20 = n - p - 1$

$20 = n - 2 - 1$

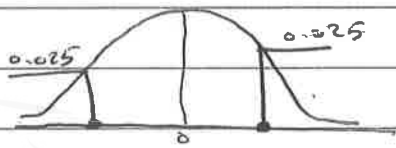
$n = 20 + 3 = 23$

e) $\alpha = 0.05$

$H_0: \beta_1 = 0$

$H_1: \beta_1 \neq 0 \qquad df = 20$

$t_1 = \frac{0.006}{0.002} = 3$



$-2.086 \qquad +2.086$

$3 > 2.086$ reject H_0

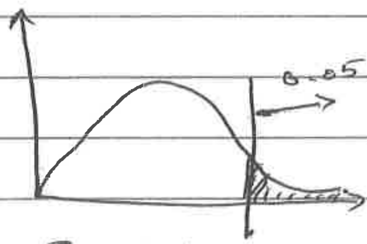
X_1 is Sign

f) (model means F test)

$\alpha = 0.05 \qquad H_0: \beta_1 = \beta_2 = 0$

$H_1: H_0$ false

	df	SS	MS	F
Reg	$2 \to p$	760	380	$380/2 = 190$
error	20	40	2	
		800		



$SSR = 800 - 40 = 760$

$MSR = \frac{760}{2} = 380$

$MSE = \frac{40}{20} = 2$

$F_{0.05}(2, 20)$ F

$= 3.49$

$190 > 3.49$

reject H_0

model is Sign

proble 3 - 382

Formulas

ch 11

1. $\frac{(n-1)S^2}{\chi^2_{\alpha/2,df}} \le \sigma^2 \le \frac{(n-1)S^2}{\chi^2_{1-\alpha/2,df}}$
2. $\chi^2\ (test\ statistic) = \frac{(n-1)S^2}{\sigma^2_0}$
3. F-statistic = $\frac{S^2(1)}{S^2(2)}$
4. $\chi^2\ (test\ statistic) = \sum_{i=1}^{k} \frac{(f_i - e_i)^2}{e_i}$
5. MSTR = $\frac{\sum_{j=1}^{k} n_j(\bar{x}_j - \bar{\bar{x}})^2}{k-1}$ → $MSTR = \frac{SSTR}{K-1}$ (between treatments)
6. SSTR = $\sum_{j=1}^{k} n_j(\bar{x}_j - \bar{\bar{x}})^2$
7. MSE = $\frac{\sum_{j=1}^{k}(n_j - 1)S^2}{n_T - k}$ → $MSE = \frac{SSE}{n_T - K}$ (within treatments)
8. SSE = $\sum_{j=1}^{k}(n_j - 1)\ S^2$
9. F-statistic = $\frac{MSTR}{MSE}$
10. SST = SSTR + SSE
11. $y = \beta_0 + \beta_1 x + \varepsilon$, $\hat{y} = b_0 + b_1 x$
12. $b_1 = \frac{\sum(X_i - \bar{X})(y_i - \bar{y})}{\sum(X_i - \bar{X})^2}$ slope
13. $b_0 = \bar{y} - b_1\bar{x}$ y intercept
14. SST = SSR + SSE $\Rightarrow \sum(Y_i - \bar{Y})^2 = \sum(\hat{Y} - \bar{Y})^2 + \sum(Y_i - \hat{Y})^2$
15. r^2 = SSR/SST → coffeint of determination
16. $r_{x,y} = (sign\ of\ b1)\sqrt{r^2}$ coffeint of Correlation
17. MSE = $\frac{\sum(Y_i - \hat{Y})^2}{n-2}$ SSE
18. (t-statistic) = $\frac{b_i - \beta_i}{se_{bi}}$, where $se_{bi} = \sqrt{\frac{MSE}{\sum(X_i - \bar{X})^2}}$ df=n-2 df=n-2
19. Confidence interval for β_i: $b_i \pm t_{\alpha/2} \cdot se_{bi}$

