Econ 380

Sep 21

Review:

Statistics? collecting, presenting & analysing DATA

why? To draw conclusions; answering questions.

DATA <
- population: collection of all elements of interest "everything".
- sample: Subset of population.

* Measure for population → "parameters"
* Measure for sample → "statistics"

Measures:

1. Mean ($\bar{X}$, μ) → Average $= \dfrac{\Sigma X}{n}$

$\bar{X}$ → sample, μ → population, mu; X → any variable

$= \dfrac{\text{sum of all values}}{\text{No. of all values}}$

Eg: assume the following population DATA:
4, 0, 3, 1, 2 → find Mean?

$$\mu = \frac{4+0+3+1+2}{5} = \frac{10}{5} = 2$$

2. Median: Value in the middle (Data in order)

A → Z, Z → A

Eg: Data set: 4, 0, 3, 1, 2 Find Median?

1. 0, 1, $\boxed{2}$, 3, 4 → 2 is the median

Case 2: Data set: 0, 1, $\boxed{2, 3}$, 4, 5

$$\text{median is} = \frac{2+3}{2} = 2.5$$

3. Variance (S^2 , σ^2)

sigma

Sample — population

التباين

Variance shows the "deviations from the mean".

$$\sigma^2 = \frac{\Sigma(X-M)^2}{n} \quad , \quad S^2 = \frac{\Sigma(X-\bar{X})^2}{n-1}$$

Eg: assume the following Population DATA: 2, 5, 6, 7, 10
Find Variance?

1. $M = \frac{2+5+6+7+10}{5} = \frac{30}{5} = 6$

X	$X-M$	$(X-M)^2$
2	-4	16
5	-1	1
6	0	0
7	1	1
10	4	16
	$\Sigma = 0$	$\Sigma = 34$

$\Sigma = 0$ → always = ZERO

$\Sigma = 34$ → it's the numerator Part

$$\sigma^2 = \frac{34}{5} = 6.8$$

If a case of sample instade of Population?

$$S^2 = \frac{\Sigma(X-\bar{X})^2}{n-1} = \frac{34}{5-1} = 8.5$$

4. Standered Deviation (σ , S) $= \sqrt{\text{Variance}}$

$\sigma = \sqrt{6.8} = 2.6 \qquad S = \sqrt{8.5} = 2.9$

5. Coefficient of Variance

$$CV = \frac{\text{St. Deviation}}{\text{mean}} * 100\%$$ (standered)

$$* CV = \frac{\sigma}{\mu} * 100\% \quad , \quad CV = \frac{S}{\bar{x}} * 100\%$$

→ CV is used when comparing DATA sets with different units

→ The smaller one shows lower variation than the other one.

Eg: assume $S = 2.61$, $\bar{X} = 6$ Find → CV_{sample};

$$CV_{sample} = \frac{2.61}{6} * 100\% = 43.5\%$$

6. Measures of relationship (x, y)

a. Coveriance (S_{xy}, σ_{xy})

b. Coefficient of correlation (r, ρ)

→ a. Coveriance (S_{xy}, σ_{xy})

Coveriance shows if x + y are positively or neg related.

if cov. > 0 → X + Y positively related.

cov. < 0 → X + Y negatively related.

$$* \sigma_{xy} = \frac{\Sigma(x - \mu_x)(y - \mu_y)}{n} \quad , \quad S_{xy} = \frac{\Sigma(x - \bar{x})(y - \bar{y})}{n-1}$$

eg: for the following sample (2, 13), (6, 20), (7, 27). Cov. ?

X	Y	$X - \bar{X}$ (2 - 5)	$Y - \bar{Y}$ (13 - 20)	$(X - \bar{X})(Y - \bar{Y})$
2	13	-3	-7	21
6	20	1	0	0
7	27	2	7	14

$\Sigma = 35$

$$\bar{X} = \frac{15}{3} = 5$$

$$\bar{Y} = \frac{60}{3} = 20$$

$$S_{xy} = \frac{35}{2} = 17.5$$

+ related

b. coefficient correlation (r, p)

(Sample → r, Pop → p)

→ It shows if $x + y$ strongly or weakly related.

$-1 \le r, p \le +1$

strong | $-\frac{1}{2}$ | weak | $\frac{1}{2}$ | strong relation

-1 .... 0 .... $+1$

$\frac{1}{2}$ is the benchmark

$$* \; p = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \quad , \quad r = \frac{S_{xy}}{S_x S_y}$$

Eg: Find coefficient correlation of prev example:

$$S_{xy} = 17.5$$

$$S_x = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}} \quad , \quad S_y = \sqrt{\frac{\Sigma(y-\bar{y})^2}{n-1}}$$

x	y	$x-\bar{x}$	$y-\bar{y}$	$(x-\bar{x})(y-\bar{y})$	$(x-\bar{x})^2$	$(y-\bar{y})^2$
2	13	−3	−7	21	9	49
6	20	1	0	0	1	0
7	27	2	7	14	4	49
				$\Sigma = 35$	$\Sigma = 14$	98

$$S_x = \sqrt{\frac{14}{2}} = 2.65$$

$$S_y = \sqrt{\frac{98}{2}} = 7$$

$$r = \frac{17.5}{(2.65)(7)} = .94$$

Strongly related

Random Variables "not sure about the outcome"

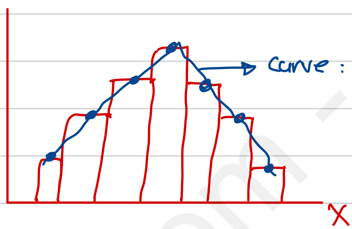
→ its value depends on chance (luck)

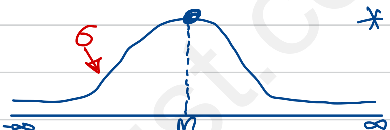
eg: Flip a coin → (H or T), roll a die (1, 2, 3, 4, 5, or 6)

Random variables
- Discrete: takes a countable value
  → easy to count 1 car
- Continuous: takes interval value
  → not easy (range) (5,7) min

this course

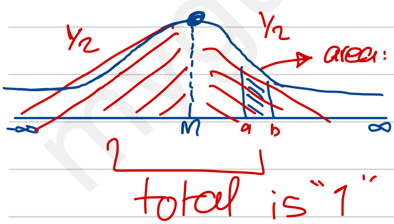
* Graphing "continuous variable" → "Histogram"





* normal distribution curve
* bell shape
* Symmetric Shape

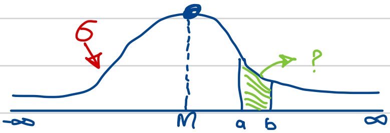
→ Total area under the curve = 1 → 100%



area : Probability : $0 \leq p(a < x < b) \leq 1$

$p(x = \text{value}) = 0$ no p @ a certain point.

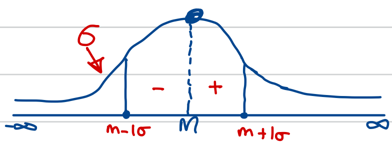
* How can we measure any area under the curve "p"?



By: 1. Empirical Rule : easy but limited.
2. Standered normal distribution "z".

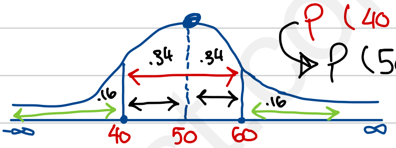
The area under the curve is almost
1. 68% within 1σ around the mean

1σ = .68
2σ = .95
3σ = 1



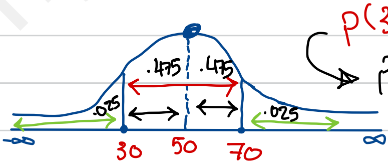
Note: > → bigger than
< → less than

eg: assume X is normally distributed with mean = 50 & St Div σ = 10



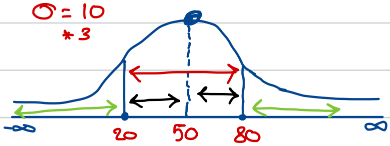
P (40 < X < 60) = .68 "68%" → in case of 1σ value
P (50 < X < 60) = .34 : half of 68%
P(40 < X < 50) = .34 : the other half
P(X > 60) = .16 : remaining of .68 from 1/2
P(X < 40) = .16
→ In total, they're equal to 1 100%.

2. 95% within 2σ around the mean : σ = 10 * 2 → 20



P(30 < X < 70) = .95
P(50 < X < 70) = .475 P (30 < X < 50) = .475
P(X > 70) = .025 : remaing of .95 from 1/2
P (X < 30) = .025

3. 100% within 3σ around the mean



$P(20 < x < 80) = 1$

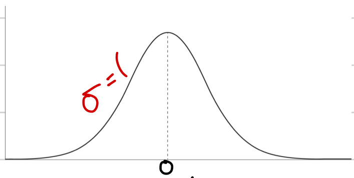
$P(x > 80) = 0$, $P(x < 20) = 0$

Second way in measuring area under the curve is:

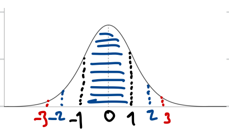
Standard Normal Distribution Z:

→ Z Normally distributed with mean = 0 , $\sigma = 1$

* important to know that...



* if we apply the empirical rule;



$P(-1 < Z < 1) \approx .68$

$P(0 < Z < 1) \approx .34$

$P(-2 < Z < 2) \approx .95$

$P(-3 < Z < 3) \approx 1$

$P(0 < Z < 3) = P(-3 < Z < 0) =$ Zero

Can Z exceed 3 or less than -3 ?

→ According to the emperical rule, the answer "NO" because there's nothing left.

How to find $P(0 < Z < 1.65)$? Z table reports any area under the curve between 0 & any Z value.

Z value → second decimal place of the Z value...

Column Row Row Row Row

Column

Column → X.XXXX

Column

Z = X.XXX

So, the $p(0<z<1.65) = .4505$

$p(0<z<9) = .3413$

→ Z table will give the exact value...

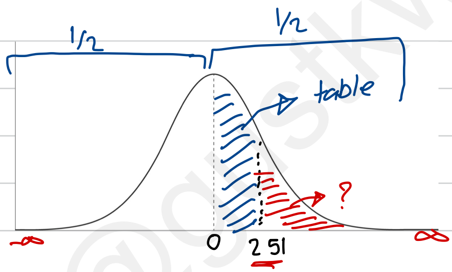
$p(-1.65<z<0) = .4505$ same

What's the $p(z>2.51)$?

$= \frac{1}{2} - p(0<z<2.51)$

$= \frac{1}{2} - .494 = .006$

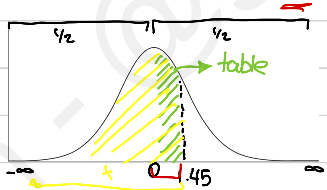
→ $p(z<-2.51)$ = same answer



* $p(z<1.45)$?

$= \frac{1}{2} + p(0<z<1.45)$

→ $\frac{1}{2} + .4265 = .9265$

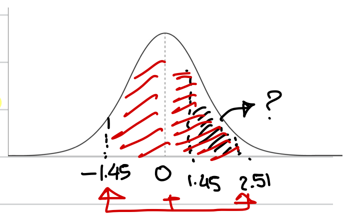


* $p(1.45<z<2.51)$?

$= p(0<z<2.51) - p(0<z<1.45)$

$= .494 - .4265$

$= .0675$



* $p(-1.45<z<2.51)$?

$= p(-1.45<z<0) + p(0<z<2.51)$

$= .4265 + .494 = .9205$

* How can we find Z-score when the area is unknown?

$p(0 < z < z_0) = .475 \rightarrow z_0 = ?$

Go Z Table and search about the value, connect it.

$z_0 = 1.96$

* $p(0 < z < z_0) = .41 \rightarrow z_0 = ?$

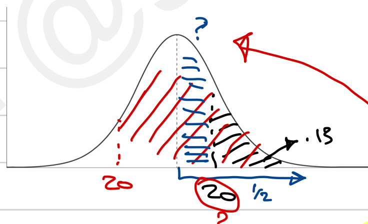
$\rightarrow 1.35$

* $p(z > z_0) = .13 \rightarrow z_0 = ?$ which one?

$p(0 < z < .13) =$

first one reports .37 is

$z_0 = 1.13$



$.5 - .13 = .37$

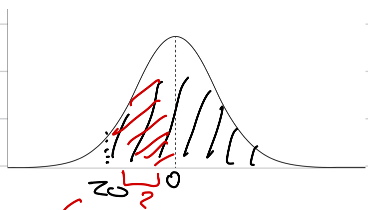
bigger than 1/2

* $p(z > z_0) = .92 \rightarrow z_0 = ?$

First one reports .42 in z table is

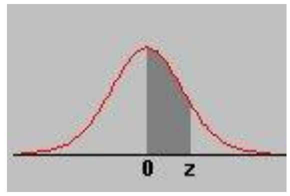
$z_0 = -1.41$

left side of the curve



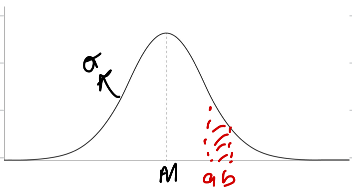
$.92 - .5 = .42$

Area between 0 and z



	0	**0.01**	**0.02**	**0.03**	**0.04**	**0.05**	**0.06**	**0.07**	**0.08**	**0.09**
0	0	0.004	0.008	0.012	0.016	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.091	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.148	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.17	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.195	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.219	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.258	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.291	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.334	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.377	0.379	0.381	0.383
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.398	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.437	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.475	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.483	0.4834	0.4838	0.4842	0.4846	0.485	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.489
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.492	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.494	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.496	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.497	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.498	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.499	0.499

* what if the normally distributed variable is not Z?
How can you find the area under the curve?



if $\sigma = 1$ & $M = 0$ are violated, there's no Z anymore. It's X

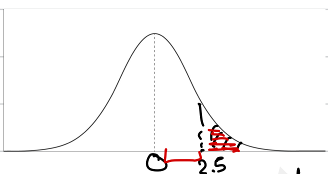
convert X into Z

$P(a<x<b) = ?$

To convert X into Z, $Z = \frac{X - M}{\sigma}$

* if X is normally distributed with $M = 50$, $\sigma = 10$,
what's the $p(X>75)$?

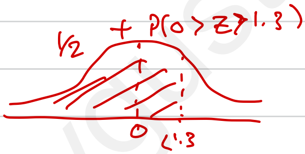
$p(X>75) = p(Z>\frac{75-50}{10}) = p(X>2.5)$.



$= 1/2 - p(0<Z<2.5)$

$= 1/2 - .4938 \quad = .0062$

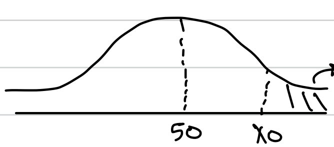
b: $p(x<63) = p(Z<\frac{63-50}{10}) = P(Z<1.3)$



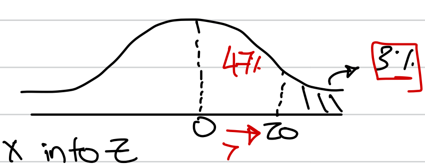
$1/2 + p(0<Z<1.3)$

$= 1/2 + .4032 = .9032$

C: if 3% of the values will receive "Merit" (good points),
what's the Maximum value for the "merit"?



convert X into Z



$P(Z>Z_0) = .03$

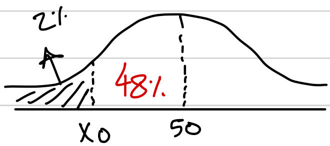
$P(Z > Z_0) = .03 \rightarrow Z_0 = ?$

$\rightarrow P(0 < Z < Z_0) = .47 \rightarrow Z_0 = 1.89$ (table)

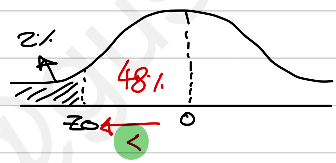
and $Z = \frac{X - M}{\sigma} \rightarrow 1.89 = \frac{X - 50}{10}$

$X - 50 = 1.89 * 10 \rightarrow X = 68.9$ max for merit

d: if 2% of value are to get punished, what's the minimum value to not be punished?



convert X into Zo →



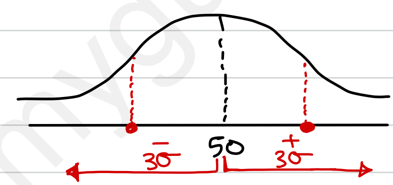
$P(Z < Z_0) = .2$

$P(Z_0 < Z < 0) = .48 \rightarrow Z_0 = -2.06$ (table)

thus $Z = \frac{X - M}{\sigma} \rightarrow -2.06 = \frac{X - 50}{10}$

$X - 50 = -20.6 \rightarrow X = 29.4$ min value to punish

e: what are the max & min values of X?



max $= 50 + 30^{10} = 50 + 30$

$= 80$

min $= 50 - 30^{10} = 50 - 30$

$= 20$

Note: Z is Vaiclated when M_X 0 & σ_X 1; therefore, it would be $\bar{X}$ ins of Z

CHAPTER 7: Sampling distribution of $\bar{X}$

→ prob distribution of $\bar{x}$

eg: assume a population (X) consists of 3 numbers 1, 3, 5
we take a sample of two numbers with replacement. → Put it back

* why replacement? Cuz the population is little & we're trying to increase them... possible samples

Sample	$\bar{X}$
1, 1	1
1, 3	2
1, 5	3
3, 1	2
3, 3	3
3, 5	4
5, 1	3
5, 3	4
5, 5	5

Possible $\bar{X}$ values

→ $\bar{X}$ population

$\bar{X}$: 1, 2, 3, 2, 3, 4, 3, 4, 5 : measures for ($\bar{X}$) Pop...

1. Mean $M_{\bar{x}} = \frac{1+2+3+2+3+4+3+4+5}{9} = \frac{27}{9} = 3$

expected value $\bar{x}$

2. Variance $\sigma^2_{\bar{x}} = \frac{\Sigma(\bar{X}-M_{\bar{x}})^2}{n}$

$\bar{X}$	$\bar{X}-M_{\bar{X}}$	$(\bar{x}-M_{\bar{x}})^2$
1	-2	4
2	-1	1
3	0	0
2	-1	1
3	0	0
4	1	1
3	0	0
4	1	1
5	2	4

$\Sigma = 12$

$= \frac{12}{9} = \frac{4}{3}$

3. St div $\sigma_{\bar{x}} = \sqrt{\sigma^2_{\bar{x}}}$

$= \sqrt{\frac{4}{3}} = \frac{2}{\sqrt{3}}$

* Going back to original population (X) 1, 3, 5

measures:..

1. Mean $(M) = \frac{1+3+5}{3} = 3$

2. Variance $\sigma^2 = \frac{\sum(X-M)^2}{n} = \frac{(1-3)^2+(3-3)^2+(5-3)^2}{3} = \frac{8}{3}$

3. St div $\sigma = \sqrt{\sigma^2} = \sqrt{\frac{8}{3}} = \frac{2\sqrt{2}}{\sqrt{3}}$

* Compare the measures...

1. $M_{\bar{x}} = M$

2. $\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}$: n is sample size

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

St error of $\bar{x}$

assume (x) has mean = 100 with st Div = 15.
if we take many samples of size 5, what's $M_{\bar{x}}, \sigma_{\bar{x}}^2, \sigma_{\bar{x}}$?

$M_{\bar{x}} = M = 100$

$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} = \frac{(15)^2}{5} = 45$

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{15}{\sqrt{5}} = 3\sqrt{5}$

if x is normally distributed, then the $\bar{x}$ will be normally distributed. But if x is not normally distributed, $\bar{x}$ will be almost normally distributed if the sample size is large enough... large enough means $(n > 30)$

size important once X is not normal..

central limit theorem C.L.T

When $\bar{X}$ is normally distributed, then we can find any area under the curve $p(a<\bar{X}<b)$.

