Example Exam Questions
Unit #3 ; Page 1 ; Problem 1 of 1

Glue

Claim: $\mu_{New} \geq \mu_{old} + 10$ (or =)

$H_0: (\mu_{New} - \mu_{old}) \geq 10$

$H_1: (\mu_{New} - \mu_{old}) < 10$

$\alpha = 0.10$ left tail

Note: "variation is about the same for both glues."

so: pool variances and add deg. of freedom

$$S^2_{pool} = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{(n_1-1)+(n_2-1)}$$

$$= \frac{(13)(10.3)^2 + 10(9.0)^2}{13 + 10}$$

$$= 2189.17/23 = 95.2$$

d.f. = 23

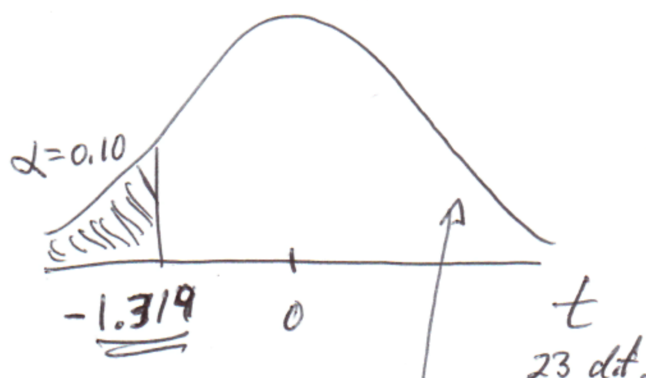
Test statistic

$$\frac{(\bar{x}_{New} - \bar{x}_{old}) - (\mu_{New} - \mu_{old})_0}{\sqrt{\frac{S^2_{pool}}{n_{New}} + \frac{S^2_{pool}}{n_{old}}}}$$

$$= \frac{(201.0 - 178.4) - 10}{\sqrt{\frac{95.2}{14} + \frac{95.2}{11}}}$$

$$= \frac{12.6}{3.93} = 3.206$$

critical Region



Do not reject H_0:

Page 2 — Problem 1 of 2

(8 points : 8 minutes)

Use the results of the experiment below to test whether the population correlation (ρ) is negative. If you cannot figure out how to get the sample correlation coefficient quickly, use r = - 0.32. (For this test, use $\alpha = 0.025$).

correlation test

$H_0: \rho \geq 0$

$H_1: \rho < 0$

Experiment results:

regression line:

intercept =	0.0	554.1
slope =	0.00	- 3.10
S_e =	0.00	63.87
n =	24	24
S_Y =	0.00	75.26

d.f. = n-2 = 22

$S_y = \sqrt{\dfrac{\text{Total Variation}}{n-1}}$ ∴ $S_y^2(n-1) = \text{Total}$

$(75.26)^2(23) = 130274$

$S_e = \sqrt{\dfrac{\text{unexplained}}{n-2}}$ ∴ $S_e^2(n-2) = \text{unexplained}$

$(63.87)^2(22) = 89746$

Explained = Total - unexplained = 40528

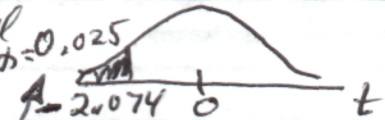
$r^2 = \text{Explained}/\text{total} = \dfrac{40528}{130274} = 0.311$

$r = \sqrt{r^2} = \sqrt{0.311} = -0.558$ Negative because slope is negative.

test statistic: $\dfrac{r}{\sqrt{\dfrac{1-r^2}{n-2}}} = \dfrac{-0.558}{\sqrt{\dfrac{1-(-.558)^2}{24-2}}}$

$= \dfrac{-0.558}{0.1769} = -3.154$

Critical Region $\alpha = 0.025$



Reject H_0:

Problem 2 of 2

(8 points : 8 minutes)

9. A company makes complicated laboratory equipment for analyzing chemical samples. To learn about the performance of their machines, the company works with 9 laboratories and gives to each four (4) identical sample of material to analyze, so a total of 36 measurements are taken.

$9 \times 4 = 36 = N$

Variability in the outcomes of all 36 tests represents differences between laboratories (laboratory is considered the "treatment") and differences from test to test within the same laboratory ("error"). Complete the Analysis of Variance table below and carry out the appropriate hypothesis test to decide whether the expected (mean) results are the same for all 9 laboratories. $k = 9$ $df(\text{lab's}) = 8$

(Use a significance level of 0.04 for this test.)

$\alpha = 0.04$

$H_0: \mu_1 = \mu_2 = \cdots = \mu_9$

H_1: Not H_1: OR "at least one of the means ≠ another"

Analysis of Variance Table

Source	Sum of Squares	Degrees of Freedom	Mean Square	F	p-value
Laborator	377.6	8	47.2	2.6077	0.0297
Error	488.7	27	18.1		
Total	866.3	35			

p-value < α, so reject H_0:

N = 36

df(total) = 35

$F = \dfrac{MS(\text{labs})}{MS(\text{Error})} = \dfrac{47.2}{18.1} = 2.6077$

$SS(\text{labs}) = MS(\text{labs}) * df(\text{labs}) = (47.2)(8) = 377.6$

$SS(\text{Error}) = SS(\text{Total}) - SS(\text{labs}) = 866.3 - 377.6 = 488.7$

$MS(\text{Error}) = SS(\text{Error}) / D.f.(\text{Error}) = 488.7/27 = 18.1$

Example Exam Questions
Unit #3; Page 3; Problem 1 of 2

Correlation

Put data in calculator:

$r = 0.6711$

$n = 5$

$d.f. = n - 2 = 3$

Claim: correlation between x and y is positive.

$\rho > 0$

$H_0: \rho \le 0$

$H_1: \rho > 0$

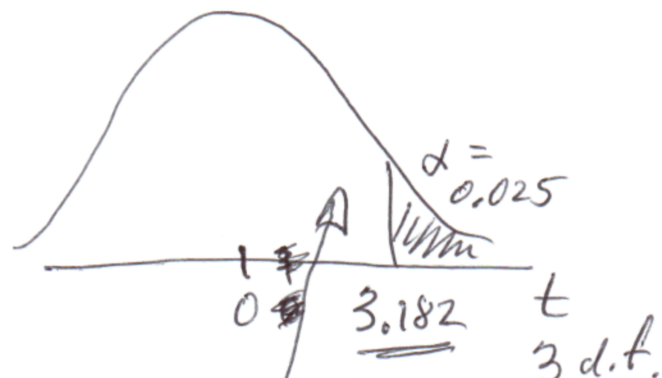
$\alpha = 0.025$ right tail

Test Statistic

$$\frac{r}{\sqrt{\dfrac{1-r^2}{n-2}}}$$

$$= \frac{0.6711}{\sqrt{\dfrac{1-(0.6711)^2}{5-2}}}$$

$$= \frac{0.6711}{0.4280} = 1.568$$



Do Not reject H_0.

