Statistics 300
Exam #3

Name: Solution

Fall 2012
Saturdays 9:00 am - 1:05 pm

(9 points; 10 minutes)

1. **Use the counts by State in the table to test the idea that the percentage of all AZ fans that have the Sonics as their favorite team is greater than the percentage of all CA fans that have the Sonics as their favorite team. Use a 2% significance level for this test.**

The data represent random samples of Suns, Kings, and Sonics fans.

Favorite Basketball Team	Home State AZ	Home State CA	Home State WA	Row Total
Phoenix Suns	129	29	26	184
Sacramento Kings	40	129	16	185
Seattle Sonics	33	21	177	231
	202	179	219	600

$P_{SON|AZ} > P_{SON|CA}$

$(P_{AZ} - P_{CA}) > 0$

H$_0$: $P_{AZ} - P_{CA} \le \boxed{0}$

H$_1$: $P_{AZ} - P_{CA} > 0$

$\alpha = 0.02$ right tail

$\bar{p} = \dfrac{33+21}{202+179} = \dfrac{54}{381} = 0.1417$

$\bar{q} = 1 - 0.1417 = 0.8583$

$\hat{p}_{AZ} = \dfrac{33}{202} = 0.1634$

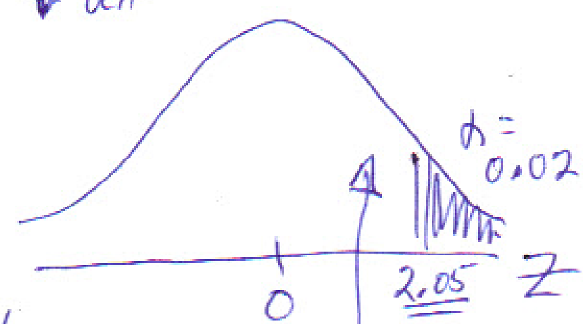
$\hat{q}_{AZ} = 0.8366$

$\hat{p}_{CA} = \dfrac{21}{179} = 0.1173$

$\hat{q}_{CA} = 1 - 0.1173 = 0.8827$

Test Statistic

$$\frac{(\hat{p}_{AZ} - \hat{p}_{CA}) - (p_{AZ} - p_{CA})_0}{\sqrt{\dfrac{\bar{p}\bar{q}}{N_{AZ}} + \dfrac{\bar{p}\bar{q}}{N_{CA}}}}$$



$$= \frac{(0.1634 - 0.1173) - 0}{\sqrt{\dfrac{(.1417)(.8583)}{202} + \dfrac{(.1417)(.8583)}{179}}} = \frac{0.0461}{0.0358} = \boxed{1.29}$$

Do Not Reject H$_0$!

Statistics 300 **Fall 2012**
Exam #3 **Name:** Solution **Saturdays 9:00 am - 1:05 pm**

(8 points; 8 minutes)

2. **Engineering students prepared a car~~e~~ so they could measure daily average speed and daily use of gasoline. They drove for 10 hours each day all around a major urban area in California. Use the data in the box to test the claim that spead and fule use are negatively correlated.** Use a Type I Error Rate of 0.025 for this test.

fuel

Day	Average Speed (mi/hour)	Fuel Use (gallons)
1	23.7	9.1
2	35.1	9.0
3	31.8	9.5
4	20.9	11.7
5	21.6	11.8
6	28.5	8.5