→ Convert $\bar{X}$ into Z. $Z_{\bar{X}} = \dfrac{\bar{X} - M_{\bar{X}}}{\sigma_{\bar{X}}}$

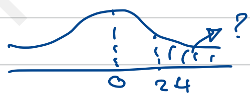
eg: if scores are normally distributed with mean = 1,200 ; St. Div = 60.

A sample of 36 scores is selected, what's the probability that the sample mean will be:

a: large than 1,224
b: less than 1,215
c: between 1190 – 1220
d: between 1205 – 1225
e: exactly 1220

a: $p(\bar{X} > 1{,}224) = p\left(Z > \dfrac{1224 - M}{\left(\frac{\sigma}{\sqrt{n}}\right)}\right) = p\left(Z > \dfrac{1224 - 1200}{\frac{60}{\sqrt{36}}}\right)$

$= p(Z > 2.4)$



$= \frac{1}{2} - p(0 < Z < 2.4)$

$= \frac{1}{2} - .4918$

$= .0082$

b: $p(\bar{X} < 1215) = p\left(Z < \dfrac{1215 - 1200}{\frac{60}{\sqrt{36}} = 10}\right) = p(Z < 1.5)$

$= \frac{1}{2} + p(0 < Z < 1.5) = \frac{1}{2} + .4332$

$= .9332$

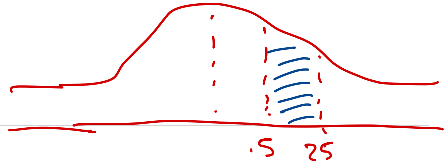
c: $p(1190 < \bar{X} < 1220) = p\left(\dfrac{1190 - 1200}{10} < Z < \dfrac{1220 - 1200}{10}\right)$

$= p(-1 < Z\ 2) = p(0 < Z < 2) + p(-1 < x < 0)$

$= .4772 + .3413 = .8185$.

$0 \to 2 + (-1 - 0)$

d: $p(1{,}205<\bar{x}<1{,}225)$



$= p\left(\frac{1{,}205-1{,}200}{10} < \bar{x} < \frac{1{,}225-1{,}200}{10}\right) = p(.5 < x < 2.5)$

$= p(0 < x < 2.5) - p(0 < x < .5) = .4938 - .1915$

$= .3023$

e: $p(\bar{x} = 1220) = 0$ cuz it's a certain value

* no area @ the point.

CHAPTER 8: Estimation of μ & $\pi(p)$ ::

population porportion %

* Estimation of μ ::

→ Finding the true μ is hard cuz we don't have full info about the population.

→ μ many many samples to be calculated, but in real life we take one sample only.

So, we take one sample & find $\bar{x}$, then use $\bar{x}$ to estimate μ.

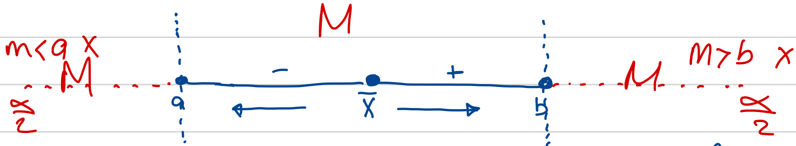
→ Estimation is DONE in two ways ::

1. Point estimation
2. Interval estimation

→ $\mu = \bar{x}$ → point estimation of μ

BUT: Different samples give different $\bar{x}$s

2. Interval estimation: we construct an interval around $\bar{X}$ & believe that it captures M.



→ we create the interval with a certain level of confidence. (confidence level) it goes between 90% – 99%.
→ there are 3 popular levels: 90%, 95%, & 99%.

→ if $M > b \Rightarrow$ mistake } significance level α
$M < a \Rightarrow$ mistake } "alfa"

* If you're 95% confidence, then α is 5%. So, it's always the rest of 100%.
→ Range of α: 1% → 10%: 3 popular levels: 1%, 5%, & 10%

* How can we create the interval?

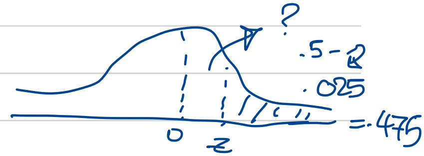
$$M = \bar{X} \pm \text{margin of error}$$

a: σ is known

$$M = \bar{X} \pm (\text{critical value})(\text{St. error of } \bar{X})$$

$$M = \bar{X} \pm Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$$

eg: $\alpha = 5\% \Rightarrow \alpha/2 = .025\ (2.5\%)$



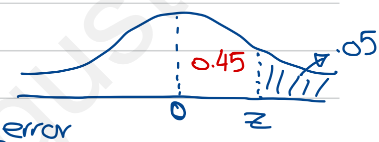
$Z = 1.96$

eg: Create a 90% confidence interval for the avarage of all college students in Kuwait, when a sample of 100 students shows an avarage age = 19.5 yrs (st Div of all ages of all students is 15 yrs.

→ $n = 100$, $\bar{X} = 19.5$, $\sigma = 15$

$M = \bar{x} \pm Z_{\frac{\alpha}{2}} \left(\frac{\sigma}{\sqrt{n}}\right)$: α is 10%, so $\frac{\alpha}{2}$ is 5% .05

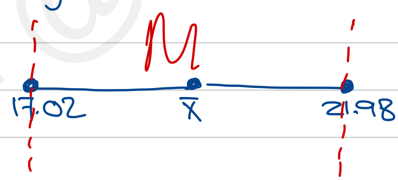
in Z table, 0.45 is 1.65



$M = 19.5 \pm 1.65\left(\frac{15}{\sqrt{100}}\right)$ → st error of $\bar{X}$

$= 19.5 \pm 2.48$ → margin of error

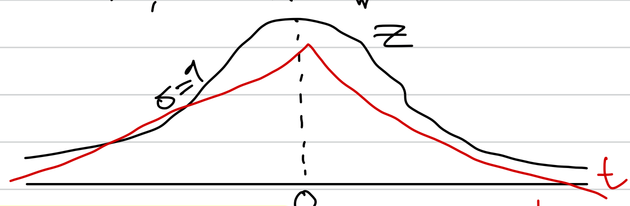
$a = 19.5 - 2.48 = 17.02$
$b = 19.5 + 2.48 = 21.98$



So, the avarage age of all college students in KWT is between (17.02 – 21.98) with a confidence of 90%.

b: When σ might not be known ! we can't use Z table in this case... Instead, a new distribution (t) will be used. t-distribution is a normally distributed with mean but variance > 1 (variance depends on n)

→ when n ↑, variance ↓



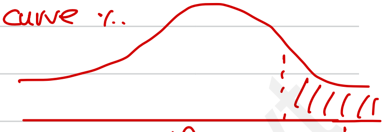
when $n \to \infty$: variance ≈ 1 → $Z \approx t$ "C.L.T"

$n > 30 \to Z \approx t$

* We use t table to find t-score according to area under the curve.

t-table   area on the right side under the curve %..

df ↓   x x x   x x x   x x x
x x x
x x x



$df = n-1$

* Interval Estimation when σ is unknown

$$\mu = \bar{x} \pm \text{margin of error}$$
$$= \bar{x} \pm t_{\frac{\alpha}{2}, df} \cdot \left(\frac{s}{\sqrt{n}}\right) \rightarrow \text{St error of } \bar{x}$$

eg: Create a 99% confidence interval for the average age of the city when a sample of 25 shows an average age = 29 with St. devi = 10 ^{s}

$$\mu = \bar{x} \pm t_{\frac{\alpha}{2}, df} \cdot \frac{s}{\sqrt{n}}$$

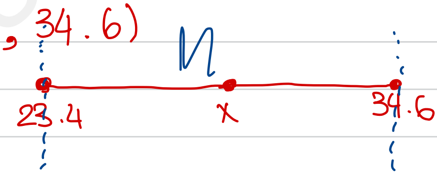
: $\alpha = 1\% \rightarrow \frac{\alpha}{2} = .005$
$df = n-1 = 25-1 = 24$ } t tab
$\rightarrow 2.7969 \approx 2.8$

$$\mu = 29 \pm (2.8) \cdot \left(\frac{10}{\sqrt{25}}\right)$$

$\rightarrow = 2$, st error of $\bar{x}$

$= 29 \pm 5.6$ margin of error

$(23.4, 34.6)$



* Sample size needed for certain margin of error (M.E). (n)

$$n = \left[\frac{Z_{\frac{\alpha}{2}} \sigma}{M.E}\right]^2$$

eg: What's the sample size needed to have a margin of error $= \pm 10$ when $\sigma = 30$ & $\alpha = 10\%$

$$\alpha = 10\% \quad : \frac{\alpha}{2} = 5\% \rightarrow Z_{\frac{\alpha}{2}} = 1.65$$

$$n = \left[\frac{(1.65)(30)}{10}\right]^2 = 24.5025 \approx 25$$ always round to the next whole number...

* Estimation of π (p) population proportion: the size of a certain group within the population.

$$\pi = \frac{X}{N}$$

But, it's hard to find the true π, cuz we don't have full information about the whole population

So, we take a sample & find sample proportion ($\hat{p}$), then use $\hat{p}$ to estimate π.

Estimation is :-

a: point estimation: $\pi = \hat{p}$ → point est of π

b: Interval estimation: $\pi = \hat{p} \pm$ margin of error

$$= \hat{p} \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$ → St error of $\hat{p}$

eg: Create a 90% confidence interval for the % of black cars in the city when a sample of 1000 cars shows 200 are black.

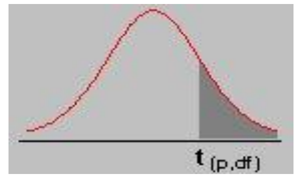
$$\hat{p} = \frac{200}{1000} = .2 : 20\%$$

$$\alpha = 10\% : \frac{\alpha}{2} = 5\%$$

$$Z = 1.65$$

$$\pi = .2 \pm (1.65) \left[\sqrt{\frac{.2(.8)}{1000}}\right] = .012 \text{ St error of } \hat{p}$$

$$\pi = .2 \pm .02 \Rightarrow (.18, .22)$$



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

Gulf University for Science & Technology
Department of Economics & Finance
ECO-380: Business Statistics

Dr. Khalid Kisswani
Room: N1-116
Tel: 2530-7339
kisswani.k@gust.edu.kw

Assignment 1(LO i)

1. The hourly wages of a sample of 130 system analysts are given below.

n

mean = 60 range = 20
mode = 73 variance = 324
median = 74

The coefficient of variation equals

$\frac{\sigma}{M} * 100 = \frac{18}{60} * 100\% = 30\%$

a. 0.30%
b. 30%
c. 5.4%
d. 54%

2. The variance of a sample of 169 observations equals 576. The standard deviation of the sample equals

n σ^2

a. 13
b. 24
c. 576
d. 28,461

$\sqrt{576}$

3. The standard deviation of a sample of 100 observations equals 64. The variance of the sample equals

n s

a. 8
b. 10
c. 6400
d. 4,096

4. Which of the following symbols represents the mean of the population?
a. σ^2
b. σ
c. μ
d. $\bar{x}$

5. Which of the following symbols represents the variance of the population?
a. σ^2
b. σ
c. μ
d. $\bar{x}$

6. The coefficient of correlation ranges between
a. 0 and 1
b. -1 and +1
c. minus infinity and plus infinity
d. 1 and 100

7. Given the following information:

Standard deviation = 8
Coefficient of variation = 64%

The mean would then be
a. 12.5
b. 8
c. 0.64
d. 1.25

$.64 = \frac{8}{M}$

$\frac{8}{.64} = \frac{.64M}{.64}$

8. The standard deviation of a sample was reported to be 20 (5). The report indicated that $\sum(x - \bar{x})^2 = 7200$. What has been the sample size?
a. 16
b. 17
c. 18
d. 19

20 − 1

Exhibit 3-2

A researcher has collected the following sample data

5	12	6	8	5
6	7	5	12	4

4 5 5 5 6 6 7 8 12 12

9. Refer to Exhibit 3-2. The median is
a. 5
b. 6
c. 7
d. 8

$\frac{12}{2} =$

10. Refer to Exhibit 3-2. The mean is
a. 5
b. 6
c. 7
d. 8

$\frac{70}{10}$

11. The probability that a continuous random variable takes any specific value = 0
a. is equal to zero
b. is at least 0.5
c. depends on the probability density function
d. is very close to 1.0

12. A normal distribution with a mean of 0 and a standard deviation of 1 is called
a. a probability density function
b. an ordinary normal curve
c. a standard normal distribution
d. None of these alternatives is correct.

bill shape, symitric

normal distribution curve

$\sigma = 1$, $M = 0$

13. In a standard normal distribution, the probability that Z is greater than zero is
a. 0.5
b. equal to 1
c. at least 0.5
d. 1.96

$P(Z > 0)$

.5 .5

14. The random variable x is known to be normally distributed. The probability of x having a value equals 80 or 95 is
a. 0.75
b. 0
c. 1
d. Can't be found, because there isn't enough information

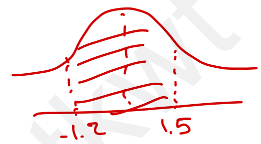
Zero at a Certain Point

15. Z is a standard normal random variable. The $P(-1.96 \le Z \le -1.4)$ equals
a. 0.8942
b. 0.0558
c. 0.475
d. 0.4192

$P(0<Z<-1.96) - P(0<Z<-1.4)$
$= .475 - .4192$

16. Z is a standard normal random variable. The $P(-1.20 \le Z \le 1.50)$ equals
a. 0.0483
b. 0.3849
c. 0.4332
d. 0.8181

$P(0<Z<1.5) + P(-1.2<Z<0)$



17. Given that Z is a standard normal random variable, what is the probability that $-2.51 \le Z \le -1.53$?
a. 0.4950
b. 0.4370
c. 0.0570
d. 0.9310

$P(-2.51<Z<0) - P(-1.53<Z<0)$
$=$

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

$1/2 - P(-2.12<Z<0)$

19. X is a normally distributed random variable with a mean of 8 (μ) and a standard deviation of 4 (σ). The probability that X is between 1.48 and 15.56 is
a. 0.0222
b. 0.4190
c. 0.5222
d. 0.9190

$\frac{15.56-8}{4} = 1.89$, $\frac{1.48-8}{4} = -1.63$
$P(0<Z<1.89) + (-1.63<Z<0)$
$=$

20. X is a normally distributed random variable with a mean of 5 and a variance of 4. The probability that X is greater than 10.52 is
a. 0.0029
b. 0.0838
c. 0.4971 ×
d. 0.9971 ×

$\sqrt{4} = 2\ \sigma$
$\frac{10.52-5}{2} = 2.76$
$1/2 - P(0<Z<2.76)$

21. Given that Z is a standard normal random variable, what is the value of Z if the area to the left of Z is 0.0559?
a. 0.4441
b. 1.59
c. 0.0000
d. 1.50

$1/2 + .0559 = .4441$ → area under curve

next go to z table find .4441, it's 1.59

Gulf University for Science & Technology
Department of Economics & Finance
ECO-380: Business Statistics

Dr. Khalid Kisswani
Room: N1-116
Tel: 2530-7339
kisswani.k@gust.edu.kw

Assignment 2 (LO i)

1. A simple random sample of 100 observations was taken from a large population. The sample mean and the standard deviation were determined to be 80 and 12 respectively. The standard error of the mean is
a. 1.20
b. 0.12
c. 8.00
d. 0.80

$$\frac{\sigma}{\sqrt{n}} = \frac{12}{\sqrt{100}} = 1.2$$

ANS: A

2. A population has a standard deviation of 16. If a sample of size 64 is selected from this population, what is the probability that the sample mean will be within ±2 of the population mean?
a. 0.6826
b. 0.3413
c. -0.6826
d. Since the mean is not given, there is no answer to this question.

ANS: A

3. As the sample size increases, the
a. standard deviation of the population decreases
b. population mean increases
c. standard error of the mean decreases
d. standard error of the mean increases

ANS: C

4. The probability distribution of the sample mean is called the
a. central probability distribution
b. sampling distribution of the mean
c. random variation
d. standard error

ANS: B

5. A population has a mean of 75 and a standard deviation of 8. A random sample of 800 is selected. The expected value of $\bar{x}$ is

$$\mu = \mu_{\bar{x}}$$

a. 8
b. 75
c. 800
d. None of these alternatives is correct.

ANS: B

6. From a population of 200 elements, a sample of 49 elements is selected. It is determined that the sample mean is 56 and the sample standard deviation is 14. The standard error of the mean is
a. 3
b. 2
c. greater than 2

$$\frac{s}{\sqrt{n}} = \frac{14}{\sqrt{49}} = 2$$

d. less than 2

ANS: B

7. A population has a mean of 300 and a standard deviation of 18. A sample of 144 observations will be taken. The probability that the sample mean will be between 297 to 303 is

a. 0.4332
b. 0.9544
c. 0.9332
d. 0.0668

$P(0 < Z < .167) + (-.167 < Z < 0)$

$= .4772 + .4772$

$\bar{X} \to Z = \frac{x - M}{\sigma} : \frac{303-300}{\frac{18}{\sqrt{144}}} = 2$

$\frac{297-300}{\frac{18}{\sqrt{144}}} = -2$

ANS: B

8. A simple random sample of 64 observations was taken from a large population. The sample mean and the standard deviation were determined to be 320 and 120 respectively. The standard error of the mean is

a. 1.875
b. 40
c. 5
d. 15

$\frac{120}{\sqrt{64}} =$

ANS: D

9. Random samples of size 81 are taken from an infinite population whose mean and standard deviation are 200 and 18, respectively. The distribution of the population is unknown. The mean and the standard error of the mean are

a. 200 and 18
b. 81 and 18
c. 9 and 2
d. 200 and 2

$M = M_{\bar{x}} = 200$

$\sigma_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{18}{\sqrt{81}} = 2$

ANS: D

10. A population has a mean of 80 and a standard deviation of 7. A sample of 49 observations will be taken. The probability that the sample mean will be larger than 82 is

a. 0.5228
b. 0.9772
c. 0.4772
d. 0.0228

$\frac{82-80}{\frac{7}{\sqrt{49}}} = 2 \rightarrow .4772$

$.5 - .4772$

ANS: D

11. A population has a mean of 180 and a standard deviation of 24. A sample of 64 observations will be taken. The probability that the sample mean will be between 183 and 186 is

a. 0.1359
b. 0.8185
c. 0.3413
d. 0.4772

$.4772 - .3413$

$= \frac{183-180}{\frac{24}{\sqrt{64}}} = 1 : = .3413$

$, \frac{186-180}{\frac{24}{\sqrt{64}}} = 2 \quad .4772$

ANS: A

12. Random samples of size 49 are taken from a population that has 200 elements, a mean of 180, and a variance of 196. The distribution of the population is unknown. The mean and the standard error of the mean are

$s = \sqrt{196} = 14$

a. 180 and 24.39
b. 180 and 28
c. 180 and 2.5
d. 180 and 2

$M = M_{\bar{X}} = 180$

$\sigma_{\bar{X}} = \frac{s}{\sqrt{n}} = \frac{14}{\sqrt{49}} = 2$

ANS: D

μ σ n

13. A population has a mean of 84 and a standard deviation of 12. A sample of 36 observations will be taken. The probability that the sample mean will be between 80.54 and 88.9 is

a. 0.0347
b. 0.7200
c. 0.9511
d. 8.3600

$$\frac{4.9}{\frac{12}{\sqrt{36}}} = 2.45 \; ; \; \frac{-3.46}{\frac{12}{\sqrt{36}}} = -1.73$$

$$P(0<z<2.45) + P(-1.73<z<0)$$

ANS: **C**

14. A population has a mean of 53 and a standard deviation of 21. A sample of 49 observations will be taken. The probability that the sample mean will be greater than 57.95 is

a. 0
b. .0495
c. .4505
d. .9505

ANS: **B**

Gulf University for Science & Technology
Department of Economics & Finance
ECO-380: Business Statistics