Example Exam Problems
Unit #3; Page 3; Problem 2 of 2

$95\%\ CI\ (\mu_{NEW} - \mu_{OLD})$

Note: Problem talks about a "claim", but this is not a hypothesis test. The problem requires a confidence interval, not a test. Also, it says to answer "the question below", but I did not copy that question.

What about variation: "Experts advise ... strength with old glue appears to be <u>more variable</u> than with the new glue."

So: Do <u>not</u> pool variances. Do <u>not</u> add the d.f., but use the smaller of the two sample d.f.

$$95\%\ CI\ (\mu_{new} - \mu_{old}) = (\bar{x}_{new} - \bar{x}_{old}) \pm t_{\alpha/2}\sqrt{\frac{s^2_{new}}{n_{new}} + \frac{s^2_{old}}{n_{old}}}$$

smaller $n = 8$
$df = 7$

confidence $= 0.95$
$\alpha = 0.05$ in 2 tails
$\alpha/2 = 0.025$ in 1 tail

$t = 2.365$

$$= (215 - 198.4) \pm 2.365\sqrt{\frac{(4.4)^2}{8} + \frac{(9.3)^2}{15}}$$

$$= 16.6 \pm (2.365)(2.861)$$

$$= 16.6 \pm 6.77$$

$$\left[\ 9.83 < (\mu_{new} - \mu_{old}) < 23.37\ \right]$$

Example Exam Problems
Unit #3; Page 4; Problem 1 of 2

$\boxed{\text{Matched Pairs: Experimental Unit = Patient}}$

follow order in the hypotheses.

patient	Tablet	Liquid	diff (Tab - Liq)
1			2.2
2			0.7
3			3.4
4			1.5
5			0.3
6			3.4

$n = 6$
$\bar{d} = 1.917$
$S_d = 1.323$

$df = 5$

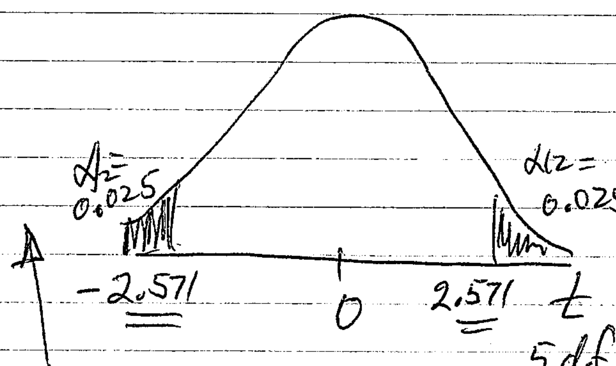
$\mu_d = (\mu_{Tab} - \mu_{Liq})$

claim: $(\mu_{Tab} = \mu_{Liq} + 5)$ or $\geq$

$H_0: (\mu_T - \mu_L) = 5$

$H_1: (\mu_T - \mu_L) \neq 5$ or $<$

$\alpha = 0.05$ in 2 tails



$\boxed{\text{Test Statistic}}$

$$\frac{\bar{d} - \mu_d}{S_d/\sqrt{n}} = \frac{1.917 - 5}{1.323/\sqrt{6}} = \frac{-3.083}{0.54} = -5.709$$

Reject Ho:

Example Exam Questions
Unit #3; Page 4; Problem 2 of 2

Confidence Interval for difference between two proportions

Data:

$$98\% \, CI(p_B - p_A) = (\hat{p}_B - \hat{p}_A) \pm Z_{\alpha/2} \sqrt{\frac{\hat{p}_A \hat{q}_A}{N_A} + \frac{\hat{p}_B \hat{q}_B}{N_B}}$$

	Good	Bad	total
Formula A	387	13	400
Formula B	354	46	400

confidence = 0.98
$\alpha = 1 - \text{confidence}$
$= 1 - 0.98 = 0.02$
$\alpha/2 = 0.01$
$Z_{\alpha/2} = 2.33$

$\hat{p}_A = \frac{13}{400} = 0.0325$ $\qquad \hat{q}_A = 0.9675$

$\hat{p}_B = \frac{46}{400} = 0.115$ $\qquad \hat{q}_B = 0.885$

$$98\% \, CI(p_B - p_A) = (0.115 - 0.0325) \pm 2.33 \sqrt{\frac{(.0325)(.9675)}{400} + \frac{(0.115)(0.885)}{400}}$$

$$= 0.0825 \pm (2.33)(0.0182)$$

$$= 0.0825 \pm 0.0424$$

$$= [0.040 < (p_B - p_A) < 0.125]$$

Example Exam Questions
Unit #3; Page 5; 1 of 2

Contingency Table

Minutes Needed before Sleep

Daily Alcohol	0-10	10-20	20-30	total
none	117 / 96.33	269 / 249.3	114 / 154.3	500
1 to 2 drinks	90 / 96.33	240 / 249.3	170 / 154.3	500
3 or more	82 / 96.33	239 / 249.3	179 / 154.3	500
Total	289	748	463	1500

$$\text{Expected} = \frac{(\text{row tot})(\text{col tot})}{\text{Grand Total}}$$

$df = (\text{rows}-1)(\text{cols}-1)$
$= (3-1)(3-1)$
$= (2)(2) = 4$

H_0: Medication effectiveness is independent of alcohol use

H_1: dependent (not H_0)

$\alpha = 0.05$ (Not stated) right tail

$$\sum_{\text{all cells}} \left[\frac{(\text{Obs} - \text{Exp})^2}{\text{Exp}} \right] \to 10.53$$

$(O-E)^2/E$

		10.53

Reject H_0

points : 10 minutes)

8. **A company makes a sleep aid medication. They are concerned that alcohol use may interfere with the medications effectiveness, so that people who drink take longer to fall asleep. Use the data below to test whether the time needed to fall asleep after taking the medication is independent of a person's level of alcohol use.**

Amount of daily alcohol use	Minutes needed before sleep 0-10	10-20	20-30	Total
none	117	269	114	500
1 to 2 drinks	90	240	170	500
3 or more drinks	82	239	179	500
Total	289	748	463	1500

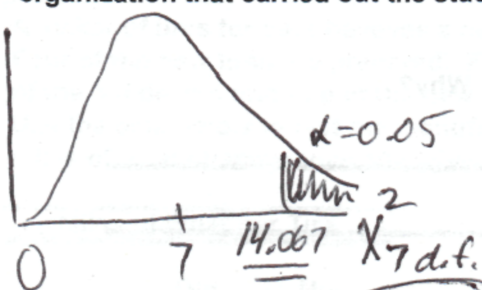
Solution on separate page (previous)

✓

Page 5 ; Problem 2 of 2 (Goodness-of-Fit)

(7 points : 10 minutes)

10. **A major news organization is interested in the public issues that registered voters think are most important. A stratified random sample of 320 registered voters is selected to represent the whole population of voters. Each voter is asked to select from a list of 8 issues the one that is most important. Compare the results to see if they are significantly different from the proportions expected by the news organization that carried out the study.** (Let α be 0.05 for the test.)