Claim: $\rho < 0$

H_0: $\rho \geq 0$

H_1: $\rho < 0$

$\alpha = 0.025$ left tail

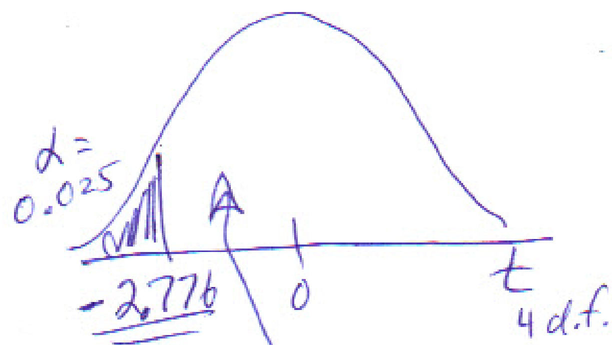
$\boxed{r = -0.7145}$

from calculator

Test Statistic

$$\frac{r}{\sqrt{\frac{1-r^2}{n-2}}} = \frac{-0.7145}{\sqrt{\frac{1-(-0.7145)^2}{6-2}}}$$

$$= \frac{-0.7145}{\sqrt{\frac{0.4895}{4}}} = \frac{-0.7145}{0.3498} = \boxed{-2.043}$$



Do Not reject Ho.

Statistics 300
Exam #3 **Name:** Solution

Fall 2012
Saturdays 9:00 am - 1:05 pm

(9 points; 10 minutes)

3. Question: Do cows give more milk in July than they do in January? Use the data below for eight cows to test the claim that cows produce at least 0.5 gallons per day more on average in July than they do in January. Experience indicates that variation in milk production per cow is the same in July and January.
Use a 5% significance level for your test.

matched Pairs

8 cows

$\mu_{JUL} \geq \mu_{JAN} + 0.5$

Claim: $(\mu_{JUL} - \mu_{JAN}) \geq 0.5$ # μ_d

H$_0$: $(\mu_{JUL} - \mu_{JAN}) \geq 0.5$ # μ_d

H$_1$: $(\mu_{JUL} - \mu_{JAN}) < 0.5$ # μ_d

$\alpha = 0.05$ left tail

Gallons of Milk per Day

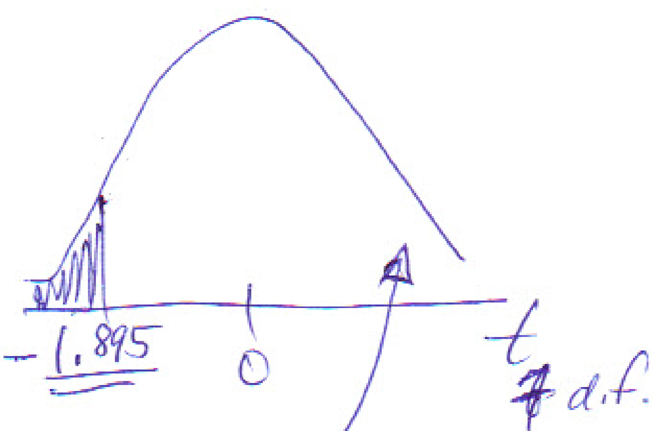
Cow	January	July	July - Jan.
1	5.3	5.1	-0.2
2	5.4	6.8	1.4
3	6.0	6.9	0.9
4	5.9	6.9	1.0
5	6.8	7.3	0.5
6	4.8	5.8	1.0
7	5.5	6.3	0.8
8	6.3	6.9	0.6
mean =	5.75	6.50	$0.75 = \bar{d}$
stdev =	0.63	0.73	$0.472 = s_d$
n =	8	8	$8 = n$
	Smaller	Bigger	$7 = d.f.$

Test Statistic

$$\frac{\bar{d} - \mu_d}{s_d/\sqrt{n}} = \frac{0.75 - 0.5}{0.472/\sqrt{8}}$$

Because matched pairs, S^2_{pooled} or not pooled does not matter

$$= \frac{0.25}{0.167} = \boxed{1.497}$$



Do Not reject Ho!

Statistics 300
Exam #3

Name: Solution

Fall 2012
Saturdays 9:00 am - 1:05 pm

(9 points; 10 minutes)

4. **Use the survey results for 600 families to test the claim that Age when autism is diagnosed is independent of whether the family had health insurance. Let alpha = 0.05 for this test. of 0.05 for this test.** Counts in categories. OBS counts in 2 categories (2 Factors) at the same time.