Dr. Khalid Kisswani
Room: N1-116
Tel: 2530-7339
kisswani.k@gust.edu.kw

Assignment 3 (LO ii)

1. When s is used to estimate σ, the margin of error is computed by using

a. normal distribution
b. t distribution
c. the mean of the sample
d. the mean of the population

ANS: B

$\sigma = \sqrt{900} = 30$

2. From a population with a variance of 900, a sample of 225 items is selected. At 95% confidence, the margin of error is

a. 15
b. 2
c. 3.92
d. 4

$\bar{X} \pm Z_{\alpha/2} \cdot \left(\frac{\sigma}{\sqrt{n}}\right)$ ← margin of error

$\alpha = 5\%$ $\alpha/2 = .025$

$Z = .475$ → 1.96

$1.96 \cdot \frac{30}{\sqrt{225}} =$

ANS: C

3. A population has a standard deviation of 50. A random sample of 100 items from this population is selected. The sample mean is determined to be 600. At 95% confidence, the margin of error is

a. 5
b. 9.8
c. 650
d. 609.8

$1.96 \cdot \frac{50}{\sqrt{100}} = 9.8$

ANS: B

4. In order to determine an interval for the mean of a population with unknown standard deviation a sample of 61 items is selected. The mean of the sample is determined to be 23. The number of degrees of freedom for reading the t value is

a. 22
b. 23
c. 60
d. 61

$df = n - 1 = 61 - 1 =$

ANS: C

5. The value added and subtracted from a point estimate in order to develop an interval estimate of the population parameter is known as the

a. confidence level
b. margin of error
c. parameter estimate
d. interval estimate

$\mu = \bar{x} \pm$ margin of error

ANS: B

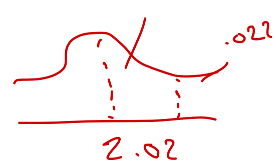
6. The z value for a 97.8% confidence interval estimation is

a. 2.02
b. 1.96

$2.2\% = .022$

$\alpha = .022 \rightarrow \alpha/2 = .011$

$.5 - .011 = .489 \rightarrow$



c. 2.00
d. 2.29

ANS: D

7. The t value for a 95% confidence interval estimation with 24 degrees of freedom is

$\alpha = .05 : \frac{\alpha}{2} = .025$

a. 1.711
b. 2.064
c. 2.492
d. 2.069

$= 2.064$

ANS: B

8. As the sample size increases, the margin of error

Variance ↓

a. increases
b. decreases
c. stays the same
d. increases or decreases depending on the size of the mean

ANS: B

n σ

9. A sample of 225 elements from a population with a standard deviation of 75 is selected. The sample mean is
M 180. The 95% confidence interval for μ is

$\alpha = .05 \quad \frac{\alpha}{2} = .025$

a. 105.0 to 225.0
b. 175.0 to 185.0
c. 100.0 to 200.0
d. 170.2 to 189.8

$z = .475 : 1.96$

$* \frac{75}{\sqrt{225}} = 9.8$

ANS: D

$\sigma = 44$ n

10. It is known that the variance of a population equals $\sqrt{1{,}936}$. A random sample of 121 has been taken from the population. There is a .95 probability that the sample mean will provide a margin of error of

$\frac{\alpha}{2} = .025$

$1.96 \cdot \frac{44}{\sqrt{121}}$

a. 7.84
b. 31.36
c. 344.96
d. 1,936

ANS: A

n $\bar{x}$

11. A random sample of 144 observations has a mean of 20, a median of 21, and a mode of 22. The population standard deviation is known to equal 4.8. The 95.44% confidence interval for the population mean is

σ

a. 15.2 to 24.8
b. 19.200 to 20.800
c. 19.216 to 20.784
d. 21.2 to 22.8

$\alpha = 4.56 : \frac{\alpha}{2} = 2.28 : .4887$

ANS: B

12. The sample size needed to provide a margin of error of 2 or less with a .95 probability when the population standard deviation equals 11 is
a. 10
b. 11
c. 116
d. 117

ANS: D

13. It is known that the population variance equals 484. With a 0.95 probability, the sample size that needs to be taken if the desired margin of error is 5 or less is

a. 25
b. 74
c. 189
d. 75

ANS: D

14. The following random sample from a population whose values were normally distributed was collected.

10 12 18 16

The 80% confidence interval for μ is

a. 12.054 to 15.946
b. 10.108 to 17.892
c. 10.321 to 17.679
d. 11.009 to 16.991

ANS: D

15. In a random sample of 144 observations, $\bar{p}$ = 0.6. The 95% confidence interval for P is

a. 0.52 to 0.68
b. 0.144 to 0.200
c. 0.60 to 0.70
d. 0.50 to 0.70

ANS: A

16. In a random sample of 100 observations, $\bar{p}$ = 0.2. The 95.44% confidence interval for P is

a. 0.122 to 0.278
b. 0.164 to 0.236
c. 0.134 to 0.266
d. 0.120 to 0.280

ANS: D

Exhibit 8-1

In order to estimate the average time spent on the computer terminals per student at a local university, data were collected for a sample of 81 business students over a one-week period. Assume the population standard deviation is 1.8 hours.

17. Refer to Exhibit 8-1. The standard error of the mean is

a. 7.50
b. 0.39
c. 2.00
d. 0.20

ANS: D

18. Refer to Exhibit 8-1. With a 0.95 probability, the margin of error is approximately

a. 0.39
b. 1.96
c. 0.20
d. 1.64

ANS: A

19. Refer to Exhibit 8-1. If the sample mean is 9 hours, then the 95% confidence interval is

a. 7.04 to 110.96 hours
b. 7.36 to 10.64 hours
c. 7.80 to 10.20 hours
d. 8.61 to 9.39 hours

ANS: D

CHAPTER: 9: testing hypotheses → one mean μ important

means: we have two arguments (hypotheses):

1. null hypotheses (H_0): represents the general pointview & beleived to be true

2. Alternative Hypothesis (H_1, H_a): Challenging statement, & disagrees with null.

So, we need to test to validate the hypo: this is done by taking a sample & running a test to reach "Decision".

accept H_0 reject H_0

* any decision made can be good or bad: so, we have 2 good vs 2 bad:

* Good ones:
  1. accept the null H_0 when it's true
  2. reject the null H_0 when it's false

* Bad decisions:
  1. accept the null H_0 when it's false
  2. reject the null H_0 when it's true

Decision Table:

Decision	H_0 true	H_0 false
Accept H_0	Good decision	type I error prob(type I) = β
Reject H_0	Type I error prob(type I) = α	Good decision

α is more bad "we're trying to Reject"

How can we run the test?

→ we use one of two methods (approaches):

1. Critical value approach (4 steps)
2. P. value approach

$\boxed{1}$ Critical value Approach (4-Steps)

1. State the hypotheses: write the correct H_0 & H_1

→ $H_0: \mu = \mu_0$; $H_1: \mu > \mu_0$ → right-sided "upper tail"

or $H_1: \mu < \mu_0$ → left-sided "lower tail"

or $H_1: \mu \neq \mu_0$ → two-sided "two tail"

2. Find "test statistics" → an indicator found using sample data.

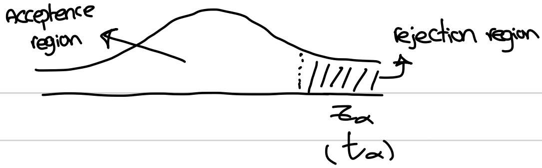
a. σ known: Z statistics $= \dfrac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$

b. σ unknown: t statistics $= \dfrac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$

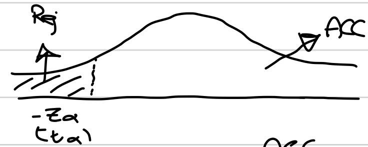
3. Create "Critical region"

↳ Show within the Z or t graph in which part you can accept the null & in which part you can reject the null (depends on: α, type of test)

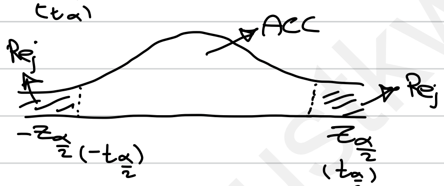
a. right - sided $>$



b. left - sided $<$



c. two - sided $\neq$



4. Make your "decision"

→ Check where is the "test statistic" going to fall within the critical region.

$H_1: M > 12$

eg: test if the population mean is bigger than 12 when a sample of 25 (n) gave an average $\bar{x} = 14$, knowing that st div of pop $\sigma = 4.32$ ($\alpha = .05 : 5\%$). Apply the 4 steps:.

z star

1. $H_0 : M = 12$

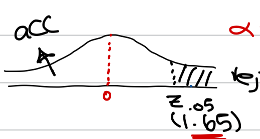
   $H_1 : M > 12$

2. test stat

   Z stat since σ is known $= \dfrac{\bar{X} - M_0}{\frac{\sigma}{\sqrt{n}}} = \dfrac{14 - 12}{\frac{4.32}{\sqrt{25}}}$

   $= 2.31$

3. Critical Rejon



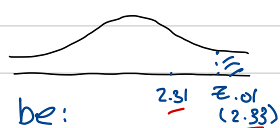
$\alpha = .05$

↳ Zval $= .5 - .05 = .45$

Z table

4. Decision : Reject Ho at $\alpha = 5\%$ since 2.31 is > 1.65 where it falls under the rej side

In case $\alpha = 1\%$:



→ Decision would be:

accept the Ho at $\alpha = 1\%$.

when it's > or < → check $\bar{x}$

$\bar{x}$ > Ho prove more

eg: test if the population mean is $\boxed{\text{different}}$ ($\neq$) from 18 when a sample of 49 (n) items shows an avarege = 17 ($\bar{x}$) with st. Div = 4.5 (s) (α = 5%).

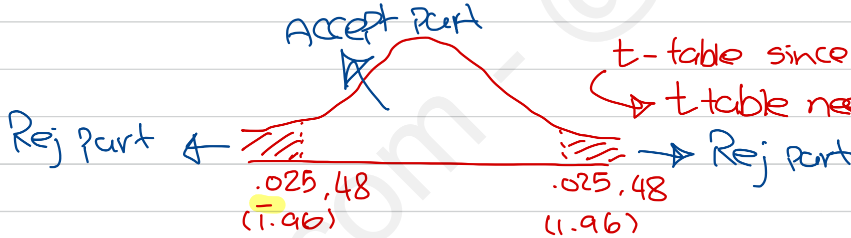
1. $H_0: \mu = 18$

   $H_1: \mu \neq 18$

2. test "statistics" t-stat since σ is unknown

$$\text{t-stat} = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}} = \frac{17 - 18}{\frac{4.5}{\sqrt{49}}} = -1.55$$

3. Critical reg $\alpha = .05 : 5\% \rightarrow \frac{\alpha}{2}$ on the right, $\frac{\alpha}{2}$ the other half.



t-table since σ is unknown

→ t table needs df = 49-1

= 48

↓

∞ in t-Tab

4. Decision

Accept Ho at α = 5% since it falls between (-1.96, 1.96)

eg: test if the population mean is less than 45 or no using a sample of 36 items that shows an average = 43 ($\bar{x}$), knowing that σ = 4.6 (α = 5%)

sample shows 43

σ is known → Z distribution

1. State the hypo

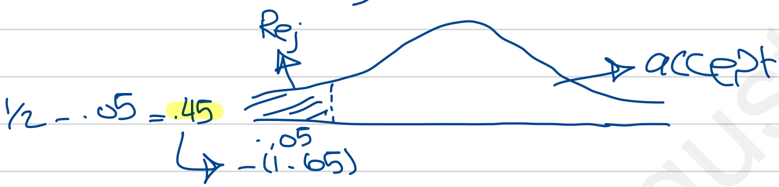
$H_0: \mu = 45$

$H_1: \mu < 45$

2. test Stat

$$Z = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} = \frac{43 - 45}{\frac{46}{\sqrt{36}}} = -2.61$$

3. Critical reg $\mu < 43$ → left sided



4. Decision

-2.61

→ Reject the "Ho" since it falls under Rej part...

* what if $\alpha = 10\%$ in this case? reject without thinking... cuz when you reject at a certain level of α, you reject all other high levels; however, when it's less than the certain α, you need to test it again...

2nd Approach (P-value) important
→ probability

P-value → an area under the curve is the rejection area according to "test stastics". we will calculate the rej part...

Decision: convert Pvalue into α (1%; 5%; 10%)

→ P-value > α : accept Ho

→ P-value ≤ α : reject Ho

finding & calculating the p-vale: it depend on σ if it known or unknown

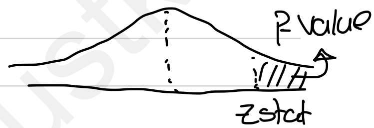
1. σ is known

p-value is found using z statistics

p value will be an area within the z distribution

a. right-side:

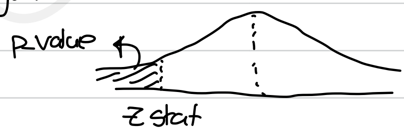
z stat has to be positive (+)



$P(z > z\,stat)$

b. left-side:

z stat has to be negative (−)

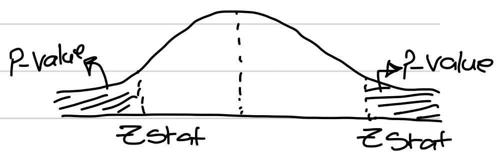


$p(z < z\,stat)$

c. two-side:

z stat can be + or −, depends on the sample...

$P\text{-value} = 2\,(P(z > z\,stat))$



2. σ is unknown

p-value depends on t-statistics

found on "interval"

will be found from "t-table"

it will use: t stat & df (n-1)

Note: in case this is two sided test, we multiply the interval by 2.

eg: assume the following test:

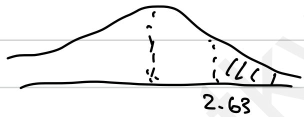
$H_0: M = 50$

$H_1: M \neq 50$

Z statistics: 2.63

find the p-value & write the decision...

$$\text{P-value} = 2(1/2 - P(Z > 2.63))$$



$$= 2(1/2 - .4957) = .0086 \rightarrow .86\%$$

Decision:

P-value < $\alpha = 1\%$ → Reject Ho at $\alpha = 1\%$.

eg: assume the following test

$H_0: M = 500$

$H_1: M > 500$

t-stat = 1.46, n = 20

find the P-value & write the Decision

t-table df = 19

P-value = between .05 & .1

→ within this range p-value < $\alpha = 10\%$

→ reject the Ho at $\alpha = 10\%$.

Gulf University for Science & Technology
Department of Economics & Finance
ECO-380: Business Statistics

Dr. Khalid Kisswani
Room: N1-116
Tel: 2530-7339
kisswani.k@gust.edu.kw

Assignment 4 (LO iii)

1. The probability of committing a Type I error when the null hypothesis is true is
a. the confidence level x
b. β → Type II
c. greater than 1 x
d. the Level of Significance α

2. The *p*-value is a probability that measures the support (or lack of support) for the
a. null hypothesis
b. alternative hypothesis
c. either the null or the alternative hypothesis
d. sample statistic

accept the
Ho when it's x

3. A Type II error is committed when
a. a true alternative hypothesis is mistakenly rejected
b. a true null hypothesis is mistakenly rejected
c. the sample size has been too small
d. a false null hypothesis is mistakenly accepted

4. The probability of making a Type I error is denoted by
a. α
b. β
c. $1 - \alpha$
d. $1 - \beta$

5. The probability of making a Type II error is denoted by
a. α
b. β
c. $1 - \alpha$
d. $1 - \beta$

6. When the following hypotheses are being tested at a level of significance of α

$$H_0: \mu \geq 500$$
$$H_a: \mu < 500$$

the null hypothesis will be rejected if the *p*-value is
a. $\leq \alpha$
b. $> \alpha$
c. $> \alpha/2$
d. $\leq 1 - \alpha/2$

P.V > α → accept
P.V ≤ α → reject.

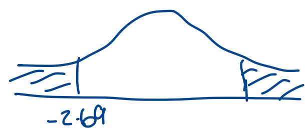
7. In order to test the following hypotheses at an α level of significance

$$H_0: \mu \leq 800$$
$$H_a: \mu > 800$$

the null hypothesis will be rejected if the test statistic Z is
a. $\geq Z_\alpha$
b. $< Z_\alpha$
c. $< -Z_\alpha$
d. $= \alpha$

8. For a lower bounds one-tailed test, the test statistic z is determined to be zero. The p-value for this test is
a. zero
b. -0.5
c. +0.5
d. 1.00

9. In a two-tailed hypothesis test situation, the test statistic is determined to be $z = -2.69$. The sample size has been 45. The p-value for this test is
a. 0.0036
b. +0.005
c. -0.01
d. +0.0072



$\frac{1}{2} - P(-2.69 < Z < 0)$
$\frac{1}{2} - .4 = .0036$
$\rightarrow 2 * .0036 =$

10. In a lower one-tail hypothesis test situation, the p-value is determined to be 0.22. If the sample size for this test is 51, the z statistic has a value of
a. 0.78
b. -0.78
c. 0.59
d. -0.59

$\frac{1}{2} - .22 = .28$
$\Rightarrow .78$

11. If a hypothesis is rejected at the 5% level of significance, it
a. will always be rejected at the 1% level ✗
b. will always be accepted at the 1% level ✗
c. will never be tested at the 1% level ✗
d. may be rejected or not rejected at the 1% level ✓

rej at all higher levels

12. For a one-tailed test (lower tail) at 93.7% confidence, Z =
a. -1.86
b. -1.53
c. -1.96
d. -1.645

$\alpha = 6.3\%$

$= \frac{1}{2} - .063 = .437$
$= 1.53$

13. Read the Z statistic from the normal distribution table and circle the correct answer. A one-tailed test (upper tail) at 87.7% confidence; Z =
a. 1.54
b. 1.96
c. 1.645
d. 1.16

$\alpha = 12.3\%$
$.5 - .123 = .377$

14. In a two-tailed hypothesis test the test statistic is determined to be $Z = -2.5$. The p-value for this test is
a. -1.25
b. 0.4938
c. 0.0062
d. 0.0124

$.5 - P(-2.5 < Z < 0)$
$.5 - .4938 = .0062 * 2 =$

15. In a one-tailed hypothesis test (lower tail) the test z-statistic is determined to be -2. The p-value for this test is
a. 0.4772
b. 0.0228
c. 0.0056
d. 0.5228

$\frac{1}{2} - P(-2 < Z < 0)$
$\frac{1}{2} - .4772 =$

Exhibit 9-1

$n = 36$ $\bar{x} = 24.6$ $S = 12$ $H_0: \mu \leq 20$
$H_a: \mu > 20$

$$t\text{-stat} = \frac{24.6 - 20}{\frac{12}{\sqrt{36}}} = 2.3$$

16. Refer to Exhibit 9-1. The test statistic is
a. 2.3 b. 0.38
c. -2.3 d. -0.38

17. Refer to Exhibit 9-1. The *p*-value is between
a. 0.005 to 0.01
b. 0.01 to 0.025
c. 0.025 to 0.05
d. 0.05 to 0.10

df = 35 'of P , 2.3

1%, 2.5%

18. Refer to Exhibit 9-1. 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.

$\alpha = .05$ X

Exhibit 9-4

The manager of a grocery store has taken a random sample of 100 customers. The average length of time it took the customers in the sample to check out was 3.1 ($\bar{x}$) minutes with a standard deviation of 0.5 (s) minutes. We want to test to determine whether or not the mean waiting time of all customers is significantly more than 3 minutes.