H0: Proportions of registered voters that rank isses most important = "expected"

H0: ______________________

Issue	Proportions Expected	Proportions In Sample	COUNT Observed	COUNT Expected	$\frac{(OBS-EXP)^2}{Exp}$
Traffic Congestion	10%	30%	96	32	128
Pollution	10%	5%	16	32	8
Taxes	10%	25%	80	32	72
Deficits	10%	5%	16	32	8
Death Penalty	5%	5%	16	16	0
Iraq War	25%	20%	64	80	3.2
Education	10%	5%	16	32	8
Health Care	20%	5%	16	64	36

$k = 8$

$df = 7$

$N = 320$

$\sum \left[\frac{(O-E)^2}{E}\right] = 263.2$

$263.2 > 14.067$, so Reject Ho:

Example Exam Questions
Unit #3; Page 5; Problem 1 of 2

Confidence Interval for difference between two means

Note: Problem says "variation is the same for both populations"; so ① pool the variances and ② add degrees of freedom

Data:

$\bar{x}_A = 111.7$	$\bar{x}_B = 103.5$
$S_A = 28.2$	$S_B = 17.9$
$n_A = 9$	$n_B = 6$

$df_A = 8$ $\quad df_B = 5$

$d.f. = 8+5 = 13$

$$S^2_{pool} = \frac{(n_A-1)S_A^2 + (n_B-1)S_B^2}{(n_A-1)+(n_B-1)} = \frac{(9-1)(28.2)^2 + (6-1)(17.9)^2}{(9-1)+(6-1)} = \frac{7963.97}{13} = 612.6$$

$$90\%\ CI(\mu_A - \mu_B) = (\bar{x}_A - \bar{x}_B) \pm t_{\alpha/2}\sqrt{\frac{S^2_{pool}}{n_A} + \frac{S^2_{pool}}{n_B}}$$

$\alpha = 0.10 = 1 - 90\%$
in 2 tails; 13 d.f.
$t = 1.771$

$$= (111.7 - 103.5) \pm 1.771\sqrt{\frac{612.6}{9} + \frac{612.6}{6}}$$

$$= 8.2 \pm (1.771)(13.04)$$

$$= 8.2 \pm 23.1$$

$$= \left[-14.9 < (\mu_A - \mu_B) < 31.3\right]$$

Q: Is claim that $\mu_B = 102$ and $\mu_A = 113$ reasonable?

A: Yes, because $(113 - 102) = 11$, and 11 is in the $CI(\mu_A - \mu_B)$ which is the "reasonable range" for $\mu_A - \mu_B$.

Example Exam Questions
Unit #3; Page 6; Problem 2 of 2

$\boxed{\text{correlation test}}$

Put data for Net Worth and Happiness into Calculator.

Get $\boxed{r = 0.372}$

Claim: correlation is positive
$\rho > 0$

$H_0: \rho \le 0$
$H_1: \rho > 0$

$\alpha = 0.05$ right tail

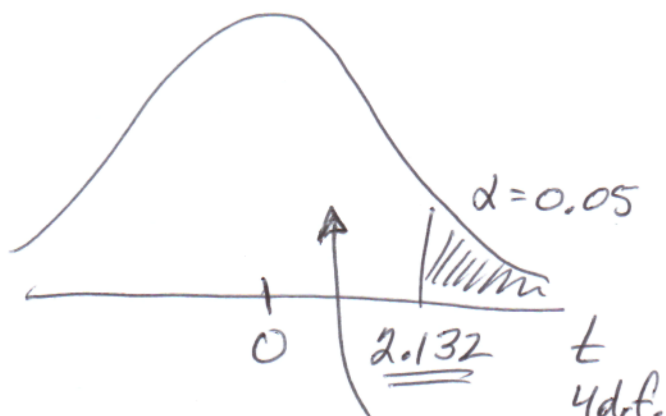
$n = 6$
$d.f. = n-2 = 4$

$t = 2.132$

Test Statistic

$$\frac{r}{\sqrt{\frac{1-r^2}{n-2}}}$$

$$= \frac{0.372}{\sqrt{\frac{1-(0.372)^2}{4}}} = \frac{0.372}{0.4641} = 0.802$$



Do Not reject H_0:

Page 7; Problem 1 of 2 — contingency table

(8 points; 10 minutes)

5. **Use the information in the contingency table to decide whether or not to reject the claim that Factor A and Factor B are independent. Let $\alpha = 0.05$ for this test.**

Expected = (row total)(col total) ÷ (grand total)

Level of Factor B	Level of Factor A: 1	2	3	Row Total
1	54.7 / 66	79.7 / 68	65.7 / 66	200
2	54.7 / 55	79.7 / 83	65.7 / 62	200
3	54.7 / 43	79.7 / 88	65.7 / 69	200
Column Total	164	239	197	600

$(O-E)^2/E$

2.3	1.7	0
0	0.14	0.21
2.5	0.9	0.97

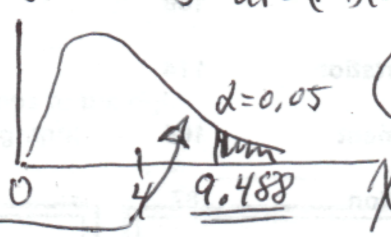
$$\sum \left[\frac{(O-E)^2}{E}\right] = 7.92$$

Claim: Factor A and Factor B are independent

H_0: A and B are independent

H_1: A and B are dependent

$\alpha = 0.05 \quad df = (r-1)(c-1) = 4$



Do not Reject H_0.

(10 points : 10 minutes)

10. **A maker of tires for cars believes a new design will wear longer than the current design. Four of the new tires are prepared. Four cars are used in an experiment where one tire of the old design and one of the new design are used on the front wheels of each car. Use the data below to test the manufacturer's claim that the new design will increase the miles of wear by more than 500 miles.** (Use a 0.10 significance level for the test.)

Miles of Wear per Tire

Car	Old Design	New Design
1	58500	59100
2	60100	60700
3	58500	59200
4	63400	63800

solution on separate page (next)

✓

Example Exam Questions
Unit#3; Page 7; Problem 2 of 2

Matched Pairs test of
$\mu_{new} - \mu_{old}$

matched because only 4 cars were used, and one old design tire and one new design tire were used on each car.

car	Old Design	New Design	(New−old) diff.
1	58500	59100	600
2	60100	60700	600
3	58500	59200	700
4	63400	63800	400

$\boxed{\alpha = 0.10 \quad df = 3 \quad t = 1.638}$

$(\bar{x}) = \bar{d} = 575$

$(s_x) = s_d = 125.8$

$n = 4$

$df = 3$

claim: $\mu_{new} > \mu_{old} + 500$

$(\mu_{new} - \mu_{old}) > 500$

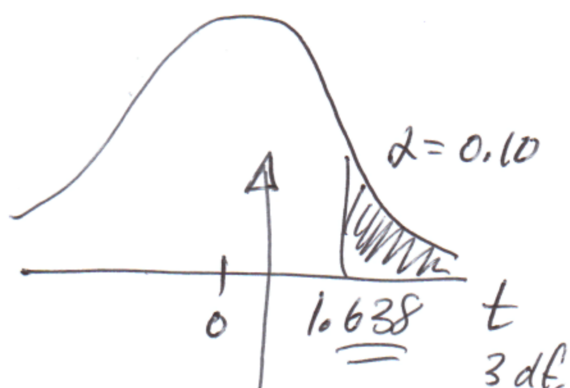
$H_0: (\mu_{new} - \mu_{old}) \le 500$

$H_1: (\mu_{new} - \mu_{old}) > 500$

$\alpha = 0.10$ right tail

Test Statistic

$$\frac{\bar{d} - (\mu_d)_0}{s_d/\sqrt{n}}$$



$$= \frac{575 - 500}{125.8/\sqrt{4}} = \frac{75}{62.9} = 1.192$$

Do Not reject H_0.