Age in Years when autism diagnosed	Family had Health Insurance		Total
	Yes	No	
< 1	71	29	100
1	69	31	100
2	69	31	100
3	77	23	100
4	70	30	100
> 4	77	23	100
Total	433	167	600

claim: age when autism diagnosed is independent of Health Ins.

H_0: age & insured are indep.

H_1: Not H_0!

$\alpha = 0.05$ right tail

$df = (\text{rows} - 1)(\text{cols} - 1) = (6-1)(2-1) = (5)(1) = 5$

$$\text{Expected} = \frac{(\text{row total})(\text{col total})}{\text{grand total}}$$

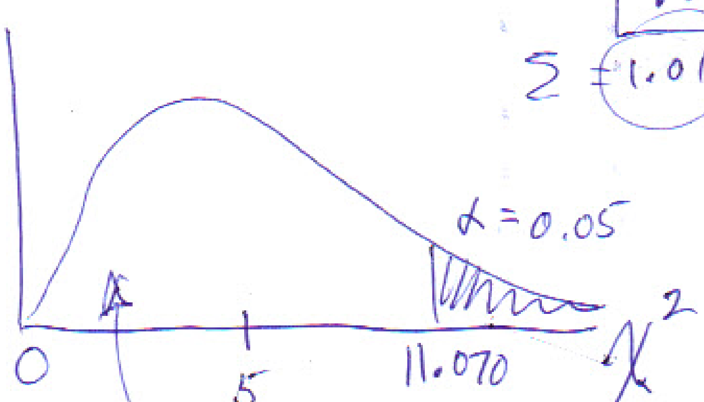
72.2	27.8
72.2	27.8
72.2	27.8
72.2	27.8
72.2	27.8
72.2	27.8

$$\frac{(\text{OBS}-\text{EXP})^2}{\text{EXP}}$$

.02	.05
.14	.37
.14	.37
.32	.83
.07	.17
.32	.83

$\Sigma = 1.01$ $\Sigma = 2.62$

$$\sum\left[\frac{(\text{OBS}-\text{EXP})^2}{\text{EXP}}\right] = 3.63$$



Do Not reject Ho!

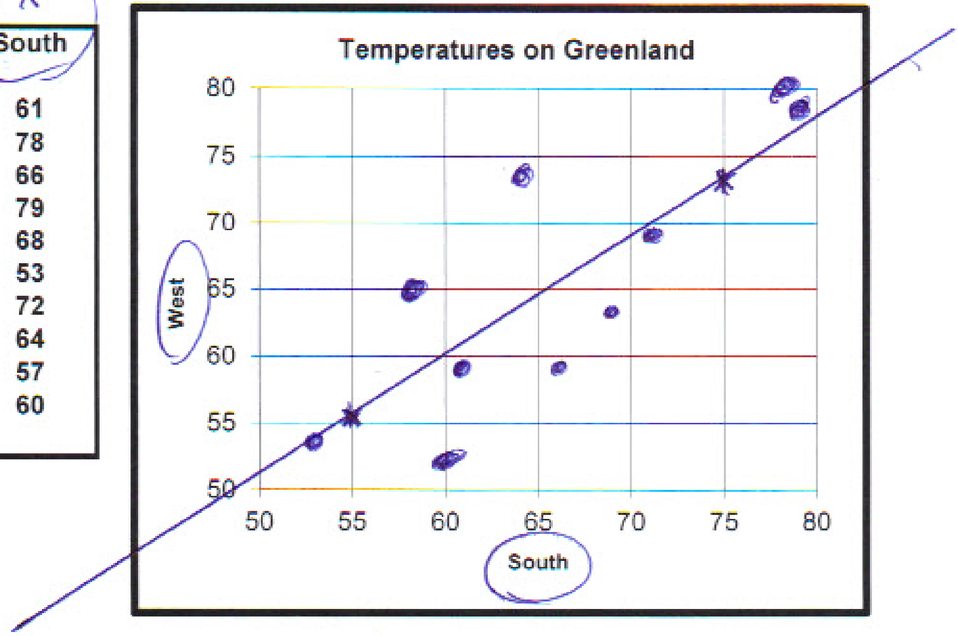
Statistics 300
Exam #3 Name: Solution

Fall 2012
Saturdays 9:00 am - 1:05 pm

(13 points; 14 minutes)