19. Refer to Exhibit 9-4. The test statistic is
a. 1.96
b. 1.64
c. 2.00
d. 0.056

$$t = \frac{3.1-3}{\frac{.5}{\sqrt{100}}} =$$

$H_0 = \mu = 3.1$

$H_1 = \mu > 3$

20. Refer to Exhibit 9-4. The *p*-value is between
a. .005 to .01
b. .01 to .025
c. .025 to .05
d. .05 to .10

21. Refer to Exhibit 9-4. At 95% confidence, it can be concluded that the mean of the population is
a. significantly greater than 3
b. not significantly greater than 3
c. significantly less than 3
d. significantly greater then 3.18

reject, ↑ 3

Exhibit 9-8

The average gasoline price of one of the major oil companies in Europe has been $1.25 ($\mu$) per liter. 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 49 of their gas stations is selected and the average price is determined to be $1.20 ($\bar{x}$) per liter. Furthermore, assume that the standard deviation of the population (σ) is $0.14.

$$Z = \frac{1.2-1.25}{\frac{.14}{\sqrt{49}}} = -2.5$$

St error of σ

22. Refer to Exhibit 9-8. The standard error has a value of
a. 0.14
b. 7
c. 2.5
d. 0.02

23. Refer to Exhibit 9-8. The value of the test statistic for this hypothesis test is
a. 1.96
b. 1.645
c. -2.5
d. -1.645

24. Refer to Exhibit 9-8. The *p*-value for this problem is
a. 0.4938
b. 0.0062
c. 0.0124
d. 0.05

$.5 - .4938 = .0062 * 100$
$= .62\%$

CHAPTER 10: Inference about two population means (μ_1 & μ_2).

Two population means: we talk about the difference between them ($\mu_1 - \mu_2$).

Two topics will be covered:

1. Estimation of ($\mu_1 - \mu_2$)
2. Testing hypothesis ($\mu_1 - \mu_2$)

Estimation of $\mu_1 - \mu_2$:

μ_1 & μ_2 are unknown & hard to find. So, we take two samples (n_1 & n_2) & find $\bar{X}_1$ & $\bar{X}_2$. Then, we use $\bar{X}_1$ & $\bar{X}_2$ to estimate (μ_1 & μ_2).

Estimation is done as:

a. Point estimation or

b. Interval estimation

a. Point estimation:

$$(\mu_1 - \mu_2) = \boxed{(\bar{X}_1 - \bar{X}_2)} \rightarrow \text{Point estimate}$$

b. Interval estimation

$$(\mu_1 - \mu_2) = (\bar{X}_1 - \bar{X}_2) \pm \text{margin of error}$$

1. σ_1 & σ_2 known

$$(\mu_1 - \mu_2) = (\bar{X}_1 - \bar{X}_2) \pm Z_{\frac{\alpha}{2}} \cdot \boxed{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \rightarrow \text{St error of } (\bar{X}_1 - \bar{X}_2)$$

notice it's variance not st dv

2. σ_1 & σ_2 unknown

$$(\mu_1 - \mu_2) = (\bar{X}_1 - \bar{X}_2) \pm t_{\frac{\alpha}{2}, df} \cdot \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}$$

$df = n_1 + n_2 - 2 \rightarrow \sigma_1 = \sigma_2$,

$$df = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)^2}{\frac{1}{n_1 - 1}\left(\frac{S_1^2}{n_1}\right)^2 + \frac{1}{n_2 - 1}\left(\frac{S_2^2}{n_2}\right)^2} \rightarrow \sigma_1 \neq \sigma_2$$

eg: Create a 95% confidence interval for the difference between the two population means, given the following information...

	Sample 1	Sample 2
n size	120	80
$\bar{X}$ mean	275	258
σ pop st Div σ	15	20

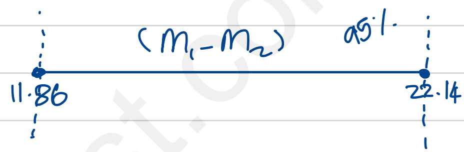
σ is known

$$(\mu_1-\mu_2) = (\bar{X}_1-\bar{X}_2) \pm Z_{\frac{\alpha}{2}} \cdot \sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$$

$\alpha = .05 \rightarrow \frac{\alpha}{2} = .025 \rightarrow \frac{1}{2} - .025 = .475$ (table)

$$(\mu_1-\mu_2) = (275-258) \pm 1.96\sqrt{\frac{(15)^2}{120}+\frac{(20)^2}{80}}$$

$$= 17 \pm 5.14 \rightarrow (11.86\ ,\ 22.14)$$



2. Testing hypotheses for two means $(\mu_1-\mu_2)$

we have a null H_0: no difference & we will Challange that using "Alternative H_1". Then we take samples & run the test through:

a. 4 steps approach (critical value)

b. P-Value approach

a. 4_ steps:

1. H_0 vs. H_1

$H_0: \mu_1 - \mu_2 = 0$

$H_1: \mu_1 - \mu_2 > 0 \rightarrow$ right side

or: $H_1: \mu_1 - \mu_2 < 0 \rightarrow$ left side

or: $H_1: \mu_1 - \mu_2 \neq 0 \rightarrow$ two side

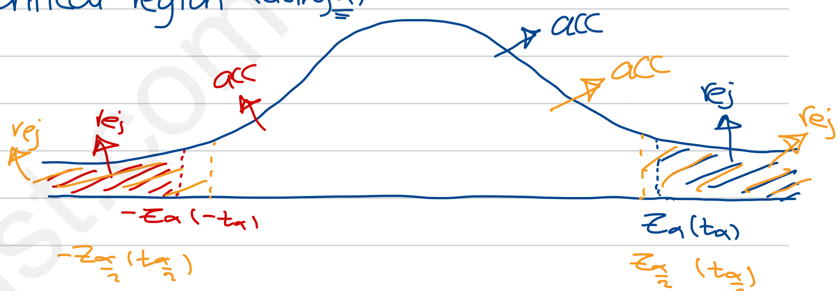
2. test statistics

a: σ_1 & σ_2 known: z stat $= \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$

b: σ_1 & σ_2 unknown: t stat $= \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{s_1^2}{n} + \dfrac{s_2^2}{n}}}$

3. Critical region (using α)

1. right side
2. left side
3. two side



4. Decision based on the value extracted in step 2...

b: P_value approach: Just like before

e.g. test if the population mean (1) is bigger than the population mean (2), using the following info..

	Sample$_1$	Sample$_2$
size (n)	40	50
Mean ($\bar{x}$)	25.2	22.8
σ	5.2	6

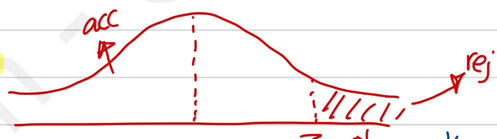
$\alpha = 1\%$

1. $H_0: M_1 - M_2 = 0 \quad H_1: M_1 - M_2 > 0$

2. Z stat $= \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{\sigma_1^2}{n} + \dfrac{\sigma_2^2}{n}}} = \dfrac{25.2 - 22.8}{\sqrt{\dfrac{(5.2)^2}{40} + \dfrac{(6)^2}{50}}} = 2.03$

Z stat

3. Critical region α

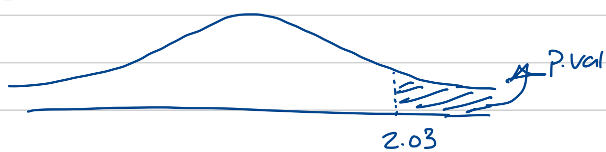


first value records (.49) from the Z table. which is 2.33.

$\frac{1}{2} - .01 = .49$
2.33 Z stat

4. Decision is accept the H_0 @ $\alpha = 1\%$ since it falls under the acceptance area.

* P- Value



$P(Z > 2.03) = \frac{1}{2} - P(0 < Z < 2.03)$

$= \frac{1}{2} - .4788 = .0212 \rightarrow 2.12\%$

P. value $< \alpha = 5\%$; reject the H_0 at $\alpha = 5\%$

eg: assume the following:

$H_0: M_1 - M_2 = 0$

$H_1: M_1 - M_2 \neq 0$

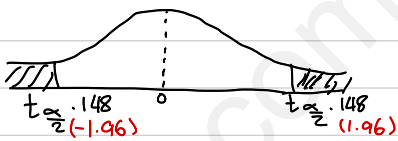
using the following DATA, run the test @ $\alpha = 5\%$

	Sam 1	Sam 2
n	80	70
$\bar{X}$	104	106
S	8.4	7.6

4 steps 1. $H_0: M_1 - M_2 = 0$

$H_1: M_1 - M_2 \neq 0$

2. $$t_{stat} = \frac{104 - 106}{\sqrt{\frac{(8.4)^2}{80} + \frac{(7.6)^2}{70}}} = -1.53$$

3. 

$\alpha = .05$, $df = 80 + 70 - 2$

$= 148$

4. Accept the Ho since it falls under the acceptance part @ $\alpha = 5\%$

NOTE: when you accept α at certain level, you accept all lower levels; however, when you reject α @ certain level, you reject all higher levels.

P. value approach: → table, inf level since 148, pick the interval that 1.53 falls between.

(.1, .05)

2 (between .05 & .1)

= between .1 & .2

P-value > $\alpha = 10\%$; accept the Ho @ 10%

eg: Sample of male & female salary information is given below:

	male	female
n	64	36
$\bar{X}$	44	41
σ^2	128	72

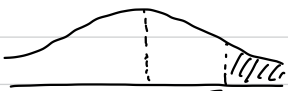
at $\alpha = 5\%$, is there evidence that male are paid more than female on average

4-steps:

1. $H_0: \mu_m - \mu_f = 0$

   $H_1: \mu_m - \mu_f > 0$

2. $$Z \text{ stat} = \frac{44 - 41}{\sqrt{\frac{128}{64} + \frac{72}{36}}} = 1.5$$

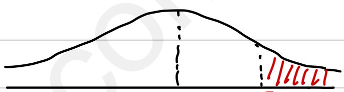
3. 

Z_α 1.65

$= .5 - .05 = .45 \rightarrow 1.65$

4. Accept the H_0 at $\alpha = 5\%$

P. value



1.5

we are trying to reject H_0

$$P.\text{value} = P(Z > 1.5) \Rightarrow 1/2 - P(0 < Z < 1.5)$$

$$= .5 - .4332 = .0668 \rightarrow 6.68\%$$

$\rightarrow$ reject the H_0 at $\alpha = 10\%$ since P. value $< \alpha = 10\%$

Gulf University for Science & Technology
Department of Economics & Finance
ECO-380: Business Statistics

Dr. Khalid Kisswani
Room: N1-116
Tel: 2530-7339
kisswani.k@gust.edu.kw

Assignment 5 (LO iii)

1. When developing an interval estimate for the difference between two sample means, with sample sizes of n_1 and n_2,
a. n_1 must be equal to n_2 X
b. n_1 must be smaller than n_2
c. n_1 must be larger than n_2
d. n_1 and n_2 can be of different sizes,

2. To construct an interval estimate for the difference between the means of two populations when the standard deviations of the two populations are unknown and it can be assumed the two populations have equal variances, we must use a t distribution with (let n_1 be the size of sample 1 and n_2 the size of sample 2)
a. $(n_1 + n_2)$ degrees of freedom
b. $(n_1 + n_2 - 1)$ degrees of freedom
c. $(n_1 + n_2 - 2)$ degrees of freedom
d. None of the above

Exhibit 10-1
Salary information regarding male and female employees of a large company is shown below.

z stat =

	Male	Female
Sample Size n	64	36
Sample Mean Salary (in \$1,000) $\bar{X}$	44	41
Population Variance (σ^2)	128	72

3. Refer to Exhibit 10-1. The point estimate of the difference between the means of the two populations is
a. -28
b. 3
c. 4
d. -4

$\bar{X}_1 - \bar{X}_2 = 44 - 41 = 3$

4. Refer to Exhibit 10-1. The standard error for the difference between the two means is
a. 4
b. 7.46
c. 4.24
d. 2.0

$= \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{128}{64} + \frac{72}{36}} = 2$

5. Refer to Exhibit 10-1. At 95% confidence, the margin of error is
a. 1.96
b. 1.645
c. 3.920
d. 2.000

$Z_{\frac{\alpha}{2}} = .5 - .025 = .475 \Rightarrow 1.96$

$1.96 * 2 =$

6. Refer to Exhibit 10-1. The 95% confidence interval for the difference between the means of the two populations is
a. 0 to 6.92
b. -2 to 2
c. -1.96 to 1.96
d. -0.92 to 6.92

$3 \pm 1.96 * 2$

$= -.92, 6.92$

7. Refer to Exhibit 10-1. If you are interested in testing whether or not the average salary of males is significantly greater than that of females, the test statistic is
a. 2.0
b. 1.5
c. 1.96
d. 1.645

$\mu_m - \mu_F > 0$

$$Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{44-41}{\sqrt{\frac{128}{64} + \frac{72}{36}}} = 1.5$$

8. Refer to Exhibit 10-1. The *p*-value is
a. 0.0668
b. 0.0334
c. 1.336
d. 1.96

Z stat

$1/2 - P(0<Z<1.5) = .0668 \rightarrow 6.68\%$ reject at 10%

$.4332$

9. Refer to Exhibit 10-1. At 95% confidence, the conclusion is the
a. average salary of males is significantly greater than females
b. average salary of males is significantly lower than females x
c. salaries of males and females are not equal
d. None of these alternatives is correct.

Exhibit 10-2

The following information was obtained from matched samples.
The daily production rates for a sample of workers before and after a training program are shown below.

Worker	Before	After
1	20	22
2	25	23
3	27	27
4	23	20
5	22	25
6	20	19
7	17	18

$\bar{X}_1 = 22$ $\bar{X}_2 = 22$

10. Refer to Exhibit 10-2. The point estimate for the difference between the means of the two populations is
a. -1
b. -2
c. 0
d. 1

11. Refer to Exhibit 10-2. The null hypothesis to be tested is H_0: $\mu_1 - \mu_2 = 0$. The test statistic is
a. -1.96
b. 1.96
c. 0
d. 1.645

12. Refer to Exhibit 10-2. Based on the results of question 11 and 5% significance level, the
a. null hypothesis should be rejected
b. null hypothesis should not be rejected
c. alternative hypothesis should be accepted
d. None of these alternatives is correct.

$\neq$

Exhibit 10-3

A statistics teacher wants to see if there is any difference in the abilities of students enrolled in statistics today and those enrolled five years ago. A sample of final examination scores from students enrolled today and from students enrolled five years ago was taken. You are given the following information.

	Today	Five Years Ago
$\bar{x}$	82	88
σ^2	112.5	54
n	45	36

13. Refer to Exhibit 10-3. The point estimate for the difference between the means of the two populations is
a. 58.5
b. 9
c. -9
d. -6

14. Refer to Exhibit 10-3. The standard error of $\bar{x}_1 - \bar{x}_2$ is
a. 12.9
b. 9.3
c. 4
d. 2

$$\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{1125}{45} + \frac{54}{36}} =$$

15. Refer to Exhibit 10-3. The 95% confidence interval for the difference between the two population means is
a. -9.92 to -2.08
b. -3.92 to 3.92
c. -13.84 to 1.84
d. -24.228 to 12.23

$$z_{\alpha/2} = .5 - .025 = .475 \Rightarrow 1.96$$

$$-6 \pm 1.96 * 2 = -9.92 , -2.08$$

16. Refer to Exhibit 10-3. The test statistic for the difference between the two population means is
a. -.47
b. -.65
c. -1.5
d. -3

$$z = \frac{-6}{2} = -3$$

17. Refer to Exhibit 10-3. The *p*-value for the difference between the two population means is
a. .0013
b. .0026
c. .4987
d. .9987

$$2(1/2 - P(-3 < z < 0))$$

$$.4987 = 2(.0013) \quad < 1\% \text{ reject}$$

18. Refer to Exhibit 10-3. What is the conclusion that can be reached about the difference in the average final examination scores between the two classes? (Use a .05 level of significance.)
a. There is a statistically significant difference in the average final examination scores between the two classes. ✓
b. There is no statistically significant difference in the average final examination scores between the two classes. x accept
c. It is impossible to make a decision on the basis of the information given. X
d. There is a difference, but it is not significant.

Exhibit 10-4

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
Sample Size n	10	12

19. Refer to Exhibit 10-4. The point estimate for the difference between the means of the two populations is
a. 0
b. 2

c. 3
d. 15

20. Refer to Exhibit 10-4. The standard error of $\bar{x}_1 - \bar{x}_2$ is

a. 3.0
b. 4.0
c. 8.372
d. 19.48

$$\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{85}{10} + \frac{90}{12}} = 4$$

21. Refer to Exhibit 10-4. The degrees of freedom for the t-distribution are

a. 22
b. 23
c. 24
d. 20

$10 + 12 - 2 =$

22. Refer to Exhibit 10-4. 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

$t_{.025, 20} = 2.086$

$3 \pm 2.086(4) = -5.3 - 11.3$

Exhibit 10-6

The management of a department store is interested in estimating the difference between the mean credit purchases of customers using the store's credit card versus those customers using a national major credit card. You are given the following information.

	Store's Card	Major Credit Card
Sample size n	64	49
Sample mean $\bar{x}$	\$140	\$125
Population standard deviation σ	\$10	\$8

23. Refer to Exhibit 10-6. A point estimate for the difference between the mean purchases of the users of the two credit cards is

a. 2
b. 18
c. 265
d. 15

24. Refer to Exhibit 10-6. At 95% confidence, the margin of error is

a. 1.694
b. 3.32
c. 1.96
d. 15

$z_{\alpha/2} = .5 - .025 = .475 \Rightarrow 1.96$

$1.96 \cdot \sqrt{\frac{10^2}{64} + \frac{8^2}{49}} =$

25. Refer to Exhibit 10-6. A 95% confidence interval estimate for the difference between the average purchases of the customers using the two different credit cards is

a. 49 to 64
b. 11.68 to 18.32
c. 125 to 140
d. 8 to 10

15 ± 3.32

Exhibit 10-9

Two major automobile manufacturers have produced compact cars with the same size engines. We are interested in determining whether or not there is a significant difference in the MPG (miles per gallon) of the two brands of automobiles. A random sample of eight cars from each manufacturer is selected, and eight drivers are selected to drive each automobile for a specified distance. The following data show the results of the test.

Driver	Manufacturer A	Manufacturer B
1	32	28
2	27	22

3	26	27
4	26	24
5	25	24
6	29	25
7	31	28
8	25	27

26. Refer to Exhibit 10-9. The mean for the differences is
a. 0.50
b. 1.5
c. 2.0
d. 2.5

$\bar{X}_1 - \bar{X}_2$

$27.625 - 25.625 = 2$

27. Refer to Exhibit 10-9. The test statistic is
a. 1.645
b. 1.96
c. 2.096
d. 1.616

2

28. Refer to Exhibit 10-9. At 90% confidence the null hypothesis
a. should not be rejected
b. should be rejected
c. should be revised
d. None of these alternatives is correct.

CHAPTER 11&8 Inference about population variance (σ^2)

→ we'll cover:

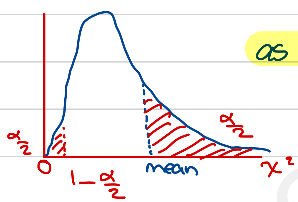
1. Estimation of σ^2 (σ)
2. Testing hypotheses → one variance
3. Testing hypotheses → two variances

1. Estimation of σ^2 (σ)

→ σ^2 is unknown & hard to find. So, we need to estimate σ^2

* We need to introduce a new distribution Chi-squared "χ^2"

Chi-squared is not symmetric, but skewed to the right.



as $n\uparrow$ → Skewness $\downarrow$

→ There's a table that reports Critical Chi-squared values...