Example Exam Questions
Unit #3; Page 8; Problem 1 of 2

$\boxed{\text{Goodness of Fit Problem}}$

H_0: all categories are equally most important

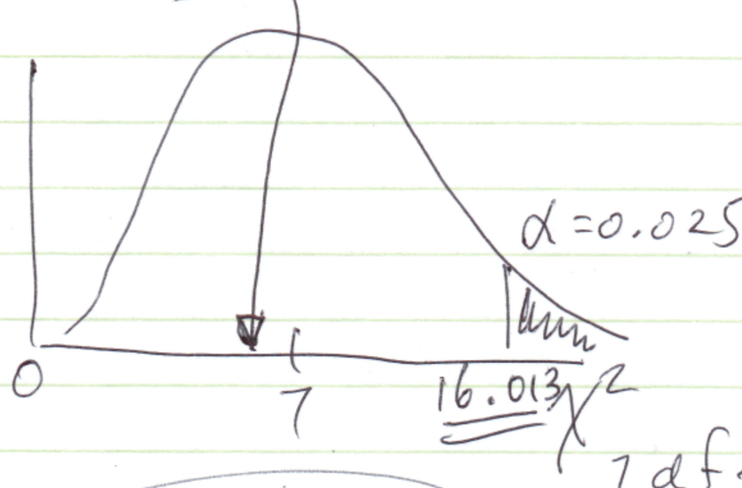
H_1: not so!

$\alpha = 0.025$; right tail

Issue	OBS counts	Exp	$\frac{(O-E)^2}{E}$
Taxes	87	100	1.69
Educ	105	100	0.25
Security	96	100	0.16
Poverty	108	100	0.64
transp.	114	100	1.96
Environ.	104	100	0.16
Immig.	87	100	1.69
SocSec/Med	99	100	0.01
	$\boxed{800}$		$\Sigma = 6.56$

$k = 8 \quad df = 7$

$Exp = \left(\frac{1}{8}\right) 800 =$



Do not reject H_0.

Example Exam Problems
Unit # 3; Page 8; Problem 2 of 2

Confidence Interval for difference of two proportions

"Hi" = High value homes
"Lo" = Low value homes

$$95\%\ CI(p_{Hi} - p_{Lo}) = (\hat{p}_{Hi} - \hat{p}_{Lo}) \pm Z_{\alpha/2}\sqrt{\frac{\hat{p}_{Hi}\hat{q}_{Hi}}{N_{Hi}} + \frac{\hat{p}_{Lo}\hat{q}_{Lo}}{N_{Lo}}}$$

$$\boxed{\begin{matrix} \hat{p}_{Hi} = 0.12 & \hat{q}_{Hi} = 0.88 & N_{Hi} = 50 \\ \hat{p}_{Lo} = 0.174 & \hat{q}_{Lo} = 0.826 & N_{Lo} = 46 \end{matrix}}$$

$$\boxed{Z_{\alpha/2} = 1.96}$$

$$95\%\ CI(p_{Hi} - p_{Lo}) = (0.12 - 0.174) \pm 1.96\sqrt{\frac{(.12)(.88)}{50} + \frac{(.174)(.826)}{46}}$$

$$= (-0.054) \pm (1.96)(0.0724)$$

$$= (-0.054) \pm 0.142$$

$$= [-0.196 < (p_{Hi} - p_{Lo}) < 0.088]$$

Example Exam Questions
Unit #3; Page 9; Problems 1, 2, & 3

Match graphs to possible correlations

#8. r = 0.070   r = -0.85   r = 0.00

#9. r = NONE   r = 0.90   r = 1.00

#10. r = -0.70   r = NONE   r = 0.90

Example Exam Questions
Unit #3; Page 10; Problems 1 and 2.

$\boxed{\text{correlation test}}$

Put data in calculator to get:

$r = -0.694 \quad n = 6$

$\boxed{df = n-2 = 4}$

$t = -3.747$

claim: correlation is Negative

$\rho < 0$

$H_0: \rho \geq 0$

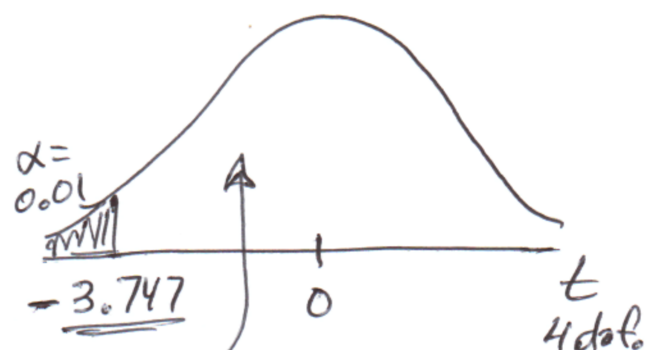
$H_1: \rho < 0$

$\alpha = 0.01$ left tail

Test Statistic

$$\frac{r}{\sqrt{\frac{1-r^2}{n-2}}} = \frac{-0.694}{\sqrt{\frac{1-(-0.694)^2}{6-2}}}$$

$$= \frac{-0.694}{0.360} = -1.928$$



Do not reject H_0.

What if n was 28 not 6, and r was still -0.694? and $t = -2.479$

$$\frac{r}{\sqrt{\frac{1-r^2}{n-2}}} = \frac{-0.694}{\sqrt{\frac{1-(-0.694)^2}{28-2}}} = \frac{-0.694}{0.1412} = -4.915$$

Reject H_0.

Page 11 ; Problem 1 of 1

(10 points - 15 minutes)

3. **The proportions of people in the U.S. that prefer 5 different kinds of entertainment are shown in the table below. A local survey of 600 people found 60 people who prefer movies, 300 who prefer to watch TV, 90 who like to listen to music, 30 who prefer dancing, and 120 that prefer to play sports. Test the claim that the true local proportions are the same as the national rates.** (Use a 0.05 significance level for the test)

$\boxed{\text{Goodness-of-Fit}}$

Claim: Local proportions = National Proportions

H_0: Local proportions = National Proportions

H_1: No they do not ; Not H_0 ; etc.