$MSR = \frac{SSR}{1}$

$MSE = \frac{SSE}{n-2}$

ANOVA table

	df	SS	MS	F
	K-1	SSTR	MSTR	$\frac{MSTR}{MSE}$
Error	n_T-K	SSE	MSE	
		SST		

$(y_i - \bar{y})^2 \to SST$

$(\hat{y} - y)^2 \to SSR$

ANOVA

	df	SS	MS	F
Reg	df=1	SSR	MSR	
error	n-2	SSE	MSE	
		SST		

$F = \frac{MSR}{MSE}$ =

$MSR = \frac{SSR}{1}$

$MSE = \frac{SSE}{n-2}$

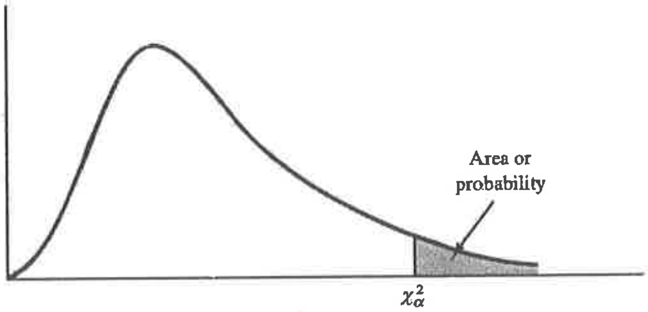
$SST = \sum(y_i - \bar{y})^2$

$SSR = (\hat{y} - \bar{y})^2$

$SSE\ \sum(y_i - \hat{y})^2$

TABLE 3 CHI-SQUARE DISTRIBUTION

Chi



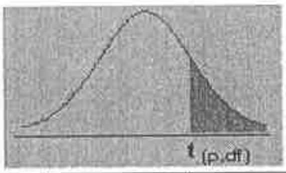
Entries in the table give χ^2_α values, where α is the area or probability in the upper tail of the chi-square distribution. For example, with 10 degrees of freedom and a .01 area in the upper tail, $\chi^2_{.01} = 23.209$.

Degrees of Freedom	Area in Upper Tail .995	.99	.975	.95	.90	.10	.05	.025	.01	.005
1	.000	.000	.001	.004	.016	2.706	3.841	5.024	6.635	7.879
2	.010	.020	.051	.103	.211	4.605	5.991	7.378	9.210	10.597
3	.072	.115	.216	.352	.584	6.251	7.815	9.348	11.345	12.838
4	.207	.297	.484	.711	1.064	7.779	9.488	11.143	13.277	14.860
5	.412	.554	.831	1.145	1.610	9.236	11.070	12.832	15.086	16.750
6	.676	.872	1.237	1.635	2.204	10.645	12.592	14.449	16.812	18.548
7	.989	1.239	1.690	2.167	2.833	12.017	14.067	16.013	18.475	20.278
8	1.344	1.647	2.180	2.733	3.490	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.041	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.041	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	14.848	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	33.196	36.415	39.364	42.980	45.558
25	10.520	11.524	13.120	14.611	16.473	34.382	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	35.563	38.885	41.923	45.642	48.290
27	11.808	12.878	14.573	16.151	18.114	36.741	40.113	43.195	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	37.916	41.337	44.461	48.278	50.994
29	13.121	14.256	16.047	17.708	19.768	39.087	42.557	45.722	49.588	52.335

TABLE 3 CHI-SQUARE DISTRIBUTION (*Continued*)

Degrees of Freedom	Area in Upper Tail									
	.995	.99	.975	.95	.90	.10	.05	.025	.01	.005
30	13.787	14.953	16.791	18.493	20.599	40.256	43.773	46.979	50.892	53.672
35	17.192	18.509	20.569	22.465	24.797	46.059	49.802	53.203	57.342	60.275
40	20.707	22.164	24.433	26.509	29.051	51.805	55.758	59.342	63.691	66.766
45	24.311	25.901	28.366	30.612	33.350	57.505	61.656	65.410	69.957	73.166
50	27.991	29.707	32.357	34.764	37.689	63.167	67.505	71.420	76.154	79.490
55	31.735	33.571	36.398	38.958	42.060	68.796	73.311	77.380	82.292	85.749
60	35.534	37.485	40.482	43.188	46.459	74.397	79.082	83.298	88.379	91.952
65	39.383	41.444	44.603	47.450	50.883	79.973	84.821	89.177	94.422	98.105
70	43.275	45.442	48.758	51.739	55.329	85.527	90.531	95.023	100.425	104.215
75	47.206	49.475	52.942	56.054	59.795	91.061	96.217	100.839	106.393	110.285
80	51.172	53.540	57.153	60.391	64.278	96.578	101.879	106.629	112.329	116.321
85	55.170	57.634	61.389	64.749	68.777	102.079	107.522	112.393	118.236	122.324
90	59.196	61.754	65.647	69.126	73.291	107.565	113.145	118.136	124.116	128.299
95	63.250	65.898	69.925	73.520	77.818	113.038	118.752	123.858	129.973	134.247
100	67.328	70.065	74.222	77.929	82.358	118.498	124.342	129.561	135.807	140.170