"Critical χ^2 value" depends on: df ($n-1$), the area under the curve to the right

Note: if df isn't available on the table, you go to the higher next df..
eg: if df is 31, then it will be 35. However, any df beyond 100 (123 eg) will be treated as 100.

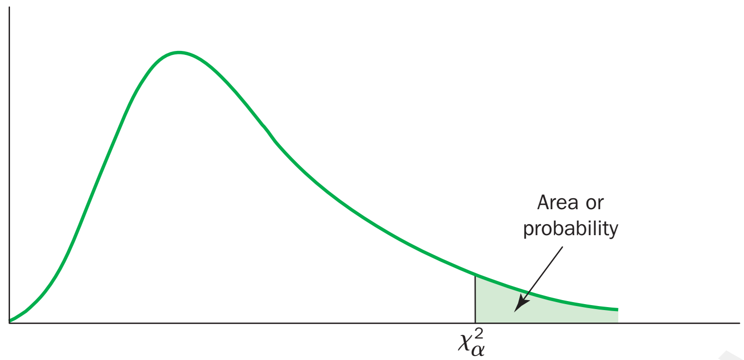
* How can we estimate σ^2: we take a sample & find S^2, then use S^2 to estimate σ^2.

Estimation is done:.

1. Point estimation: $\sigma^2 = S^2$ → Point estimate of σ^2 → $\sigma = S$ → Point est of σ
2. Interval Estimation:

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

TABLE 3 Chi-squared distribution



Entries in the table give χ^2_α values, where α is the area or probability in the upper tail of the chi-squared distribution. For example, with ten degrees of freedom and 0.01 area in the upper tail, $\chi^2_{0.01}=23.209$

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

(Continued)

TABLE 3 (*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

eg: Create a 95% confidence interval for the population variance when a sample of 25 (n) gave sample st deviation (S) $= 10$ (s).

$$\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}}} \quad : \alpha = 5\%$$

$$\frac{24(10)^2}{\chi^2_{.025,24}} \leq \sigma^2 \leq \frac{24(10)^2}{\chi^2_{.975,24}} \Rightarrow \frac{24(10)^2}{39.364} \leq \sigma^2 \leq \frac{24(10)^2}{12.401}$$

$$60.96 \leq \sigma^2 \leq 193.5$$

* We are 95% confidant the the population variance "σ^2" falls between (60.96 – 193.5).

eg: Create a 90% confidence interval for the population st deviation (σ) when a sample of 16 (n) gave sample variance "s^2" $= 7$

$$\sqrt{\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}}}} \leq \sigma \leq \sqrt{\frac{(n-1)s^2}{\chi^2_{1-\frac{\alpha}{2}}}} \quad : \alpha = 10\%$$

$$\sqrt{\frac{15(7)}{\chi^2_{.05}}} \leq \sigma \leq \sqrt{\frac{15(7)}{\chi^2_{.95}}} \Rightarrow \sqrt{\frac{15(7)}{24.996}} \leq \sigma \leq \sqrt{\frac{15(7)}{7.261}}$$

$$\sqrt{4.2} \leq \sigma \leq \sqrt{14.5}$$

* We're 90% confident that the population st deviation "σ" falls between ($\sqrt{4.2}$ – $\sqrt{14.5}$).

2. Testing hypotheses of one variance σ^2

→ We have a null (variance = value), but we will change this using alternative (bigger, smaller, or different..). Then, we take a sample & run the test...

a. 4-steps

b. P. value

4_steps method :

1. State the hypo

$H_0: \sigma^2 = \sigma_0^2$

$H_1: \sigma^2 > \sigma_0^2$ → Right sided

or $H_1: \sigma^2 < \sigma_0^2$ → Left sided

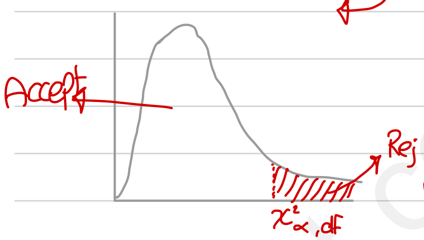
or $H_1: \sigma^2 \neq \sigma_0^2$ → Two sided

2. Test stat (χ^2 chart)

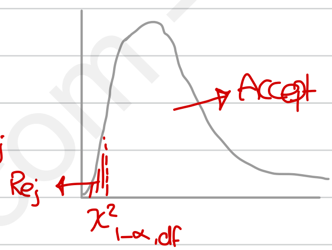
$$\chi^2 \text{ Stat} = \frac{(n-1)\,S^2}{\sigma_0^2}$$

3. Critical Region : $\underline{\underline{\alpha}}$

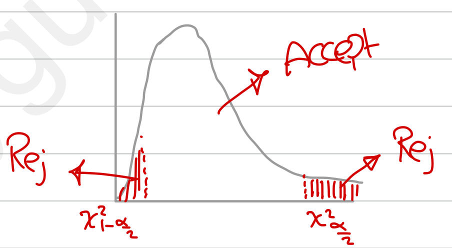
a. Right side



b. Left side



c. Two side split the α



4. Decision

→ take χ^2stat value (step2)

& look at it within the graph

We have to know df & test stat

χ^2 stat (step two)
→ by default you'll ν or α 1,5,10

P. value Method...
(you don't need α when using P.value)

→ is found as interval "range" from χ^2-table, using χ^2 stat & df.

NOTE: if it is a two-sided test using the p.value, you need to multiply by "2"

e.g: test if the population variance exceeds 50 when a sample of 16 (n) gave a st. deviation = 10 (s) ($\alpha = 1\%$)

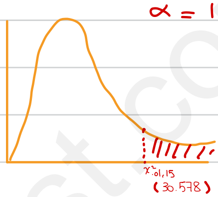
4 steps:

1. $H_0 : \sigma^2 = 50$

   $H_1 : \sigma^2 > 50$ → right-sided test

2. $\chi^2 \text{stat} = \dfrac{(n-1)s^2}{\sigma_0^2} = \dfrac{15(10)^2}{50} = 30$

3. $\alpha = 1\%$:



4. Decision is Accept the Ho at $\alpha = 1\%$

P. value:

df = 15 , χ^2 stat = 30

→ So, go to χ^2 chart df 15 & find 30 as an interval...

* So, 30 falls between .01 & .025 (1% - 2.5%).

Decision: if we pick a number between the interval (2%); P.value < α = 5%

→ So, Reject Ho @ $\alpha = 5\%$

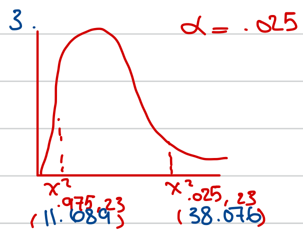
" σ " we know how to test σ^2 so we just have to square the σ...

eg: test if the st. deviation of the population is different from 2 σ_0 when a sample of 24 n gave a variance = 9 s^2 (α = 5%).

$\hookrightarrow \sigma = 2 \rightarrow \sigma^2 = 2^2 = \underline{4}$

1. $H_0 : \sigma^2 = 4$
   $H_1 : \sigma^2 \neq 4$

2. χ^2 stat $= \frac{(n-1)s^2}{\sigma_0^2} = \frac{23(9)}{4} = \underline{51.75}$

3. 

4. Reject the Ho at $\alpha = .05$

Using the P-value approach: two sided $* 2$

$H_0 : \sigma^2 = 4$ $H_1 : \sigma^2 \neq 4$ χ^2 stat $= \frac{23(9)}{4} = 51.75$

find it as an interval on the table..

P. Value = 2 (less than .005)
= less than .01 $\rightarrow$ 1%

Decision is: Reject the Ho at α = 1% since it is less than 1%

Assume that $\chi^2 = \underline{22.8}$ using the previous example...

$\rightarrow$ P. value = 2 (between .1 & .9)
= (between .2 & 1) cuz the value under goes between 0-1 (shouldn't exceed 1)

Decision: bigger than 10% (Accept the Ho).

Testing hypotheses of two variances... (σ_1^2, σ_2^2)
↳ we have a null: two variances are equal, but we will challange this using the alternative (bigger, smaller, different)

So, we take a sample & run the test:.

* 4 steps.

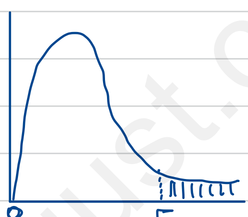
1. $H_0: \sigma_1^2 = \sigma_2^2$
   $H_1: \sigma_1^2 > \sigma_2^2$ right sided
   or $H_1: \sigma_2^2 > \sigma_1^2$ right sided σ^2 family doesn't allow us to do the comparision in terms of "smaller than"
   or $H_1: \sigma_1^2 \neq \sigma_2^2$

2. test statistics:.

$$F \text{ stat} = \frac{S_B^2}{S_S^2}$$

→ bigger s^2 : more than $\underline{1}$
→ smaller s^2

3. Critical region:. F table (no negative values...)



F value depends on $\underline{2}$ df

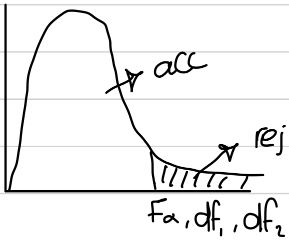
F → df_1 (n_1-1) , df_2 (n_2-1)

↓ link them, ↓ then there will be 5 values (αs) : .1, .05, .025, .001 (10%, 5%, 2.5% & .1%)

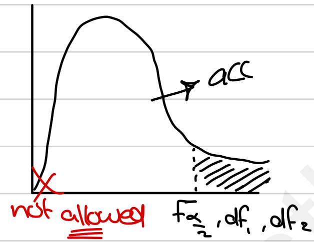
* what sample considered to be $\underline{1}$?
→ The one that has a bigger variance.

if an exact df is not available (a) → you go to the next higher level & vice versa...

a. right sided



b. two sided



Note: in one case you use the whole α & the other case you use half of α.

4. Decision

* P_value is found from F-table as an interval using Fstat & df_1 & df_2

e.g.: use the following DATA to test if variance of population 1 is different from variance of population 2.

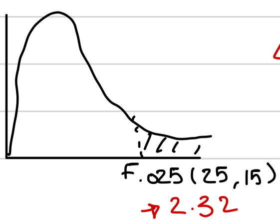
	sample 1	sample 2
size	26	16
variance	48	20

: $\alpha = 5\%$

4 steps

1. $H_0: \sigma_1^2 = \sigma_2^2$, $H_1: \sigma_1^2 \neq \sigma_2^2$

2. F stat $= \frac{S_B^2}{S_S^2} = \frac{48}{20} = 2.4$

3. 

4. Reject Ho @ $\alpha = 5\%$

P.value

F stat = 2.4 , $df_1 = 25$, $df_2 = 15$

P.value = 2 (between .1 & .025)
= (.2 & .05)

→ P.value < α = 5% → reject...

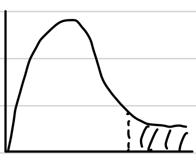
eg: test if variance Population (1) exceeds variance of Pop (2) at α = 5%.

	sample 1	sample 2
n	41	31
s	120	80

4 steps

$H_0: \sigma_1^2 = \sigma_2^2$, $H_1: \sigma_1^2 > \sigma_2^2$

$$F\,stat = \frac{S_B^2}{S_S^2} = \frac{(120)^2}{(80)^2} = 2.25$$



reject Ho at α = 5%

P_value = 2.25 (between .001 & .01) (.1% & 1%)

P-value < α = 1% → reject Ho at α = 1%

F distribution critical value landmarks

Table entries are critical values for *F** with probably *p* in right tail of the distribution.

Figure of *F* distribution (like in Moore, 2004, p. 656) here.

Degrees of freedom in denominator (df2)	*p*	Degrees of freedom in numerator (df1)										
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**12**	**24**	**1000**
1	0.100	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	60.71	62.00	63.30
	0.050	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	243.9	249.1	254.2
	0.025	647.8	799.5	864.2	899.6	921.8	937.1	948.2	956.6	976.7	997.3	1017.8
	0.010	4052	4999	5404	5624	5764	5859	5928	5981	6107	6234	6363
	0.001	405312	499725	540257	562668	576496	586033	593185	597954	610352	623703	636101
2	0.100	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.41	9.45	9.49
	0.050	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.41	19.45	19.49
	0.025	38.51	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.41	39.46	39.50
	0.010	98.50	99.00	99.16	99.25	99.30	99.33	99.36	99.38	99.42	99.46	99.50
	0.001	998.38	998.84	999.31	999.31	999.31	999.31	999.31	999.31	999.31	999.31	999.31
3	0.100	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.22	5.18	5.13
	0.050	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.74	8.64	8.53
	0.025	17.44	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.34	14.12	13.91
	0.010	34.12	30.82	29.46	28.71	28.24	27.91	27.67	27.49	27.05	26.60	26.14
	0.001	167.06	148.49	141.10	137.08	134.58	132.83	131.61	130.62	128.32	125.93	123.52
4	0.100	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.90	3.83	3.76
	0.050	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	5.91	5.77	5.63
	0.025	12.22	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.75	8.51	8.26
	0.010	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.37	13.93	13.47
	0.001	74.13	61.25	56.17	53.43	51.72	50.52	49.65	49.00	47.41	45.77	44.09
5	0.100	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.27	3.19	3.11
	0.050	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.68	4.53	4.37
	0.025	10.01	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.52	6.28	6.02
	0.010	16.26	13.27	12.06	11.39	10.97	10.67	10.46	10.29	9.89	9.47	9.03
	0.001	47.18	37.12	33.20	31.08	29.75	28.83	28.17	27.65	26.42	25.13	23.82
6	0.100	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.90	2.82	2.72
	0.050	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.00	3.84	3.67
	0.025	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.37	5.12	4.86
	0.010	13.75	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.72	7.31	6.89
	0.001	35.51	27.00	23.71	21.92	20.80	20.03	19.46	19.03	17.99	16.90	15.77
7	0.100	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.67	2.58	2.47
	0.050	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.57	3.41	3.23
	0.025	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.67	4.41	4.15
	0.010	12.25	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.47	6.07	5.66
	0.001	29.25	21.69	18.77	17.20	16.21	15.52	15.02	14.63	13.71	12.73	11.72
8	0.100	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.50	2.40	2.30
	0.050	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.28	3.12	2.93
	0.025	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.20	3.95	3.68
	0.010	11.26	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.67	5.28	4.87
	0.001	25.41	18.49	15.83	14.39	13.48	12.86	12.40	12.05	11.19	10.30	9.36
9	0.100	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.38	2.28	2.16
	0.050	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.07	2.90	2.71
	0.025	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	3.87	3.61	3.34
	0.010	10.56	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.11	4.73	4.32
	0.001	22.86	16.39	13.90	12.56	11.71	11.13	10.70	10.37	9.57	8.72	7.84

Critical values computed with Excel 9.0

Degrees of freedom in denominator (df2)		Degrees of freedom in numerator (df1)										
	p	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**12**	**24**	**1000**
10	0.100	3.29	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.28	2.18	2.06
	0.050	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	2.91	2.74	2.54
	0.025	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.62	3.37	3.09
	0.010	10.04	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.71	4.33	3.92
	0.001	21.04	14.90	12.55	11.28	10.48	9.93	9.52	9.20	8.45	7.64	6.78
12	0.100	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.15	2.04	1.91
	0.050	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.69	2.51	2.30
	0.025	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.28	3.02	2.73
	0.010	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.16	3.78	3.37
	0.001	18.64	12.97	10.80	9.63	8.89	8.38	8.00	7.71	7.00	6.25	5.44
14	0.100	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.05	1.94	1.80
	0.050	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.53	2.35	2.14
	0.025	6.30	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.05	2.79	2.50
	0.010	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	3.80	3.43	3.02
	0.001	17.14	11.78	9.73	8.62	7.92	7.44	7.08	6.80	6.13	5.41	4.62
16	0.100	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	1.99	1.87	1.72
	0.050	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.42	2.24	2.02
	0.025	6.12	4.69	4.08	3.73	3.50	3.34	3.22	3.12	2.89	2.63	2.32
	0.010	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.55	3.18	2.76
	0.001	16.12	10.97	9.01	7.94	7.27	6.80	6.46	6.20	5.55	4.85	4.08
18	0.100	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	1.93	1.81	1.66
	0.050	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.34	2.15	1.92
	0.025	5.98	4.56	3.95	3.61	3.38	3.22	3.10	3.01	2.77	2.50	2.20
	0.010	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.37	3.00	2.58
	0.001	15.38	10.39	8.49	7.46	6.81	6.35	6.02	5.76	5.13	4.45	3.69
20	0.100	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.89	1.77	1.61
	0.050	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.28	2.08	1.85
	0.025	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.68	2.41	2.09
	0.010	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.23	2.86	2.43
	0.001	14.82	9.95	8.10	7.10	6.46	6.02	5.69	5.44	4.82	4.15	3.40
30	0.100	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.77	1.64	1.46
	0.050	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.09	1.89	1.63
	0.025	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.41	2.14	1.80
	0.010	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	2.84	2.47	2.02
	0.001	13.29	8.77	7.05	6.12	5.53	5.12	4.82	4.58	4.00	3.36	2.61
50	0.100	2.81	2.41	2.20	2.06	1.97	1.90	1.84	1.80	1.68	1.54	1.33
	0.050	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	1.95	1.74	1.45
	0.025	5.34	3.97	3.39	3.05	2.83	2.67	2.55	2.46	2.22	1.93	1.56
	0.010	7.17	5.06	4.20	3.72	3.41	3.19	3.02	2.89	2.56	2.18	1.70
	0.001	12.22	7.96	6.34	5.46	4.90	4.51	4.22	4.00	3.44	2.82	2.05
100	0.100	2.76	2.36	2.14	2.00	1.91	1.83	1.78	1.73	1.61	1.46	1.22
	0.050	3.94	3.09	2.70	2.46	2.31	2.19	2.10	2.03	1.85	1.63	1.30
	0.025	5.18	3.83	3.25	2.92	2.70	2.54	2.42	2.32	2.08	1.78	1.36
	0.010	6.90	4.82	3.98	3.51	3.21	2.99	2.82	2.69	2.37	1.98	1.45
	0.001	11.50	7.41	5.86	5.02	4.48	4.11	3.83	3.61	3.07	2.46	1.64
1000	0.100	2.71	2.31	2.09	1.95	1.85	1.78	1.72	1.68	1.55	1.39	1.08
	0.050	3.85	3.00	2.61	2.38	2.22	2.11	2.02	1.95	1.76	1.53	1.11
	0.025	5.04	3.70	3.13	2.80	2.58	2.42	2.30	2.20	1.96	1.65	1.13
	0.010	6.66	4.63	3.80	3.34	3.04	2.82	2.66	2.53	2.20	1.81	1.16
	0.001	10.89	6.96	5.46	4.65	4.14	3.78	3.51	3.30	2.77	2.16	1.22

Use StaTable, WinPepi > WhatIs, or other reliable software to determine specific *p* values

Gulf University for Science & Technology
Department of Economics & Finance
ECO-380: Business Statistics

Dr. Khalid Kisswani
Room: N1-116
Tel: 2530-7339
kisswani.k@gust.edu.kw

Assignment 6 (LO iii)

1. A sample of 51 elements is selected to estimate a 95% confidence interval for the variance of the population. The chi-square values to be used for this interval estimation are
a. -1.96 and 1.96
b. 32.357 and 71.420
c. 34.764 and 67.505
d. 12.8786 and 46.9630

2. We are interested in testing whether the variance of a population is significantly less than 1.44. The null hypothesis for this test is Ho:
a. H_o: $\sigma^2 < 1.44$
b. H_o: $s^2 = 1.44$
c. H_o: $\sigma < 1.20$
d. H_o: $\sigma^2 = 1.44$