$\alpha = 0.05$

Preferred Entertainment	National Rates	Local Survey Count (Observed)	Expected	$(Obs-Exp)^2/Exp$
Watching Movies	0.15	60	90 = (0.15)600	10
Watching Television	0.35	300	210 = (0.35)600	38.57
Listening to Music	0.10	90	60	15
Dancing	0.10	30	60	15
Playing Sports	0.30	120	180	20
Survey Total =		600		

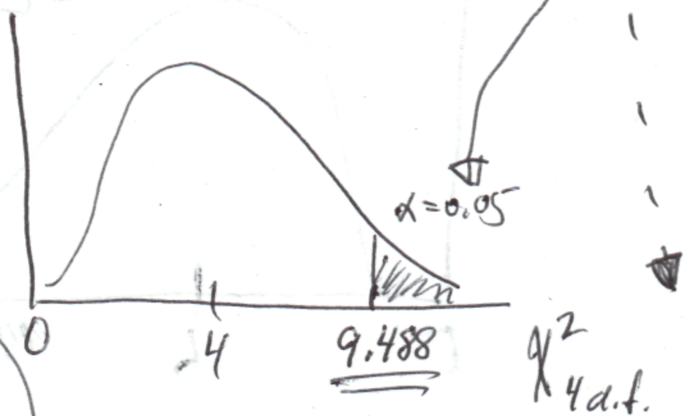
5 categories

$k = 5$

$d.f. = k - 1 = 4$

$$\sum\left[\frac{(O-E)^2}{E}\right] = 98.57$$

Test statistic formula required



The first cell in the table is enough to reject H_0:

Reject H_0:

(10 points - 20 minutes) Page 12; Problem 1 of 1

4. Use the data in the table to test the idea that the use of some "slang" terms is independent of age. The data represent a stratified random sample of 400 people from Los Angeles.

(Use $\alpha = 0.025$ for this test)

$\boxed{\text{Contingency Table}}$

Expected counts $= \dfrac{(\text{row total})(\text{col. total})}{\text{grand total}}$

$\boxed{\text{Expected counts are written in each cell.}}$

Most used Slang Term	Age Group 10 to 20	21 to 40	41 to 60	> 60	Total
"I'm like ... "	37.5 \ 88	37.5 \ 50	37.5 \ 10	37.5 \ 2	150
"totally"	25 \ 10	25 \ 40	25 \ 40	25 \ 10	100
"far out"	37.5 \ 2	37.5 \ 10	37.5 \ 50	37.5 \ 88	150
Total	100	100	100	100	400

Claim: Slang used is independent of Age

H0: Slang and Age are indep.

H1: Dependent; not Ho:; etc.

$\alpha = 0.025$

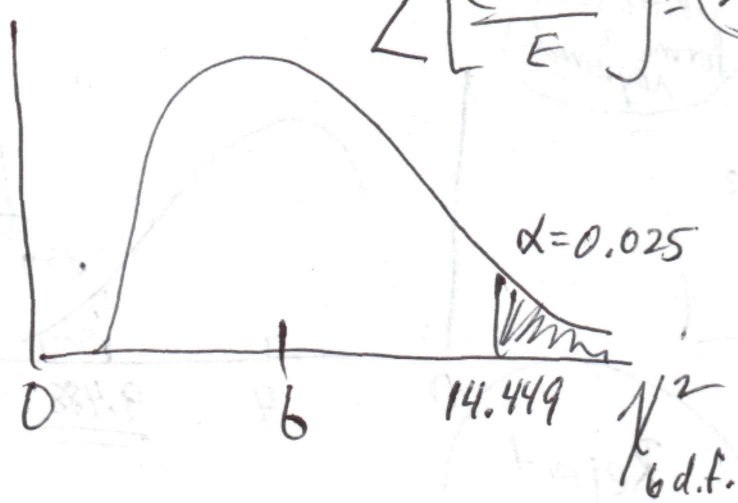
$(\text{Obs.} - \text{Exp.})^2 / \text{Exp.}$

68.0	4.2	20.2	33.6
9.0	9.0	9.0	9.0
33.6	20.2	4.2	68.0

This one cell is big enough. It is OK to stop

$$df = (\text{rows}-1)(\text{cols}-1)$$
$$= (3-1)(4-1) = (2)(3) = 6$$

$$\sum\left[\frac{(O-E)^2}{E}\right] = 288$$



Reject Ho:

Example Exam Questions
Unit #3; Page 13; Problem 1 of 2

$\boxed{\text{Confidence Interval for the difference between two proportions}}$

	Cream A	Cream B
Less Pain	38	51
Not less Pain	12	9
TOTAL	50	60

$\hat{p}_A = \frac{38}{50} = 0.76$

$\hat{q}_A = 0.24$

$n_A = 50$

$\hat{p}_B = 0.85$

$\hat{q}_B = 0.15$

$n_B = 60$

$$\boxed{\begin{aligned} \text{Confidence} &= 0.98 \\ \alpha &= 0.02 \\ \alpha/2 &= 0.01 \\ Z_{\alpha/2} &= 2.33 \end{aligned}}$$

$$98\%\ CI(p_B - p_A) =$$

$$(\hat{p}_B - \hat{p}_A) \pm Z_{\alpha/2}\sqrt{\frac{\hat{p}_A\hat{q}_A}{n_A} + \frac{\hat{p}_B\hat{q}_B}{n_B}}$$

$$= (0.85 - 0.76) \pm 2.33\sqrt{\frac{(.76)(.24)}{50} + \frac{(.85)(.15)}{60}}$$

$$= 0.09 \pm (2.33)(0.07598)$$

$$= 0.09 \pm 0.177$$

$$= [-0.087 < (p_B - p_A) < 0.267]$$

Example Exam Problems
Page 13; Unit #3; Problem 2 of 2

$\boxed{\text{Matched Pairs}}$

	Pain ~~Level~~ Reported Using Cream		diff
Person	A	B	B−A
1	7	9	2
2	6	5	−1
3	9	7	−2
4	1	2	1
5	4	5	1

$0.2 = \bar{d}$
$1.64 = s_d$
$5 = n$

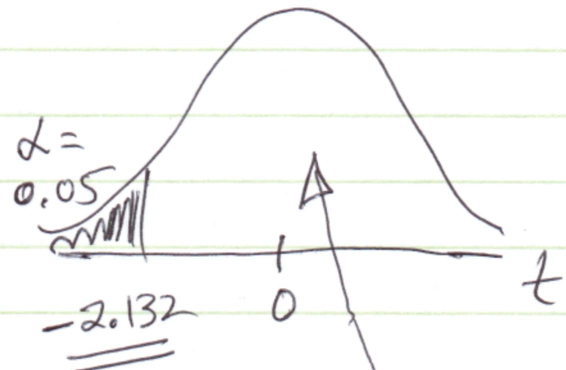
df = 4

Claim: B is Better than A
[but "better" means less pain]
$\mu_B < \mu_A$

$H_0: \mu_B - \mu_A \geq 0$

$H_1: (\mu_B - \mu_A) < 0$

$\alpha = 0.05$ left tail



Do not reject H_0.