5. Plot daily temperatures for West (y) and South (x) parts of Greenland. Each row in the data set is for a different day. Then answer parts b, c, d, e, f, and g.

	East	West (Y)	North	South (X)
1	53	59	58	61
2	78	80	79	78
3	53	59	51	66
4	74	79	78	79
5	79	63	66	68
6	53	54	50	53
7	80	69	76	72
8	53	74	58	64
9	58	65	61	57
10	56	52	55	60



(a) Plot the points on the graph.

(b) Use your calculator to determine the equation of the line that best predicts the East temperature based on the North temperature.

equation of your line : $\hat{y} = 4.71 + 0.922(X)$

(c) Plot your line on the graph.

if $X = 55$, $\hat{y} = 55.4$; $X = 75$, $\hat{y} = 73.9$

(d) What is the linear correlation for the given ~~North~~ WEST and ~~East~~ SOUTH data? $r = 0.8023$

(e) Provide the symbolic expressions for Total, Explained, and Unexplained variation in "Y".

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

Total Variation = Explained Variation + Unexplained Variation

(f) Provide the values for Total, Explained, and Unexplained variation in "Y" for the graphed data.

882.4 = 568.0 + 314.4

Total Variation = Explained Variation + Unexplained Variation

$s_y^2(n-1)$ r^2(Total)   Total − Explained

(g) Provide symbolic expression and the value of the "Standard Error of Estimate."

Symbolic Expression: $\sqrt{\frac{\Sigma(y-\hat{y})^2}{n-2}}$ ← unexplained

Value: 6.27

$\sqrt{\frac{314.4}{8}} = 6.27$

Statistics 300
Exam #3 Name: Solution

Fall 2012
Saturdays 9:00 am - 1:05 pm

(9 points; 10 minutes)

6. For questions "a" through "c", check all the circles that are true.

(a) A hypothesis test had the following parts:

H_1: $(p_1 - p_2) > 0.012$
Significance level = 0.025
Conclusion: Reject H_0:

right tail Z

- ✓ The p-value was less than 0.025
- O The critical value was from the t table
- ✓ The critical value was for alpha = 0.025 in the right tail
- ✓ The critical value was greater than 1.28
- ✓ The test statistic value was greater than 1.96

(b) A hypothesis test had the following parts:

H_0: $(p_1 - p_2) = 0$
Significance level = 0.10
Conclusion: Do not reject H_0:

H_1: ≠
2 tails
Z

- O The p-value was less than 0.10
- ✓ The critical value was from the Z table
- O The critical value was for 0.025 in the right tail
- ✓ The critical values were -1.645 and 1.645
- O The test statistic value was greater than 1.96

(c) A hypothesis test had the following parts:

H_1: $(\mu_1 - \mu_2) < 12$
Significance level = 0.01
Conclusion: Reject H_0:

left tail
t dist.