H_0: $\sigma^2 = 1.44$
H_1: $\sigma^2 < 1.44$

3. A sample of 41 observations yielded a sample standard deviation of 5. If we want to test H_0: $\sigma^2 = 20$, the test statistic is
a. 100
b. 10
c. 51.25
d. 50

$$\frac{(n-1)s^2}{\sigma_0^2} = \frac{40 * 25}{20}$$

4. The value of $F_{0.05}$ with 8 numerator and 19 denominator degrees of freedom is
a. 2.45
b. 2.51
c. 3.12
d. 3.28

5. The bottler of a certain soft drink claims their equipment to be accurate and that the variance of all filled bottles is 0.05 (or even less). The null hypothesis in a test to confirm the claim would be written as
a. H_0: $\sigma^2 \neq 0.05$ x
b. H_0: $\sigma^2 > 0.05$ x
c. H_0: $\sigma^2 < 0.05$ x
d. H_0: $\sigma^2 = 0.05$

Ho? H_1

6. A sample of 20 cans of tomato juice showed a standard deviation of 0.4 ounces. A 95% confidence interval estimate of the **variance** for the population is
a. 0.2313 to 0.8533
b. 0.2224 to 0.7924
c. 0.3042 to 0.5843
d. 0.0925 to 0.3413

$$\frac{(20-1).4^2}{8.907} \leq \sigma^2 \leq \frac{19*(.4)^2}{32.852}$$
$.341 \leq \sigma^2 \leq .092$

7. The manager of the service department of a local car dealership has noted that the service times of a sample of 15 new automobiles has a standard deviation of 4 minutes. A 95% confidence interval estimate for the variance of service times for all their new automobiles is
a. 8.576 to 39.796
b. 4 to 16
c. 4 to 15
d. 2.93 to 6.31

$$\frac{14*(4)^2}{5.629} \leq \sigma^2 \leq \frac{14\ (4)^2}{26.119}$$

8. The manager of the service department of a local car dealership has noted that the service times of a sample of 30 new automobiles has a standard deviation of 6 minutes. A 95% confidence interval estimate for the standard deviation of the service times for all their new automobiles is

a. 16.047 to 45.722
b. 4.778 to 8.066
c. 2.93 to 6.31
d. 22.833 to 65.059

$\frac{29(6)^2}{16.047} \le \sigma^2 \le \frac{29(6)^2}{45.722} \Rightarrow$

9. The producer of a certain medicine claims that their bottling equipment is very accurate and that the standard deviation of all their filled bottles is 0.1 ounce or less. A sample of 20 bottles showed a standard deviation of 0.11. The test statistic to test the claim is = σ = s

a. 400
b. 22.99
c. 4.85
d. 20

$\frac{19(.11)^2}{(.1)^2}$

10. The producer of a certain bottling equipment claims that the variance of all their filled bottles is 0.027 or less. A sample of 30 bottles showed a standard deviation of 0.2. The *p*-value for the test is σ^2 s

a. between 0.025 to 0.05
b. between 0.05 to 0.01
c. 0.05 ✗
d. 0.025 ✗

$\frac{29(.2)^2}{.027} = 42.96$

11. The chi-square values (for interval estimation) for a sample size of 21 at 95% confidence are

a. 9.591 and 34.170
b. 2.700 and 19.023
c. 8.260 and 37.566
d. -1.96 and 1.96

df = 20   .975   .025

12. The chi-square value for a one-tailed (right tail) hypothesis test at 95% confidence and a sample size of 25 is

a. 33.196
b. 36.415
c. 39.364
d. 37.652

13. The chi-square value for a one-tailed test (left 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

$1-\alpha = .9$, df = 14] =

Exhibit 11-1

Last year, the standard deviation of the ages of the students at UA was 1.8 years. Recently, a sample of 61 students had a standard deviation of 2.1 years. We are interested in testing to see if there has been a significant change in the standard deviation of the ages of the students at UA. σ n s

14. Refer to Exhibit 11-1. The test statistic is ≠

a. 44.08
b. 79.08
c. 81.67
d. 3.24

$\frac{60(2.1)^2}{(1.8)^2} =$

15. Refer to Exhibit 11-1. The *p*-value for this test is

a. 0.05
b. between 0.025 and .05

$2^*(.025 - .05)$

c. between .05 and .01
d. 1.96

16. Refer to Exhibit 11-1. 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.

$\neq$

Exhibit 11-2

We are interested in determining whether or not the variances of the sales at two music stores (A and B) are equal. A sample of 26 days of sales at store A has a sample standard deviation of 30 while a sample of 16 days of sales from store B has a sample standard deviation of 20.

17. Refer to Exhibit 11-2. The test statistic is
a. 1.50
b. 0.67
c. 1.56
d. 2.25

	A	B
n	26	16
s	30	20

$F = \frac{(30)^2}{(20)^2}$

18. Refer to Exhibit 11-2. The *p*-value for this test is
a. between 0.05 and 0.10
b. between 0.10 and 0.2
c. between 0.2 and 0.3
d. None of these alternatives is correct.

between $2^{*}(2.5\% - 5\%)$

19. Refer to Exhibit 11-2. 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 11-4

n = 81 $\quad$ $s^2 = 625$ $\quad$ H_0: $\sigma^2 = 500$
H_a: $\sigma^2 \neq 500$

20. Refer to Exhibit 11-4. The test statistic for this problem equals
a. 100
b. 101.88
c. 101.25
d. 64

$\frac{80(625)}{500}$

21. Refer to Exhibit 11-4. The *p*-value is between
a. 0.025 and 0.05
b. 0.05 and 0.1
c. 0.1 and 0.2
d. 0.2 and 0.3

2^{*} .05 – .1

22. Refer to Exhibit 11-4. 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 11-6

	Sample A	Sample B
s^2	40	96
n	16	26

We want to test the hypothesis that the population variances are equal. $\neq$

23. Refer to Exhibit 11-6. The test statistic for this problem equals
a. 0.417 ✗
b. .843 ✗
c. 2.4
d. 1.500

$\frac{96}{40} =$

24. Refer to Exhibit 11-6. The *p*-value is between
a. 0.01 and 0.025
b. 0.02 and 0.05
c. 0.025 and 0.05
d. 0.00 and 0.01

$2*(.01-.025)$

25. Refer to Exhibit 11-6. 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 11-7

	Sample A	Sample B
s^2	12.1	5
n	11	10

We want to test the hypothesis that population A has a larger variance than population B.

26. Refer to Exhibit 11-7. The test statistic for this problem equals
a. 0.4132
b. 1.96
c. 2.42
d. 1.645

$\frac{12.1}{5} =$

27. Refer to Exhibit 11-7. The *p*-value is between
a. 0.05 and 0.10
b. 0.025 and 0.05
c. 0.01 and 0.025
d. Less than 0.01

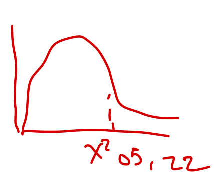
Exhibit 11-8

n = 23 $S^2 = 60$ $H_0: \sigma^2 \leq 66$
$H_a: \sigma^2 > 66$

28. Refer to Exhibit 11-8. The test statistic has a value of
a. 20.91
b. 24.20
c. 24.00
d. 20.00

$\frac{22(60)}{66}$

29. Refer to Exhibit 11-8. At 95% confidence, the critical value(s) from the table is(are)
a. 10.9823 and 36.7897
b. 33.924
c. 12.338
d. 33.924



30. Refer to Exhibit 11-8. The *p*-value is

a. less than 0.025

b. less than 0.05

c. less than 0.10

d. greater than 0.10

(.1 - .9)

31. Refer to Exhibit 11-8. The null hypothesis

a. should be rejected

b. should not be rejected

c. should be revised

d. None of these alternatives is correct.

CHAPTER 12: Testing hypothesis → population proportion (π).

→ We test if an action, policy, event, a crisis, ... had affected the population proportions.

eg:

* Did Covid-19 affect women participation in the labor market?
* Did the USD exchange rate affect the % of US made cars in Kuwait.

→ So, we take a sample & run the test.

We'll have a null: no change
the alternative: there is a change

All sizes of groups = 100%: at least two groups have changed.

* 4-steps

1. $H_0: \pi_1 = a, \pi_2 = b, \pi_3 = c, ... \pi_K = z$

in case they've same n: $(\pi_1 = \pi_2 = \pi_3 = ... \pi_K = 1/K)$

$H_1: \pi_1 \neq a, \pi_2 \neq b, \pi_3 \neq c, ... \pi_K \neq z$

in case they've same n: $(\pi_1 \neq \pi_2 \neq \pi_3 \neq ... \neq \pi_K \neq 1/K)$

2. Test Statistics

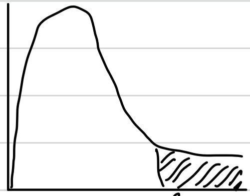
i → a group

$$\chi^2 \text{ stat} = \sum_{i=1}^{k} \frac{(f_i - e_i)^2}{e_i}$$

f_i: observed frequency of group i

e_i: expected frequency of group i

3. Critical Region (assumed as a right-sided test)



$\chi^2_{\alpha, df} \Rightarrow (K-1)$

4. Decision

* P. Value: just like before...

eg: Before the presidential debates it was expected that the % of voters in favor of candidates to be as follows

Candidates	%
Democrates	48%
Republicans	38%
Independent	4%
Undecided	10%

After the debates, a random sample of 1200 showed that 540 in favor of democratic candidate, 480 in favor of the republican candidate, 40 in favor of independent candidate, & 140 are undecided.

At $\alpha = 5\%$, test if the proportions have changed.

$\Rightarrow$

1 $H_0: \pi_1 = .48,\ \pi_2 = .38,\ \pi_3 = .04,\ \pi_4 = .1$

$H_1: \pi_1 \neq .48,\ \pi_2 \neq .38,\ \pi_3 \neq .04,\ \pi_4 \neq .1$

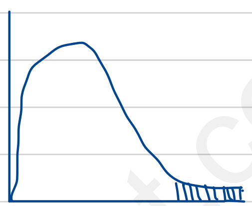
2. test statistics:. the sum of "i" goes 1 to k "4"

$$\chi^2\text{stat} = \sum_{i=1}^{4} \frac{(f_i - e_i)^2}{e_i}$$

(1200*%) : 1200*.48

group	π	f	e	$f_i - e_i$	$(f_i - e_i)^2$	$\frac{(f_i-e_i)^2}{e_i}$
1	.48	540	576	-36	1296	1296/576 = 2.25
2	.38	480	456	24	576	576/456 = 1.26
3	.04	40	48	-8	64	1.33
4	.1	140	120	20	400	3.33
						$\Sigma = 8.17$

3. Critical region



$\alpha = .05$, $df = k-1 \Rightarrow 4-1$

$\chi^2_{(.05,3)}$ 7.815 4. $\Rightarrow$ Reject the Ho @ $\alpha = 5\%$

Meaning: there's evidence at 5% level that the proportions have changed.

P-value : 8.17 is between 2.25% – 5% on χ^2 table;

P-value $< \alpha = 5\% \Rightarrow$ reject at 5%

eg: the HR department reported 60 resignations during the last year, the following table groups the resignations according to the season in which it happened.

season	resignation	
W	10	1/4
S	22	1/4
S	19	1/4
F	9	1/4

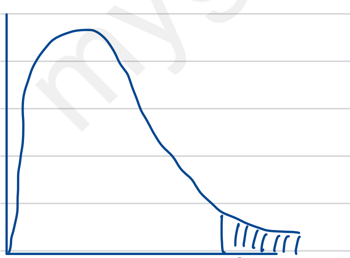
test if the number of resignations is uniform over the seasons if $\alpha = .01$.

$H_0: \pi_1 = \pi_2 = \pi_3 = \pi_4 = 1/4$

$H_1: \pi_1 \neq \pi_2 \neq \pi_3 \neq \pi_4 \neq 1/4$

$\chi^2 \text{ stat} = \sum_{i=1}^{4} \frac{(f_i - e_i)^2}{e_i} = 8.41$

group	π	f	e	$f_i - e_i$	$(f_i - e_i)^2$	$\frac{(f_i-e_i)^2}{e_i}$
1	1/4	10	$60 * \frac{1}{4} = 15$	-5	25	1.67
2	1/4	22	15	7	49	3.27
3	1/4	19	15	4	16	1.07
4	1/4	9	15	-6	36	2.4
						$= 8.41$



$\chi^2(.01, 3)$

11.345

Using the P-value:

8.41 falls between .025 – .05

P-value < α = 5%, reject at 5%

→ Accept the Ho.

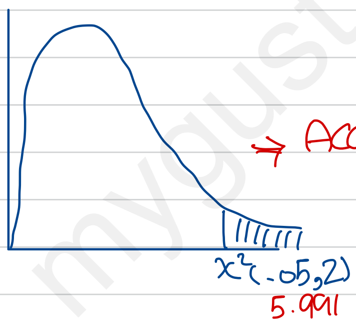
eg: before the rush began for Ramadan shopping, a department store had noted that the % of customers paying with store credit cards, % of customers paying with a major credit card, & customers paying in cash are the same... During the Ramadan rush a sample of 150 shoppers 46 used store credit card, 43 major credit card, & 61 paid cash... @ $\alpha = 5\%$ test if methods of payments have changed over the Ramadan rush.

$H_0: \pi_1 = \pi_2 = \pi_3 = 1/3$

$H_1: \pi_1 \neq \pi_2 \neq \pi_3 \neq 1/3$

$$\chi^2_{stat} = \sum_{i=1}^{3} \frac{(f_i - e_i)^2}{e_i} = 3.72$$

group	π	f	e (150 * %)	$f_i - e_i$	$(f_i - e_i)^2$	$\frac{(f_i - e_i)^2}{e_i}$
1	1/3	46	50	-4	16	.32
2	1/3	43	50	-7	49	.98
3	1/3	61	50	11	121	2.42
						= 3.72



→ Accept the H_0

Using the P-value:

3.72 falls between .1 – .9

P. value > $\alpha = 10\%$

→ reject at $\alpha = 10\%$

Gulf University for Science & Technology
Department of Economics & Finance
ECO-380: Business Statistics

Dr. Khalid Kisswani
Room: N1-116
Tel: 2530-7339
kisswani.k@gust.edu.kw

Assignment 7 (LO iv)

1. The sampling distribution for a goodness of fit test (testing hypotheses about proportions) is the
a. Poisson distribution
b. t distribution
c. normal distribution
d. chi-square distribution

2. A goodness of fit test is always conducted as a
a. lower-tail test
b. upper-tail test
c. middle test
d. None of these alternatives is correct.

Exhibit 12-1

When individuals in a sample of 150 were asked whether or not they supported capital punishment, the following information was obtained.

Do you support Capital punishment?	Number of individuals
Yes	40
No	60
No Opinion	50

3 groups

We are interested in determining whether or not the opinions of the individuals (as to Yes, No, and No Opinion) are uniformly distributed.

3. Refer to Exhibit 12-1. The expected frequency for each group is
a. 0.333
b. 0.50
c. 1/3
d. 50

e_i

$150 * 1/3$

4. Refer to Exhibit 12-1. The calculated value for the test statistic equals
a. 2
b. -2
c. 20
d. 4

5. Refer to Exhibit 12-1. The number of degrees of freedom associated with this problem is
a. 150
b. 149
c. 2
d. 3

$K-1 \Rightarrow 3-1$

6. Refer to Exhibit 12-1. The *p*-value is
a. larger than 0.1
b. less than 0.1
c. less than 0.05
d. larger than 0.9

7. Refer to Exhibit 12-1. The conclusion of the test (at 95% confidence) is that the
a. distribution is uniform
b. distribution is not uniform
c. test is inconclusive
d. None of these alternatives is correct.

Exhibit 12-2

Last school year, the student body of a local university consisted of 30% freshmen, 24% sophomores, 26% juniors, and 20% seniors. A sample of 300 students taken from this year's student body showed the following number of students in each classification.

Freshmen	83 f
Sophomores	68
Juniors	85
Seniors	64

We are interested in determining whether or not there has been a significant change in the classifications between the last school year and this school year.

8. Refer to Exhibit 12-2. The expected number of freshmen is
a. 83
b. 90
c. 30
d. 10

9. Refer to Exhibit 12-2. The expected frequency of seniors is
a. 60
b. 20%
c. 68
d. 64

10. Refer to Exhibit 12-2. The calculated value for the test statistic equals
a. 0.5444
b. 300
c. 1.6615
d. 6.6615

11. Refer to Exhibit 12-2. The *p*-value is
a. less than .005
b. between .025 and 0.05
c. between .05 and 0.1
d. greater than 0.1

12. Refer to Exhibit 12-2. 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.

Exhibit 12-4

In the past, 35% of the students at ABC University were in the Business College, 35% of the students were in the Liberal Arts College, and 30% of the students were in the Education College. To see whether or not the proportions have changed, a sample of 300 students was taken. Ninety of the sample students are in the Business College, 120 are in the Liberal Arts College, and 90 are in the Education College.

13. Refer to Exhibit 12-4. The expected frequency for the Business College is
a. 0.3
b. 0.35
c. 90
d. 105

14. Refer to Exhibit 12-4. The calculated value for the test statistic equals
a. 0.01
b. 0.75
c. 4.29
d. 4.38

15. Refer to Exhibit 12-4. The hypothesis is to be tested at the 5% level of significance. The critical value from the table equals
a. 1.645
b. 1.96
c. 5.991
d. 7.815

16. Refer to Exhibit 12-4. The *p*-value is
a. greater than 0.1
b. between 0.05 and 0.1
c. between 0.025 and 0.05
d. between 0.01 and .025

17. Refer to Exhibit 12-4. The conclusion of the test is that the
a. proportions have changed significantly X
b. proportions have not changed significantly
c. test is inconclusive
d. None of these alternatives is correct.

Exhibit 12-8

The following shows the number of individuals in a sample of 300 who indicated they support the new tax proposal.

Political Party	**Support**
Democrats	100
Republicans	120
Independents	80

We are interested in determining whether or not the opinions of the individuals of the three groups are uniformly distributed.

18. Refer to Exhibit 12-8. The expected frequency for each group is
a. 0.333
b. 0.50
c. 50
d. None of these alternatives is correct. ⇒ 100

19. Refer to Exhibit 12-8. The calculated value for the *test statistic* equals
a. 300
b. 4
c. 0
d. 8

20. Refer to Exhibit 12-8. The number of degrees of freedom associated with this problem is
a. 2
b. 3
c. 300
d. 299

K-1 ⇒ 3-1

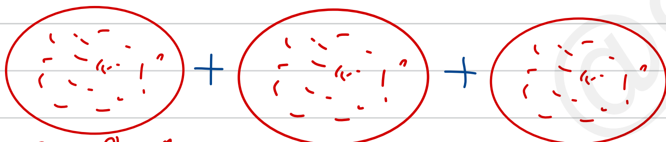
CHAPTER 13: Analysis of Variance (ANOVA)

Testing if 3 or more means are equal or no.

We have a null: All means are equal, but the alternative says not all the means are equal.
So, we take samples & we run the test.

* New concepts · Conceptual view:

assume we're testing 3 means.



= one big sample (combined)

sample 1: $n_1, \bar{x}_1, s_1^2$ sample 2: $n_2, \bar{x}_2, s_2^2$ sample 3: $n_3, \bar{x}_3, s_3^2$ → Total observations

$n_T = n_1 + n_2 + n_3$

↓

we can find the AVG (diff sources)

↓

overall mean (grand mean) $\bar{\bar{x}}$

* In this kind of analysis, they refer to the sample as treatment.

$$\bar{\bar{x}} = \frac{\text{sum of all observations}}{\text{total number of observations}} = \frac{\sum X_s}{n_T}$$

or: $\bar{\bar{x}} = \frac{\sum \bar{x}}{k}$ if all samples of same size

* Assume the following samples:

Sample	n	$\bar{x}$	$\sum x = n\bar{x}$
A	10	5	50
B	20	4	80
C	15	10	150
$\sum =$	45	19	280

Find $\bar{\bar{x}}$? using: $\bar{\bar{x}} = \frac{\sum X_s}{n_T}$ since $n_s \neq$

$\Rightarrow \bar{\bar{x}} = \frac{\sum X_s}{n_T} = \frac{280}{45} = 6.22$

$\Rightarrow \frac{19}{3} = 6.33$ the results are not the same

* Assume the following samples:.