Test statistic

$$\frac{\bar{d} - \mu_d}{s_d/\sqrt{n}} = \frac{0.2 - 0}{1.64/\sqrt{5}} = \frac{0.2}{0.733} = 0.273$$

(8 points; 8 minutes) Page 14; Problem 1 of 2

2. **The popularity of TV shows is important to advertisers. A random sample of 1500 TV viewers in California was studied with the results shown below. Use these results to decide whether the popularity of the selected TV shows is the same in CA and NY or different.**
(Use a Type I error rate of 0.01 to make your decision.)

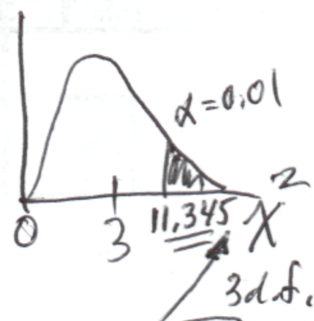
H_0: CA props = NY props
H_1: " $\neq$ "
$\alpha = 0.01$

Goodness - of - Fit

Popularity of Selected TV Shows				
	Monday 6 p.m. Audience			
Show	**Share in NY**	**Viewers in CA sample** (OBS)	Expected	$\frac{(O-E)^2}{E}$
Lost	**31%**	**426**	465	3.27
American Idol	**26%**	**414**	390	1.48
Boston Legal	**19%**	**333**	285	8.08
Friends	**24%**	**327**	360	3.03

$k = 4$; d.f. $= 3$ $N = 1500$

$$\sum\left[\frac{(O-E)^2}{E}\right] = 15.86$$



Reject H_0.

(8 points; 8 minutes) Problem 2 of 2

3. **The popularity of TV shows is important to advertisers. A random sample of 1600 TV viewers in California was studied with the results shown below. Use these results to decide whether the selected TV shows are equally popular.**
(Use a Type I error rate of 0.01 to make your decision.)

Goodness-of-Fit

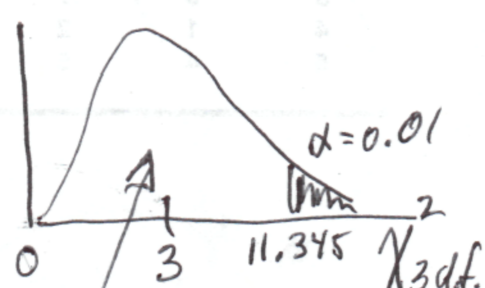
H_0: All four shows are equally popular

H_1: Not all equally popular

Popularity of Selected TV Shows Monday 6 p.m. Audience			
Show	**CA sample** (OBS.)	Expected	$\frac{(O-E)^2}{E}$
Lost	**416**	400	0.64
American Idol	**414**	400	0.49
Boston Legal	**383**	400	0.72
Friends	**387**	400	0.42

$k = 4$; $df = 3$ $N = 1600$

$$\sum\left[\frac{(O-E)^2}{E}\right] = 2.27$$



Do Not reject H_0.

(8 points; 12 minutes) Page 15; Problem 1 of 2

4. Random samples of 500 women and 500 men were asked to evaluate their job satisfaction as "Excellent," "High," "Moderate," or "Low." Use the data below to test the claim that women and men have the same proportions of these ratings. (Use α = 10% for this test.)

Job Satisfac	Gender Female	Male	Row Total
Excellen	E 68 O 64 $(O-E)^2/E$ 0.24	68 72 0.24	136
High	114.5 103 1.16	114.5 126 1.16	229
Moderat	189.5 201 0.70	189.5 178 0.70	379
Low	128 132 0.13	128 124 0.13	256
Col. Tot	500	500	1000

$df = (rows-1)(cols-1)$
$= (4-1)(2-1) = 3$

Claim: Men and women have these job satisfaction levels in (=) proportions

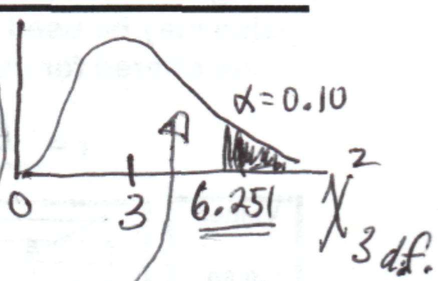
H0: Job Satisfaction levels for men and for women are the same

H1: Not so!

$\alpha = 0.10$

$$\sum\left[\frac{(O-E)^2}{E}\right] = 4.46$$

Do Not Reject H_0!



(8 points; 8 minutes) Problem 2 of 2

1. A study of different "rehabilitation" programs needed the participation of ~~500~~ 1000 inmates from California prisons. A randomized list of all prisoners was prepared and the first ~~250~~ 500 names on the list were assigned to rehab method #1 and the rest were assigned to rehab method #2. The rates of recidivism (later conviction and return to prison for another crime) were studied. Use the results below to make a 90% confidence interval for the difference between the recidivism rates for the two methods.

proportions

	Rehabilitation Method Method 1	Method 2
Returned to Prison	112	175
Did not return to Prison	388	325

$$CI(p_1-p_2) = (\hat{p}_1-\hat{p}_2) \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{N_1}+\frac{\hat{p}_2\hat{q}_2}{N_2}}$$

$$= (0.224-0.35) \pm 1.645\sqrt{\frac{(.224)(.776)}{500}+\frac{(.35)(.65)}{500}}$$

$$= (-0.126) \pm 0.047$$

$$\boxed{-0.173 < (p_1-p_2) < -0.079}$$

$\alpha = 0.10$
$\alpha/2 = 0.05$
$z_{\alpha/2} = 1.645$

$N_1 = 500$ $\quad N_2 = 500$

$\hat{p}_1 = 112/500 = 0.224$ $\quad \hat{p}_2 = 175/500 = 0.35$

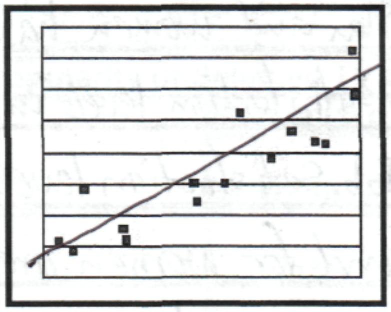
$\hat{q}_1 = 0.776$ $\quad \hat{q}_2 = 0.65$

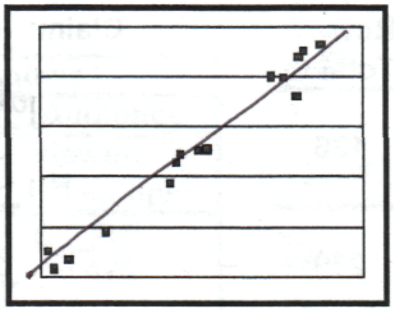
(3 points; 2 minutes) (Page 16; Problems 1, 2, and 3)