t table with right tail probabilities



df\p	0.4	0.25	0.1	0.05	0.025	0.01	0.005	0.0005
1	0.3249	1.0000	3.0777	6.3138	12.7062	31.8205	63.6567	636.6192
2	0.2887	0.8165	1.8856	2.9200	4.3027	6.9646	9.9248	31.5991
3	0.2767	0.7649	1.6377	2.3534	3.1825	4.5407	5.8409	12.9240
4	0.2707	0.7407	1.5332	2.1318	2.7765	3.7470	4.6041	8.6103
5	0.2672	0.7267	1.4759	2.0150	2.5706	3.3649	4.0321	6.8688
6	0.2648	0.7176	1.4398	1.9432	2.4469	3.1427	3.7074	5.9588
7	0.2632	0.7111	1.4149	1.8946	2.3646	2.9980	3.4995	5.4079
8	0.2619	0.7064	1.3968	1.8595	2.3060	2.8965	3.3554	5.0413
9	0.2610	0.7027	1.3830	1.8331	2.2622	2.8214	3.2498	4.7809
10	0.2602	0.6998	1.3722	1.8125	2.2281	2.7638	3.1693	4.5869
11	0.2596	0.6974	1.3634	1.7959	2.2010	2.7181	3.1058	4.4370
12	0.2590	0.6955	1.3562	1.7823	2.1788	2.6810	3.0545	4.3178
13	0.2586	0.6938	1.3502	1.7709	2.1604	2.6503	3.0123	4.2208
14	0.2582	0.6924	1.3450	1.7613	2.1448	2.6245	2.9768	4.1405
15	0.2579	0.6912	1.3406	1.7531	2.1315	2.6025	2.9467	4.0728
16	0.2576	0.6901	1.3368	1.7459	2.1199	2.5835	2.9208	4.0150
17	0.2573	0.6892	1.3334	1.7396	2.1098	2.5669	2.8982	3.9651
18	0.2571	0.6884	1.3304	1.7341	2.1009	2.5524	2.8784	3.9216
19	0.2569	0.6876	1.3277	1.7291	2.0930	2.5395	2.8609	3.8834
20	0.2567	0.6870	1.3253	1.7247	2.0860	2.5280	2.8453	3.8495
21	0.2566	0.6864	1.3232	1.7207	2.0796	2.5177	2.8314	3.8193
22	0.2564	0.6858	1.3212	1.7171	2.0739	2.5083	2.8188	3.7921
23	0.2563	0.6853	1.3195	1.7139	2.0687	2.4999	2.8073	3.7676
24	0.2562	0.6849	1.3178	1.7109	2.0639	2.4922	2.7969	3.7454
25	0.2561	0.6844	1.3163	1.7081	2.0595	2.4851	2.7874	3.7251
26	0.2560	0.6840	1.3150	1.7056	2.0555	2.4786	2.7787	3.7066
27	0.2559	0.6837	1.3137	1.7033	2.0518	2.4727	2.7707	3.6896
28	0.2558	0.6834	1.3125	1.7011	2.0484	2.4671	2.7633	3.6739
29	0.2557	0.6830	1.3114	1.6991	2.0452	2.4620	2.7564	3.6594
30	0.2556	0.6828	1.3104	1.6973	2.0423	2.4573	2.7500	3.6460
inf	0.2533	0.6745	1.2816	1.6449	1.9600	2.3264	2.5758	3.2905

2.0930