Sample	n	$\bar{X}$
A	10	5
B	10	4
C	10	10

$\Rightarrow$ Find $\bar{\bar{X}}$? because n_s are the same, use $\bar{\bar{X}} = \frac{\Sigma \bar{X}}{k} = \frac{19}{3} = 6.33$

* Why there are differences between $\bar{X}_s$ & $\bar{\bar{X}}_s$? two reasons

1. Due to differences between the groups (treatments)

These differences can be measured by "Sum Square Treatments" (SSTR)

$$SSTR = \sum_{i=1}^{k} n_i (\bar{X}_i - \bar{\bar{X}})^2$$

Then, we change SSTR into "Mean Square Treatment" (MSTR)

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

2. Due to differences within the groups (Error)

These differences can be measured by "Sum Square Error" SSE

$$SSE = \sum_{i=1}^{k} (n_i - 1) S_i^2$$

Then, we convert the SSE into "Mean Square Error" MSE

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

How do we run the test?

4. steps:.

1. $H_0: M_1 = M_2 = M_3 = \dots = M_K$

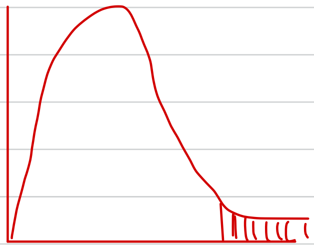
$H_1: M_1 \neq M_2 \neq M_3 \neq \dots \neq M_K$

2. Test Statistics:.

$$F \text{ stat} = \frac{MSTR}{MSE} \quad \begin{matrix} \Rightarrow \text{kind of variance} \\ \Rightarrow \text{kind of variance} \end{matrix}$$

Compare two variances to make a decision

3. Critical Region "treat is as a right sided"



$df_1 = K-1$
$df_2 = n_T - K$

4. Decision

P. value: Just like before

The ANOVA table:

	Source of variation	df	Sum square SS	Mean square MS	F. Stat
diff among sample	Treatment	$K-1$	SSTR	MSTR	$\left(\frac{MSTR}{MSE}\right)$
within sample	Error	$n_T - K$	SSE	MSE	
	Total	$n_T - 1$	SST → SSTR + SSE	*	

eg: test if the mean of all 4 cities is the same or no using the following

$\alpha = 5\%$

Sample	n	$\bar{X}$	S^2	$\Sigma X = n\bar{x}$
1	8	10.25	14.2	82
2	5	16.4	4.7	82
3	7	16	10.7	112
4	10	11.4	11.5	114
	=30			=390

4. steps:.

1. $H_0: M_1 = M_2 = M_3 = M_4$

   $H_1: M_1 \neq M_2 \neq M_3 \neq M_4$

2. $F\ stat = \frac{MSTR}{MSE}$

$MSTR = \frac{SSTR}{K-1} \Rightarrow SSTR = \sum_{i=1}^{4} n_i(\bar{x} - \bar{\bar{x}})^2 \Rightarrow \bar{\bar{X}} = \frac{\Sigma X_s}{n_T} = \frac{390}{30} = 13$

$SSTR = 8(10.25-13)^2 + 5(16.4-13)^2 + 7(16-13)^2 + 10(11.4-13)^2 = 206.9$

So, $MSTR = \frac{206.9}{4-1} = 68.97$ → first part

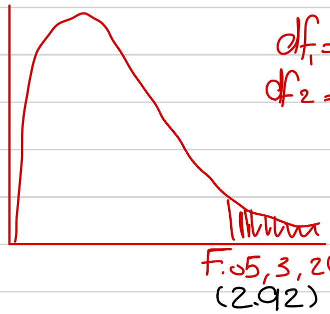
$MSE = \frac{SSE}{n_T - K} \Rightarrow SSE = \sum_{i=1}^{4}(n_i - 1)S_i^2$

$= (8-1)14.2 + (4)4.7 + (6)10.7 + (9)11.5 = 285.9$

$\therefore MSE = \frac{285.9}{(30-4)} = 11$ → second part

→ $F\ stat = \frac{MSTR}{MSE} = \frac{68.97}{11} = 6.27$

3. Critical region



$df_1 = n-1 = 4-1 = 3$

$df_2 = n_T - k = 30 - 4 = 26$

$\Longrightarrow$ Reject at $\alpha = 5\%$

P-value :

Between .1% - 1% reject at 1%

ANOVA Table :

Sources of Variation	df	SS	MS	F Stat
Treatment	3	206.9	68.97	6.27
Error	26	285.9	11	
Total	29	492.8		

eg: given the following ANOVA, answer the questions..

SOV	SS	df	MS	F
Treatment	** 36	42-40 2	18^*	3
Error	*** 240	40	6	
Total	276	42		

k-1

a. find missing values
b. write Ho & H_1
c. Find p-value & write the decision.

$* \; F = \frac{MSTR}{MSE}$

$3 = \frac{MSTR}{6}$

$** \; MSTR = \frac{SSTR}{df} \Rightarrow 18 = \frac{MSTR}{2} = 36$

$*** \; MSE = \frac{SSE}{df} \Rightarrow 6 = \frac{SSE}{40} = 240$

b. $H_0: \mu_1 = \mu_2 = \mu_3$, $H_1: \mu_1 \neq \mu_2 \neq \mu_3$

c. p-value is between .05 & .1 $\Rightarrow$ p-value $< \alpha = 10\%$; reject @ $\alpha = 10\%$.

eg: given the following ANOVA, answer the questions:

SOV	SS	df	MS	F
Treatment	90	3	* 30	*** 5
Error	120	20	** 6	
Total	210	23		

$* \; MSTR = \frac{SSTR}{df} = \frac{90}{3} = 30$

$** \; MSE = \frac{SSE}{df} = \frac{120}{20} = 6$

$*** \; F = \frac{MSTR}{MSE} = \frac{30}{6} = 5$

a. Find missing values
b. Write Ho & H_1 $\rightarrow$ $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$; $H_1: \mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4$
c. Find p-value & Decision.

$\hookrightarrow$ p-value is between .001 & .01 $\Rightarrow$ p-value is less than .1 ; reject @ $\alpha = 1\%$.

Gulf University for Science & Technology
Department of Economics & Finance
ECO-380: Business Statistics

Dr. Khalid Kisswani
Room: N1-116
Tel: 2530-7339
kisswani.k@gust.edu.kw

Assignment 8 (LO v)

1. In an analysis of variance problem involving 3 treatments and 10 observations per treatment, SSE = 399.6. The MSE for this situation is

30

a. 133.2
b. 13.32
c. 14.8
d. 30.0

$MSE = \frac{SSE}{n_T - k} = \frac{399.6}{30-3} =$

2. When an analysis of variance is performed on samples drawn from K populations, the mean square between treatments (MSTR) is

a. $SSTR/n_T$
b. $SSTR/(n_T - 1)$
c. SSTR/K
d. SSTR/(K - 1)

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

3. In an analysis of variance where the total sample size for the experiment is nT and the number of populations is K, the mean square within treatments is

a. $SSE/(n_T - K)$
b. $SSTR/(n_T - K)$
c. SSE/(K - 1)
d. SSE/K ×

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

4. The F ratio in a completely randomized ANOVA is the ratio of

a. MSTR/MSE
b. MST/MSE
c. MSE/MSTR
d. MSE/MST

$\frac{MSTR}{MSE}$

5. The critical F value with 6 numerator and 60 denominator degrees of freedom at $\alpha = .05$ is

a. 3.74
b. 2.25
c. 2.37
d. 1.96

6. An ANOVA procedure is applied to data obtained from 6 samples where each sample contains 20 observations. The degrees of freedom for the critical value of F are

×6

a. 6 numerator and 20 denominator degrees of freedom
b. 5 numerator and 20 denominator degrees of freedom
c. 5 numerator and 114 denominator degrees of freedom
d. 6 numerator and 20 denominator degrees of freedom

$df_1 = k-1 = 6-1 = 5$

$df_2 = n_T - k = 120 - 6 = 114$

7. In an analysis of variance problem if SST = 120 and SSTR = 80, then SSE is

a. 200
b. 40
c. 80
d. 120

SSTR 80
SSE ?
SST 120

8. In a completely randomized design involving three treatments, the following information is provided:

	Treatment 1	Treatment 2	Treatment 3	
Sample Size n	5	10	5	$= 20 \leftarrow n_T$
Sample Mean $\bar{x}$	4	8	9	
×	20	80	45	= 145

The overall mean for all the treatments is $\bar{\bar{x}} = \frac{145}{20}$

a. 7.00
b. 6.67
c. 7.25
d. 4.89

Exhibit 13-1

SSTR = 6,750 H$_0$: $\mu_1=\mu_2=\mu_3=\mu_4$
SSE = 8,000 H$_a$: at least one mean is different
n_T = 20 SST 14.75

9. Refer to Exhibit 13-1. The mean square between treatments (MSTR) equals

a. 400
b. 500
c. 1,687.5
d. 2,250

$MSTR = \frac{SSTR}{k-1} = \frac{6{,}750}{3}$

10. Refer to Exhibit 13-1. The mean square within treatments (MSE) equals

a. 400
b. 500
c. 1,687.5
d. 2,250

$MSE = \frac{SSE}{n_T - k} = \frac{8{,}000}{16}$

11. Refer to Exhibit 13-1. The test statistic to test the null hypothesis equals

a. 0.22
b. 0.84
c. 4.22
d. 4.5

$\frac{MSTR}{MSE} = \frac{2{,}250}{500}$

12. Refer to Exhibit 13-1. The null hypothesis is to be tested at the 5% level of significance. The *p*-value is

a. less than .01
b. between .01 and .025
c. between .025 and .05
d. between .05 and .10

13. Refer to Exhibit 13-1. The null hypothesis

a. should be rejected
b. should not be rejected
c. was designed incorrectly
d. None of these alternatives is correct.

4 4 4

Exhibit 13-3

To test whether or not there is a difference between treatments A, B, and C, a sample of 12 observations has been randomly assigned to the 3 treatments. You are given the results below.

Treatment			**Observation** "Xs"		$\bar{x}$
A	20	30	25	33	27
B	22	26	20	28	24
C	40	30	28	22	30

$\bar{\bar{x}} = \frac{324}{12}$ $\frac{81}{3} = 27$

SSTR $n(\bar{x}_i - \bar{\bar{x}})^2$: 0, 36, 36, 72

SSE $(n_i - 1)s^2$

14. Refer to Exhibit 13-3. The null hypothesis for this ANOVA problem is

a. $\mu_1=\mu_2$
b. $\mu_1=\mu_2=\mu_3$
c. $\mu_1=\mu_2=\mu_3=\mu_4$
d. $\mu_1=\mu_2= \dots =\mu_{12}$

$s^2 = \frac{\sum(x-\bar{x})^2}{n-1}$

15. Refer to Exhibit 13-3. The mean square between treatments (MSTR) equals
a. 1.872
b. 5.86
c. 34
d. 36

$MSTR = \frac{SSTR}{k-1} \rightarrow \frac{72}{2} =$

16. Refer to Exhibit 13-3. The mean square within treatments (MSE) equals
a. 1.872
b. 5.86
c. 34
d. 36

17. Refer to Exhibit 13-3. The test statistic to test the null hypothesis equals
a. 0.944 x
b. 1.059
c. 3.13
d. 19.231

$\frac{MSTR}{MSE} = \frac{36}{34} =$

18. Refer to Exhibit 13-3. The null hypothesis is to be tested at the 1% level of significance. The *p*-value is
a. greater than 0.1
b. between 0.1 to 0.05
c. between 0.05 to 0.025
d. between 0.025 to 0.01

19. Refer to Exhibit 13-3. The null hypothesis
a. should be rejected
b. should not be rejected
c. should be revised
d. None of these alternatives is correct.

Exhibit 13-4

In a completely randomized experimental design involving five treatments, 13 observations were recorded for each of the five treatments (a total of 65 observations). The following information is provided.

SSTR = 200 (Sum Square Between Treatments)
SST = 800 (Total Sum Square)

$n_T = 5 * 13 = 65$

20. Refer to Exhibit 13-4. The sum of squares within treatments (SSE) is
a. 1,000
b. 600
c. 200
d. 1,600

800-200

21. Refer to Exhibit 13-4. The number of degrees of freedom corresponding to between treatments is
a. 60
b. 59
c. 5
d. 4

k-1

22. Refer to Exhibit 13-4. The number of degrees of freedom corresponding to within treatments is
a. 60
b. 59
c. 5
d. 4

$n_T - k$

23. Refer to Exhibit 13-4. The mean square between treatments (MSTR) is
a. 3.34
b. 10.00
c. 50.00
d. 12.00

$$MSTR = \frac{SSTR}{k-1} = \frac{200}{5-1}$$

24. Refer to Exhibit 13-4. The mean square within treatments (MSE) is
a. 50
b. 10
c. 200
d. 600

$$MSE = \frac{SSE}{n_T - k} = \frac{600}{60} = 10$$

25. Refer to Exhibit 13-4. The test statistic is
a. 0.2
b. 5.0
c. 3.75
d. 15

$$F.stat = \frac{MSTR}{MSE} = \frac{50}{10} = 5$$

26. Refer to Exhibit 13-4. If at 95% confidence we want to determine whether or not the means of the five populations are equal, the *p*-value is
a. between 0.05 to 0.10
b. between 0.025 to 0.05
c. between 0.01 to 0.025
d. less than 0.01

Exhibit 13-5

Part of an ANOVA table is shown below.

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F
Between Treatment	180	3 ($k-1$)	60	3
Within Treatment (Error)	300	15 ($n_T - k$)	20	
TOTAL	480	18	*	

27. Refer to Exhibit 13-5. The mean square between treatments (MSTR) is
a. 20
b. 60
c. 300
d. 15

$$MSTR = \frac{SSTR}{k-1} = \frac{180}{3}$$

28. Refer to Exhibit 13-5. The mean square within treatments (MSE) is
a. 60
b. 15
c. 300
d. 20

$$MSE = \frac{SSE}{n_T - k} = \frac{300}{15}$$

29. Refer to Exhibit 13-5. The test statistic is
a. 2.25
b. 6
c. 2.67
d. 3

$$\frac{60}{20}$$

30. Refer to Exhibit 13-5. If at 95% confidence, we want to determine whether or not the means of the populations are equal, the *p*-value is
a. between 0.01 to 0.025
b. between 0.025 to 0.05
c. between 0.05 to 0.1
d. greater than 0.1

Exhibit 13-6

Part of an ANOVA table is shown below.

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square	F
Between Treatments	64	4	16	8
Within Treatments Error	36	18	2	
Total	100	22		

$$MSTR = \frac{SSTR}{k-1} \quad \left(16 = \frac{64}{k-1}\right)$$

$$MSE = \frac{SSE}{n_T - k} \quad \left(2 = \frac{36}{n_T - k}\right)$$

$$F = \frac{MSTR}{MSE} \qquad 8 = \frac{16}{2}$$

31. Refer to Exhibit 13-6. The number of degrees of freedom corresponding to between treatments is
a. 18
b. 2
c. 4
d. 3

$k-1$ 64/16

32. Refer to Exhibit 13-6. The number of degrees of freedom corresponding to within treatments is
a. 22
b. 4
c. 5
d. 18

$n_T - k$ 36/2

33. Refer to Exhibit 13-6. The mean square between treatments (MSTR) is
a. 36
b. 16
c. 64
d. 15

$$MSTR = \frac{64}{4} =$$

34. Refer to Exhibit 13-6. If at 95% confidence we want to determine whether or not the means of the populations are equal, the *p*-value is
a. greater than 0.1
b. between 0.05 to 0.1
c. between 0.025 to 0.05
d. less than 0.01

35. Refer to Exhibit 13-6. The conclusion of the test is that the means
a. are equal accept
b. may be equal
c. are not equal strong evidence
d. None of these alternatives is correct. x

CHAPTER 14: Simple linear regression

Estimating relationship between a dependent variable (Y) & an independent variable (X)

Y depends on X

$$y = f(x)$$

$$\hookrightarrow y = \beta_0 + \beta_1 x$$

β_0: Intercept (constant): Y value when $X = 0$

↳ Connection point between the line & Y.

β_1: $\frac{\Delta Y}{\Delta X}$ → slope: moving the same direction.

if $\beta_1 > 0$ → X & Y positively related same direction

$\beta_1 < 0$ → X & Y Negatively related opposite

eg: $\beta_1 = -3$: X & Y will move to the opposite direction.

↳ This says: if X ↑ by 1 unit, then Y ↓ by 3

$y = \beta_0 + \beta_1 x$ → "Mathematics"

in Statistics:

$$y = \underbrace{\beta_0 + \beta_1 x}_{\text{Explained}} + \underbrace{\varepsilon}_{\text{Unexplained}}$$

→ "error term" → Captures the effect of unobserved variables"

↳ $E(Y|X)$ expected Y given the X

$E(Y|X) = \beta_0 + \beta_1 x$ → Regression is about finding values for β_0 & β_1

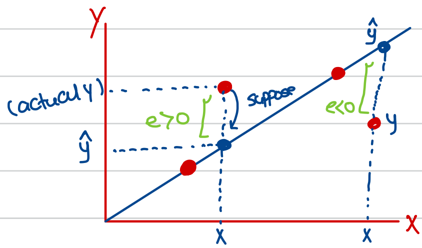
↳ population level

Because it is population, finding true β_0 & β_1 is hard

So, we take a sample (x, y) & estimate a similar relationship

$\hat{y} = b_0 + b_1 x$ → So, we need to estimate b_0 & b_1

Assume we took a sample (x & y) & we plotted the points



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

→ No line can connect all the points

Note that the line is an assumption...

the difference between actual y & expect $\hat{y}$ → residual (e) $e = y - \hat{y}$

e above the line → + residual ; e below the line → − e

* $\Sigma e = 0$ = sum all residuals = zero

How can we choose the best line (best b_0 & b_1)?

→ we use "Ordinary Least squares: OLS"

OLS:

the best line is the one that minimizes Σe^2

$$b_1 = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(x-\bar{x})^2} , \quad b_0 = \bar{y} - b_1 \bar{x}$$

eg: for the following sample, find the least square line...

X	Y	$(x-\bar{x})$	$(y-\bar{y})$	$(x-\bar{x})(y-\bar{y})$	$(x-\bar{x})^2$	$\hat{y}$	$e = y-\hat{y}$	e^2	$(y-\bar{y})^2$
1	14	−1	−6	6	1	15	−1	1	36
3	24	1	4	4	1	25	−1	1	16
2	18	0	−2	0	0	20	−2	4	4
1	17	−1	−3	3	1	15	2	4	9
3	27	1	7	7	1	25	2	4	49
				$\Sigma = 20$	$\Sigma = 4$			$\Sigma e^2 = 14$ → SSE	$\Sigma = 114$ → SST

$\bar{x} = \frac{10}{5} = 2$, $\bar{y} = \frac{100}{5} = 20$

best slope

$b_1 = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(x-\bar{x})^2} = \frac{20}{4} = 5$; $b_0 = \bar{y} - b_1\bar{x} = 20 - 5(2) = 10$

$\therefore \hat{y} = 10 + 5x$ → The best fit the DATA → (the best line)

↳ the min sum of e^2

* Coefficient of determination (R^2)

↳ We use it to check if the model is good or no.