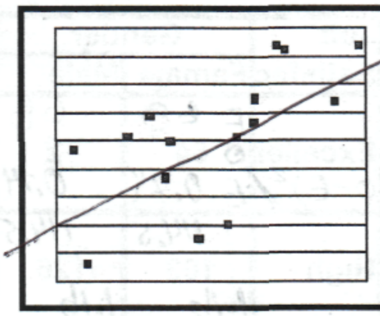
11. Assign the three sample correlation coefficients to the three pictures. A correlation value may be used more than once or not at all. If a picture has no appropriate value offered for its correlation, select the "N/A" option.

r = 0.90   r = 0.98   r = 0.60

Values
0.98
0.60
0.90
N/A





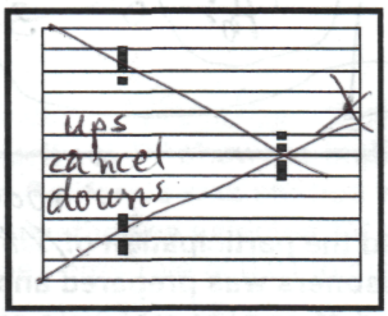


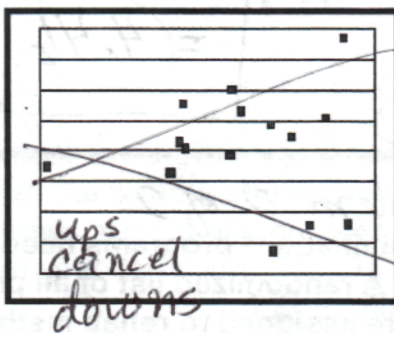
(3 points; 2 minutes)

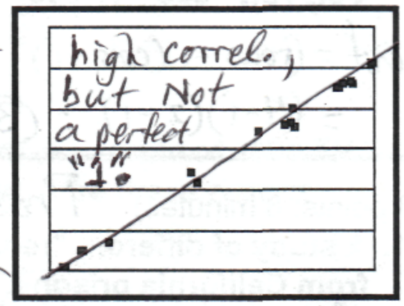
12. Assign the three sample correlation coefficients to the three pictures. A correlation value may be used more than once or not at all. If a picture has no appropriate value offered for its correlation, select the "N/A" option.

r = 0.00   r = 0.00   r = ~~NONE~~ N/A

Values
-0.80
1.00
0.00
N/A





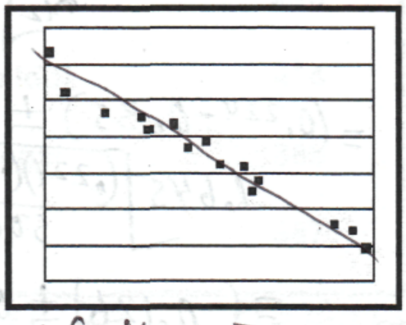


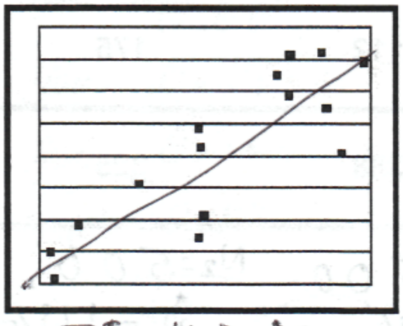
(3 points; 2 minutes)

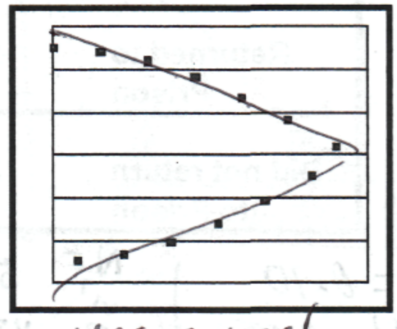
13. Assign the three sample correlation coefficients to the three pictures. A correlation value may be used more than once or not at all. If a picture has no appropriate value offered for its correlation, select the "N/A" option.

r = N/A (−0.70 with)   with   r = 0.85 (N/A)   r = 0.00

Values
0.00
-0.70
0.85
N/A







if this is −0.70, then the Next cannot be +0.85.

If this is +0.85, then the first cannot be −0.70.

ups cancel downs

Example Exam Problems
Unit #3; Page 17; Problem 1 of 1

$\boxed{\text{A.O.V.}}$

Anova Table

Source	d.f.	Sum of Squares	Mean Square	F	p-val
States	4	13714112	3428528	$\boxed{3.2565}$	0.0183
Error	54	56852688	1052828		
Total	58	70566800			

$N = 59$

$H_0: \mu_1 = \mu_2 = \cdots = \mu_5$

H_1: not H_0.

$\alpha = 0.02$

$0.0182 < 0.02$

p-value $< \alpha$

so reject H_0.

(10 points - 15 minutes) Page 18; Problem 1 of 1

5. **The following data are "random" measurements of responses to eight different "treatments". An incomplete Analysis of Variance table is given. Use the data to complete the ANOVA table (but do not include a p-value). Then use the results in your your table to carry out the appropriate test of the claim that the true means of the eight populations are all equal.**
(Use $\alpha = 0.025$ for this test)

8 treatments

	Treatment							
	A	B	C	D	E	F	G	H
	107	100	108	104	101	96	111	110
	104	98	95	101	100	97	104	117
	97	93	102	109	101	103	112	115
	100	95	99	98	106	99	110	109
	105		98	97	95	100	119	
	102		96	89		105		
	101		96	103		108		
	96					97		
	Sample Statistics for each Treatment							
Mean	101.5	96.5	99.1	100.1	100.6	100.6	111.2	112.8
Std. Dev.	3.82	3.11	4.56	6.34	3.91	4.31	5.36	3.86
N	8	4	7	7	5	8	5	4

$N = 48$

Overall Mean = 102.25

H1: $\mu_1 = \mu_2 = \cdots = \mu_8$

H0: Not Ho:

$\alpha = 0.025$

For Edition #3, let $\alpha = 0.05$.

Analysis of Variance

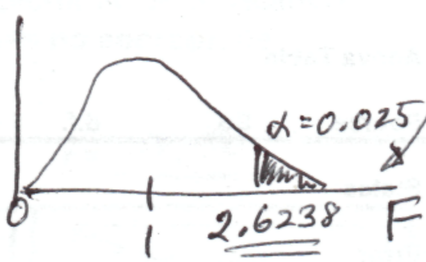
Source	df	SS	MS	F
Treatments	7	1111.67	158.81	7.498
Error	40	847.33	21.18	
Total	47	1959.0		

SS(Treatments) = MS(Treat) * d.f(treat)
= (158.81)(7) =

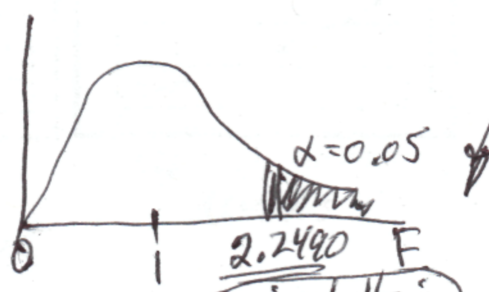
SS(Error) = SS(Total) − SS(Treat)
= (1959.0 − 1111.67) =