- ✓ The p-value was less than 0.01
- ✓ The critical value was from the t table
- O The critical value was for 0.01 in the right tail
- ✓ The critical value was negative
- ✓ The test statistic value was left of the critical value

Statistics 300
Exam #3

Name: Solution

Fall 2012
Saturdays 9:00 am - 1:05 pm

(10 points; 10 minutes)

7. **Two formulas for glue, Formula A and Formula B, are used to join pieces of wood together. Standard wood joints are made with each glue and tested for strength. Use the statistics given here to make a 98% confidence interval for the mean strength of glue A (μ_A) minus the mean strength of glue B (μ_B). Variabilty in the strengths of the joints is about the same for both glues.**

pool variances

Strength of Glue Joints		
Sample Statistic	Glue A	Glue B
N =	8	18
df =	7 +	17 = 24
Average =	863	874
Std. Deviation =	6.5	7.4

$t = 2.492$

$$S^2_{pool} = \frac{S_A^2(n_A-1) + S_B^2(n_B-1)}{(n_A-1)+(N_B-1)} = \frac{(6.5)^2(7) + (7.4)^2(17)}{7+17} = 51.11$$

$$98\%\ CI(\mu_A - \mu_B) = (\bar{x}_A - \bar{x}_B) \pm t\sqrt{\frac{S^2_{pool}}{N_A} + \frac{S^2_{pool}}{N_B}}$$

$$= (863 - 874) \pm 2.492\sqrt{\frac{51.11}{8} + \frac{51.11}{18}}$$

$$= (-11) \pm 7.57 = \left[-18.57 < (\mu_A - \mu_B) < -3.43\right]$$

Based on your interval is it reasonable to claim that joints made with glue A are stronger on average than joints made with glue B?

Yes **No** **Why?** if $\mu_A > \mu_B$ then $(\mu_A - \mu_B) > 0$, but there are No values in the CI that are > 0.

Based on your interval is it reasonable to claim that joints made with glue B are stronger on average than joints made with glue A?

Yes **No** **Why?** if $\mu_A < \mu_B$ then $(\mu_A - \mu_B) < 0$, and the whole CI is < 0 (Negative)

Statistics 300
Exam #3

Name: Solution

Fall 2012
Saturdays 9:00 am - 1:05 pm

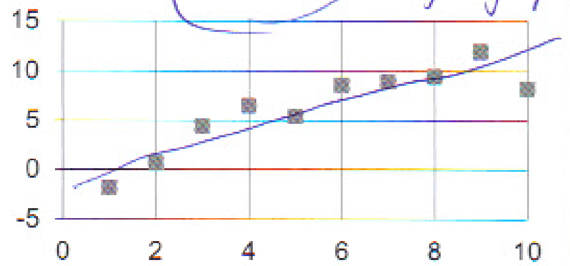
(6 points; 6 minutes)

8. **Connect each picture with one of the candidate "r" values by writing the appropriate candidate "r" value in the space at the top of each graph.**

Candidate values of "r", the sample correlation coefficient.						
0.00	-0.70	-0.90	-1.00	0.70	0.90	1.00

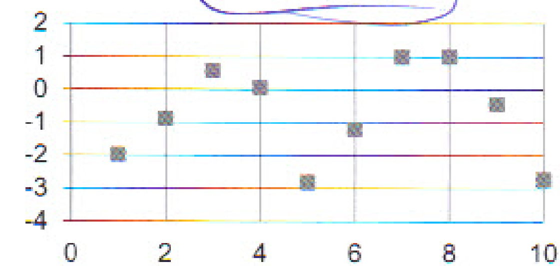
(+) about the same |r| as for bottom right graph

0.90



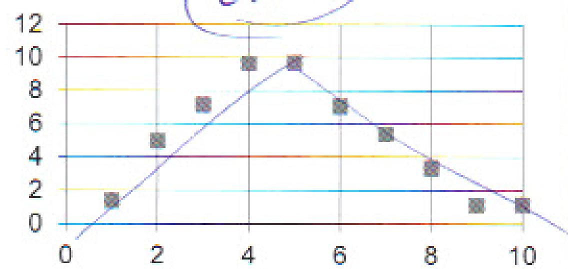
Neither up nor down

0.00



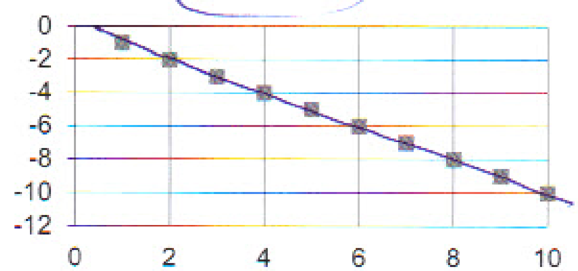
ups cancel downs

0.00



perfect (−)