R^2: Shows how much (X) explains of the variations of (Y)

total variation of Y → Measured by $SST = \Sigma(Y-\bar{Y})^2$ ← Sum squares total

X (regression) — Sum squares regression ↳ measured by $SSR = \Sigma(\hat{Y}-\bar{Y})^2$

error — Sum squares Error ↳ measured by $SSE = \Sigma(Y-\hat{Y})^2 = \Sigma e^2$

$$\therefore\ SST = SSR + SSE$$

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

$0 \leq R^2 \leq 1$ → $R^2 \geq .5$ → model is good

* for pervious example, find R^2

$SSE = 14$

$SST = 114$] → $SSR = 100$

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} = 1 - \frac{14}{114} = .88 \rightarrow 88\%$$

→ (X) explains 88% of the total variations of Y.

* Coefficient of correlation (r)

previously: $r = \frac{Cov_{XY}}{S_X S_Y} \rightarrow -1 \leq r \leq +1$

Now: $r = (\text{sign of } b_1)\sqrt{R^2}$

eg: assume $\hat{y} = 20 - \frac{1}{2}x$, $R^2 = .81$ → $r = ?$

$r = -\sqrt{R^2} = -\sqrt{.81} = -.9$

eg: assume the following sample:

$\hat{Y} = 11 - X$

they have to add up to ZERO

X	Y	$(X-\bar{X})$	$(Y-\bar{Y})$	$(X-\bar{X})(Y-\bar{Y})$	$(X-\bar{X})^2$	$(Y-\bar{Y})^2$	$\hat{Y}$	$e = Y-\hat{Y}$	e^2
4	12	0	5	0	0	25	7	5	25
6	3	2	-4	-8	4	16	5	-2	4
2	7	-2	0	0	4	0	9	-2	4
4	6	0	-1	0	0	1	7	-1	1
				$\Sigma = -8$	$\Sigma = 8$	$\Sigma = 42$ SST		$= 0$	$= 34$ SSE

SSR = 8

a. Find the least square line

b. Find: R^2 & r c. test if the slope is significant ($\alpha = 5\%$)

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

$$b_1 = \frac{\Sigma(X-\bar{X})(Y-\bar{Y})}{\Sigma(X-\bar{X})^2} \quad . \quad \bar{X} = \frac{16}{4} = 4 \quad ; \quad \bar{Y} = \frac{28}{4} = 7$$

$b_1 = \frac{-8}{8} = -1$ → if x goes ↑ by 1 unit then y goes ↓ by 1 unit

$b_0 = \bar{Y} - b_1\bar{X} = 7 - (-1)4 = 7 + 4 = 11$

→ $\hat{Y} = 11 - X$ our equation (the best line for the DATA we have)

b. R^2 ; r (good model or no / good independent)

$$R^2 = \frac{SSR}{SST} \text{ or } = 1 - \frac{SSE}{SST} :$$

$SSR = \Sigma(\hat{Y} - \bar{Y})^2$

→ $SSE = \Sigma(Y - \hat{Y})^2 = \Sigma e^2$

→ $SST = \Sigma(Y - \bar{Y})^2 = SSR + SSE$

$R^2 = \frac{8}{42} = .19 : 19\%$ → X explains 19% of the total variation of Y : not good cuz the rest 81% is explained by the error (unknown variable)

→ we pick a weak variable

→ $r = (\text{sign of } b_1)\sqrt{R^2} = -\sqrt{.19} = -.44$ → weak correlation

c. test if β_1 significant or no:.

4. steps:.

$H_0: \beta_1 = 0 \rightarrow$ not significant (x doesn't affect y)

$H_1: \beta_1 \neq 0$

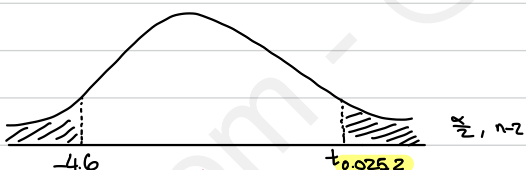
2. test stat: $t.stat = \dfrac{b_1}{se(b_1)} \rightarrow$ st error of b_1

$$se(b_1) = \sqrt{\frac{MSE}{\Sigma(x-\bar{x})^2}} = \sqrt{\frac{\left(\frac{\Sigma e^2}{n-2}\right)}{\Sigma(x-\bar{x})^2}}$$ → 2 cuz estimating 2 things

$$Seb_1 = \sqrt{\frac{\frac{32}{2}}{8}} = 1.46$$

$$\therefore t.stat = \frac{-1}{1.46} = -.69$$

3. Critical region



∴ accept the H_0 at $\alpha = 5\%$ ∵ there's no evidence that β_1 is significant.

P-value:

→ Between $2*(.25 - .4)$ → 50% - 80% ⇒ greater than 10% accept

CHAPTER: 15 : Multiple linear Regression

$Y = f(x_1, x_2, x_3, \dots x_p)$

$E(Y|X) \rightarrow$ expected y given xs (explained) unexplained

$$Y = \underbrace{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots \beta_p x_p}_{E(Y|X)} + \underbrace{\varepsilon}_{\text{unexplained}} \rightarrow \text{error}$$

↳ many independent variables & everyone has its slope.

\# of Xs : P

\# of β : P+1 → K "how many Betas"

$$E(Y|X) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_p x_p$$

β_0 : Intercept → y value when all Xs = 0

↳ where the line gonna cross the "y" access

$$\beta_1 = \frac{\Delta Y}{\Delta X_1} \Big| \text{other variables are constant}$$

$$\beta_2 = \frac{\Delta Y}{\Delta X_2} \Big| \text{other variables are constant.}$$

Finding : $\beta_1, \beta_2, \beta_3, \dots \beta_p$ is hard → Population

↳ So, we take a sample & estimate.

$$\hat{Y} = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + \dots b_p x_p$$

we use "OLS" to get the best line ('minimize se^2')

Now: the estimated outcomes of regression will be given

eg:

variable	coeffecient	st. error
intercept	2.5	.95
X_1	1.8	1.03
X_2	-3.6	5.88
X_3	4.1	.02

Also, ANOVA table will be given (missing values)

Sources of variation	df	SS	MS	F-stat
Regression	K-1 (p)	SSR	MSR	$\frac{MSR}{MSE}$
Error	n-K	SSE	MSE	
Total	n-1	SST		

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

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

→ What's needed from a student?

1. write the model
2. Find R^2
3. Create a confidence interval for any β
4. Testing hypotheses

1. Writing the model:

variable	coeffecient	st. error
intercept	2.5	.95
X_1	1.8	1.03
X_2	-3.6	5.88
X_3	4.1	.02

$\hat{y} = b_0 + b_1x_1 + b_2x_2 + b_3x_3$

$\hat{y} = 2.5 + 1.8x_1 \underline{-3.6x} + 4.1x_3$

When x_2 ↑ by 1 unit, y ↓ by 3.6, keeping all other variables constant.

2. Finding R^2 → it shows how much Xs are explaining total variation of y

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

3. Creating a confidence interval for any "β" question will specify

$$\beta_i = b_i \pm t_{\frac{\alpha}{2}, df} \cdot se(b_i)$$

eg: $\beta_7 = b_7 \pm t_{\frac{\alpha}{2}, df} \cdot se(b_7)$: df = n-K

4. Testing hypotheses:

1. Individual significance test
2. Overall significance test

1. Individual significance test tests if a certain specific β is significant

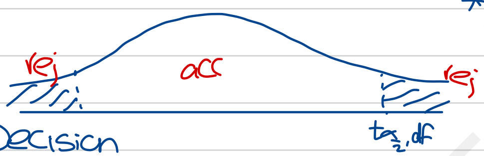
4-steps: 1. $H_0: \beta_i = 0$ → meaning that β_i is not significant

$H_1: \beta_i \neq 0$

2. Test stat:

$$t\text{-stat} = \frac{b_i}{se(b_i)}$$

3. Critical region



4. Decision

* P. value

make sure to multiply by "2" since it's a two-sided test

2. overall significance test tests if all βs are significant

→ if ALL slopes are zero

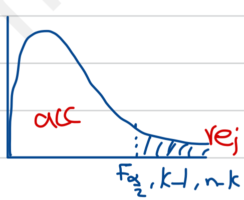
4-steps: 1. $H_0: \beta_1 = \beta_2 = \beta_3 = \dots = \beta_p = 0$

$H_1: \beta_1 \neq \beta_2 \neq \beta_3 \neq \dots \neq \beta_p \neq 0$ at least 1 is not zero "at least one of the Xs affects Y"

2. Test stat:

$$F\text{.stat} = \frac{MSR}{MSE}$$

3. Critical region: treated "as a right-sided..."



4. Decision

* P. value → no * by 2

eg: assume estimating the effect of X_1, X_2, X_3 on Y using a sample of 10 observations. The following result coming from the estimation.

Variable	Coefficient	St. error
intercept "b_0"	4.09	1.44
X_1	10.02	1.65
X_2	.10	.12
X_3	−4.48	1.44

K=4
P=3

ANOVA:

S.O.V	df	SS	MS	F.stat
Regression	K−1 3	360	120	30
Error	n−K 6	24	4	
Total	9	384		

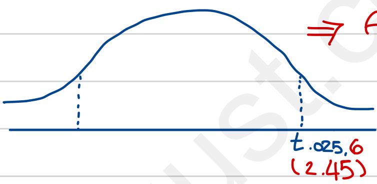
1. Write the estimated regression line
2. Find R^2 3. Test if the coefficient of X_2 is significant ($\alpha = 5\%$)
4. Test if the overall significance of the model ($\alpha = 5\%$)

1. $\hat{Y} = b_0 + b_1x_1 + b_2x_2 + b_3x_3 \rightarrow \hat{Y} = 4.09 + 10.02x_1 + .10x_2 - 4.48x_3$

2. $R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} = \frac{360}{384} = 93.8\%$: meaning that the X_s explain 93.8% of y total variation.

3. $H_0: \beta_2 = 0$, $H_1: \beta_2 \neq 0$

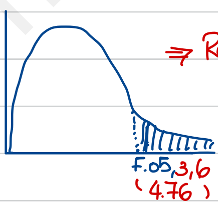
$t.stat = \frac{b_2}{Se\,b_2} = \frac{.1}{.12} = .83$



⇒ Accept the H_0: there's no evidence that β_2 is significant.

4. $H_0: \beta_1 = \beta_2 = \beta_3 = 0$, $H_1: \beta_1 \neq \beta_2 \neq \beta_3 \neq 0$

$F.stat = \frac{MSR}{MSE} = \frac{120}{4} = 30$



⇒ Reject the H_0: There's evidence that at least one of the β has an effect over Y.

Gulf University for Science & Technology
Department of Economics & Finance
ECO-380: Business Statistics

Dr. Khalid Kisswani
Room: N1-116
Tel: 2530-7339
kisswani.k@gust.edu.kw

Assignment 9 (LO vi)

1. In the following estimated regression equation $\hat{y} = b_0 + b_1 x$

a. b_1 is the slope
b. b_1 is the intercept ×
c. b_0 is the slope ×
d. None of these alternatives is correct.

2. A regression analysis between sales (Y in \$) and price (X in \$) resulted in the following equation:
$\hat{Y} = 30 - 4X$. The equation implies that an

a. increase of \$1 in price is associated with an increase of \$4 in sales
b. increase of \$4 in price is associated with an increase of \$1 in sales
c. decrease of \$1 in price is associated with an decrease of \$4 in sales
d. decrease of \$1 in price is associated with an increase of \$4 in sales

3. In a regression and correlation analysis if $r^2 = 1$, then

a. SSE = SST
b. SSE = 1
c. SSR = SSE
d. SSR = SST

4. In a regression analysis, the regression equation is given by y = 12 - 6x. If SSE = 510 and SST = 1000, then the coefficient of correlation is r

a. -0.7
b. +0.7
c. 0.49
d. -0.49

$R^2 = \frac{510}{1000} = .51$

$\therefore r = -\sqrt{.51}$

5. In a regression analysis if SSE = 200 and SSR = 300, then the coefficient of determination is $R^2 = \frac{300}{500}$

a. 0.6667
b. 0.6000
c. 0.4000
d. 1.5000

SST = 500

6. If the coefficient of determination is equal to 1, then the coefficient of correlation

a. must also be equal to 1
b. can be either -1 or +1
c. can be any value between -1 to +1
d. must be -1

7. Regression analysis was applied between demand for a product (Y) and the price of the product (X), and the following estimated regression equation was obtained: $\hat{Y} = 120 - 10X$
Based on the above estimated regression equation, if price is increased by 2 units, then demand is expected to

a. increase by 120 units
b. increase by 100 units
c. increase by 20 units
d. decease by 20 units

$-10(2) = -20$

8. If the coefficient of correlation is 0.8, the percentage of variation in the dependent variable explained by the variation in the independent variable is

r y x

a. 0.80%
b. 80%
c. 0.64%
d. 64%

9. In a regression analysis if SST = 500 and SSE = 300, then the coefficient of determination is
a. 0.20
b. 1.67
c. 0.60
d. 0.40

$R^2 = 1 - \frac{3}{5}$

Exhibit 14-1

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

Y	X	$(X-\bar{X})$	$(Y-\bar{Y})$	$(X-\bar{X})(Y-\bar{Y})$	$(X-\bar{X})^2$
4	2	-1	-1	1	1
3	1	-2	-2	4	4
4	4	1	-1	-1	1
6	3	0	1	0	0
8	5	2	3	6	4
				10	10

SSE = 6. SST = 16

$\bar{Y} = 5$ $\bar{X} = 3$

10. Refer to Exhibit 14-1. The least squares estimate of the slope is
a. 1
b. 2
c. 3
d. 4

$b_1 = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{\Sigma(x-\bar{x})^2} = \frac{10}{10}$

11. Refer to Exhibit 14-1. The least squares estimate of the Y intercept is
a. 1
b. 2
c. 3
d. 4

$b_0 = \bar{Y} - b_1\bar{X} = 5 - 1(3) = 2$

12. Refer to Exhibit 14-1. The coefficient of determination is
a. 0.7096
b. - 0.7906
c. 0.625
d. 0.375

$R^2 = 1 - \frac{6}{16} =$

13. Refer to Exhibit 14-1. The coefficient of correlation is
a. 0.7906
b. - 0.7906
c. 0.625
d. 0.375

$r = -\sqrt{.625}$

14. Refer to Exhibit 14-1. The MSE is
a. 1
b. 2
c. 3
d. 4

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

$\hat{y}$ 4 0 0 0 1

$\hat{y} = 6 - 1x$

Exhibit 14-2

You are given the following information about y and x.

1 4 3 2 1

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

$\bar{Y} = 3$
$\bar{X} = 3$

$= -10$ $= -1$
$= 10$

15. Refer to Exhibit 14-2. The least squares estimate of b_1 (slope) equals
a. 1
b. -1
c. 6
d. 5

$\frac{-10}{10}$

16. Refer to Exhibit 14-2. The least squares estimate of b_0 (intercept)equals
a. 1
b. -1
c. 6
d. 5

3+1(3)

17. Refer to Exhibit 14-2. The point estimate of y when x = 10 is
a. -10
b. 10
c. -4 ?
d. 4

18. Refer to Exhibit 14-2. The sample correlation coefficient equals
a. 0
b. +1
c. -1
d. -0.5

$R^2 =$ ____

19. Refer to Exhibit 14-2. The coefficient of determination equals
a. 0
b. -1
c. +1
d. -0.5

Exhibit 14-4

Regression analysis was applied between sales data (Y in \$1,000s) and advertising data (x in \$100s) and the following information was obtained: $\hat{Y} = 12 + 1.8$ x. n = 17, SSR = 225, SSE = 75, $S_{b1} = 0.2683$

20. Refer to Exhibit 14-4. Based on the above estimated regression equation, if advertising is \$3,000, then the point estimate for sales (in dollars) is
a. \$66,000
b. \$5,412
c. \$66
d. \$17,400

12 + 1.8 (3,000)
= 5,412

21. Refer to Exhibit 14-4. The F statistic computed from the above data is
a. 3
b. 45
c. 48
d. 50

22. Refer to Exhibit 14-4. The t statistic for testing the significance of the slope is
a. 1.80
b. 1.96
c. 6.708
d. 0.555

23. Refer to Exhibit 14-4. The critical t value for testing the significance of the slope at 95% confidence is
a. 1.753
b. 2.131
c. 1.746
d. 2.120

24. A multiple regression model has
a. only one independent variable
b. more than one dependent variable
c. more than one independent variable
d. at least 2 dependent variables

Exhibit 15-1

In a regression model involving 44 observations, the following estimated regression equation was obtained.

$\hat{Y} = 29 + 18X_1 + 43X_2 + 87X_3$. For this model SSR = 600 and SSE = 400.

25. Refer to Exhibit 15-1. The coefficient of determination for the above model is
a. 0.667
b. 0.600
c. 0.336
d. o.400

26. Refer to Exhibit 15-1. MSR for this model is
a. 200
b. 10
c. 1,000
d. 43

27. Refer to Exhibit 15-1. The computed F statistics for testing the significance of the above model is
a. 1.500
b. 20.00
c. 0.600
d. 0.6667

Exhibit 15-6

Below you are given a partial computer output based on a sample of 16 observations.

	Coefficient	Standard Error
Constant	12.924	4.425
X_1	-3.682	2.630
X_2	45.216	12.560

Analysis of Variance

3
k-1 (p)

Source of Variation	**Degrees of Freedom**	**Sum of Squares**	**Mean Square**	**F**
Regression	2	4,853	2,426.5	
Error	13		485.3	5

$485.3 = \frac{SSE}{n-k}$

16-3

28. Refer to Exhibit 15-6. The estimated regression equation is
a. $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$
b. $E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2$
c. $\hat{Y} = 12.924 - 3.682X_1 + 45.216X_2$
d. $\hat{Y} = 4.425 + 2.63X_1 + 12.56X_2$

29. Refer to Exhibit 15-6. The interpretation of the coefficient of X_1 is that
a. a one unit change in X_1 will lead to a 3.682 unit decrease in Y ✓
b. a one unit increase in X_1 will lead to a 3.682 unit decrease in Y when all other variables are held constant
c. a one unit increase in X_1 will lead to a 3.682 unit decrease in X_2 when all other variables are held constant
d. It is impossible to interpret the coefficient.

30. Refer to Exhibit 15-6. We want to test whether the parameter β_1 is significant. The test statistic equals
a. -1.4
b. 1.4
c. 3.6
d. 5

$$\frac{b_1}{s_{b_1}} = \frac{-3.682}{2.630} =$$

31. Refer to Exhibit 15-6. The t value obtained from the table which is used to test an individual parameter at the 1% level is
a. 2.65
b. 2.921
c. 2.977
d. 3.012

32. Refer to Exhibit 15-6. Carry out the test of significance for the parameter β_1 at the 1% level. The null hypothesis should be
a. rejected
b. not rejected
c. revised
d. None of these alternatives is correct.

33. Refer to Exhibit 15-6. The degrees of freedom for the sum of squares explained by the regression (SSR) are
a. 2
b. 3
c. 13
d. 15

34. Refer to Exhibit 15-6. The sum of squares due to error (SSE) equals
a. 37.33
b. 485.3
c. 4,853
d. 6,308.9

35. Refer to Exhibit 15-6. The test statistic used to determine if there is a relationship among the variables equals
a. -1.4
b. 0.2
c. 0.77
d. 5

36. Refer to Exhibit 15-6. The F value obtained from the table used to test if there is a relationship among the variables at the 5% level equals
a. 5.10
b. 3.89
c. 3.74
d. 4.86

37. Refer to Exhibit 15-6. Carry out the test to determine if there is a relationship among the variables at the 5% level. The null hypothesis should
a. be rejected
b. not be rejected
c. revised
d. None of these alternatives is correct.