MS(Error) = SS(Error) ÷ DF(Error)
= 847.33 ÷ 40 = 21.18

F = MS(Treat) ÷ MS(Error)
= 7.498



Reject Ho:



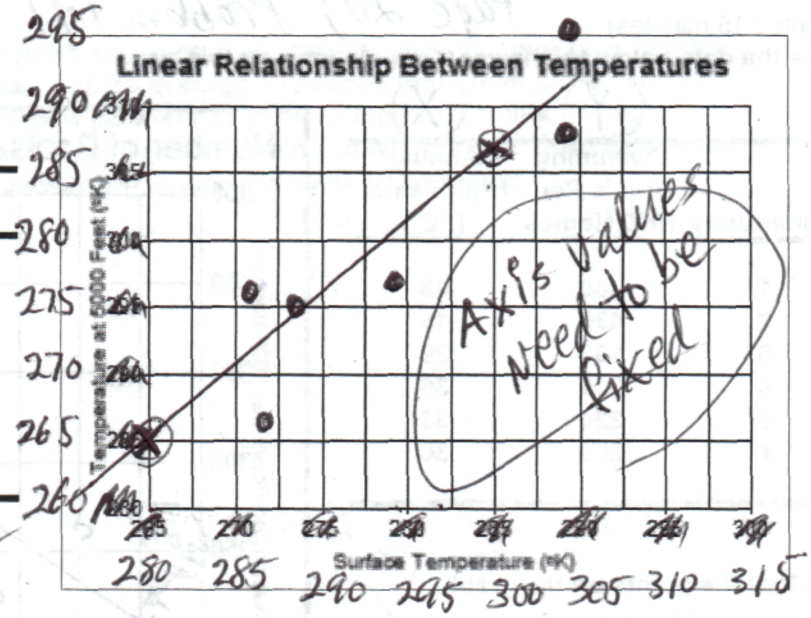
Reject Ho:

Page 19; Problem 1 of 1

(6 points; 7 minutes)

2. Based on the data given below, do parts (a) through (d). (Y) (X)

Observation	Temperature (°K) at 5000 feet	Surface
1	296	304
2	277	294
3	275	287
4	288	304
5	276	286
6	267	287
	(Y)	(X)



(a) Plot the data points on the graph. ✓

(b) Enter data in calculator and write the equation for the best-fitting line: ✓ $\hat{y} = -43.36 + 1.10(X)$

(c) Plot the line on the graph. ✓ $(x = 280 \quad y = 265)$ $(x = 300 \quad y = 287)$

(d) Predict the temperature at 5000 feet when the surface temperature is 280 °K?. ~~265~~ 264.8

$-43.36 + 1.10(280) = 265$

(e) What is the proportion of the variability in Y that is "explained" by the temperature at the surface? $0.8122 = r^2$

(b) The expression for the total variability in Y is: $\sum (y - \bar{y})^2$

(c) The value of the total variability in Y is: $538.8 = S_y^2 (n-1)$

(d) The expression for the explained variability in Y is: $\sum (\hat{y} - \bar{y})^2$

(e) The value of the explained variability in Y is: $437.65 = (r^2)(Total)$

(f) The expression for the unexplained variability in Y is: $\sum (y - \hat{y})^2$

(g) The value of the unexplained variability in Y is: 101.15 = Total − Explained

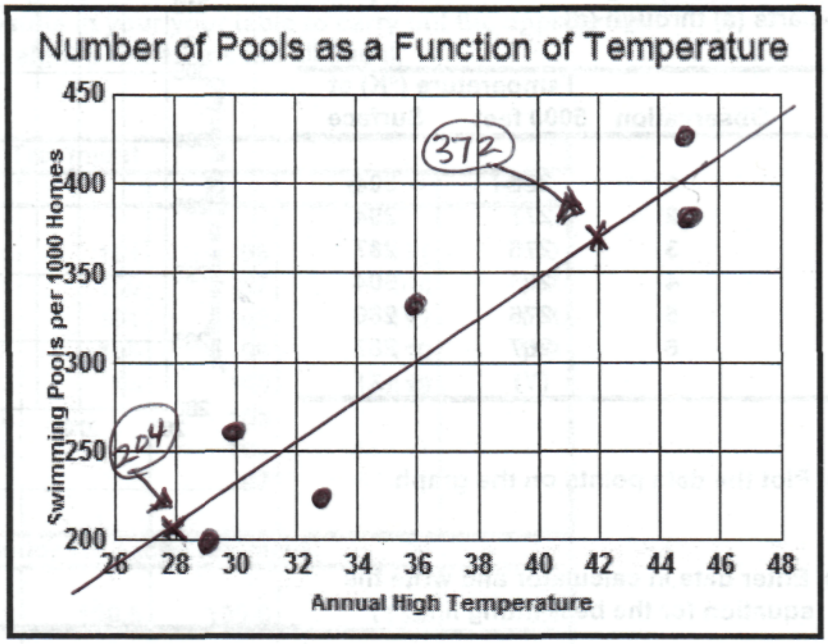
(h) The expression for the Standard Error of Estimate is: $S_e = \sqrt{\dfrac{\sum (y-\hat{y})^2}{n-2}} = \sqrt{\dfrac{\text{unexplained}}{n-2}}$

(i) The value of the Standard Error of Estimate is: 5.03

(15 points : 15 minutes)

2. Use the data below to answer the questions on this page.

Community	(Y) Swimming Pools Per 1000 Homes	(X) Annual High Temp. (°C)
1	380	45
2	430	45
3	199	29
4	331	36
5	224	33
6	260	30



(a) Plot the points on the graph. ✓

Plot the line ✓

(b) Determine the equation of the line that fits the data best and plot it:

intercept = -131 slope = 11.97 equation: $-131 + 11.97(x) = \hat{y}$

(c) For a new community, what is the estimated number of swimming pools per 1000 homes if the annual high temperature is 45 °C? 407.8 $= (-131) + 11.97(45)$

(d) What is the value of the linear correlation coefficient for the two variables? 0.9364 $= r$

(e) What percentage of the total variation in number of pools is explained by your line? 87.7% $= r^2$

$r^2 = 0.877$

(f) Write the symbolic expressions and give the values for the three items below:

	Total variation in number of pools	Explained variation in number of pools	Unexplained variation in number of pools
Symbolic expression	$\Sigma(y-\bar{y})^2$	$\Sigma(\hat{y}-\bar{y})^2$	$\Sigma(y-\hat{y})^2$
value	41742	36600	5142
	$S_y^2(n-1)$	r^2(total)	(total - explained)

(g) Write the symbolic expression and give the value for the standard error of estimate:

Symbolic expression $S_e = \sqrt{\dfrac{\Sigma(y-\hat{y})^2}{n-2}}$ 35.85 $= \sqrt{\dfrac{\text{unexplained}}{n-2}}$