−1.00

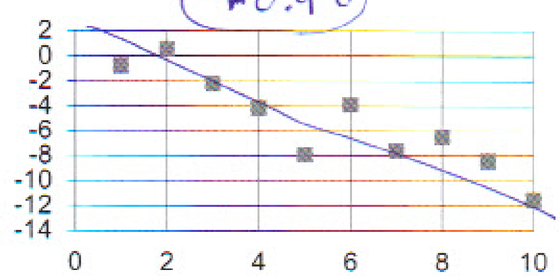


(−)

−0.70

(−)

−0.90



Statistics 300
Exam #3 **Name:** solution **Fall 2012**
Saturdays 9:00 am - 1:05 pm

(9 points; 9 minutes)

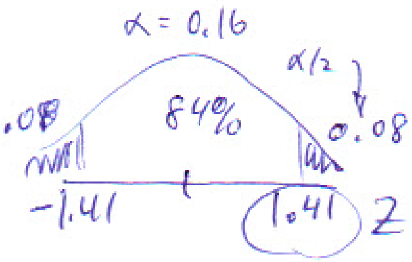
9. Based on the data shown below from a random sample of 800 people, construct an 84% confidence interval for the difference between the proportion of meat-eaters die from heart disease and the proportion of vegans that die of heart disease.

	Cause of Death is Heart Disease				
	Yes	**No**	N	$\hat{p}$	$\hat{q}$
Meat-eaters	**53**	**347**	400	$\frac{53}{400} = .1325$	.8675
Vegans	**48**	**352**	400	0.12	0.88
			800		

$$\boxed{Z_{\alpha/2} = 1.41}$$

$$84\%\ CI(p_m - p_v) = (\hat{p}_m - \hat{p}_v) \pm Z_{\alpha/2}\sqrt{\frac{\hat{p}_m\hat{q}_m}{N_m} + \frac{\hat{p}_v\hat{q}_v}{N_v}}$$

$$= (0.1325 - 0.12) \pm 1.41\sqrt{\frac{(.1325)(.8675)}{400} + \frac{(.12)(.88)}{400}}$$



$$= 0.0125 \pm (1.41)(0.0235)$$

$$= 0.0125 \pm 0.0331$$

$$= [-0.0206 < (p_{meat} - p_{veg}) < 0.0456]$$

Based on your interval is it reasonable to claim that the percentage of Vegans that die of heart disease is the same as the percentage of meat-eaters that die of heart disease?

Yes **No** **Why?** If $p_{meat} = p_{vegan}$ then $(p_{meat} - p_{veg}) = 0$ and $\emptyset$ is in the CI, the reasonable range for the truth.

Statistics 300
Exam #3 **Name:** Solution **Fall 2012**
Saturdays 9:00 am - 1:05 pm

(9 points; 7 minutes) N

10. Use the 320 values on the next page to complete the Analysis of Variance table and test the claim that milk from the ten different producers has the same average amount of butter fat per 10 liters of milk. Use an 8% significance level for the test.

$\alpha = 0.08$

AOV Table

Source	SS	df	MS	F	p-value
Producer	1548.999	9	172.111	1.2198	0.282018
Error	43743	310	141.01		
Total	45292	319			

$> \alpha$, so Do Not Reject H_0

H_0: $\mu_1 = \mu_2 = \cdots = \mu_{10}$

H_1: Not H_0

Based on the completed table, the value of the "pooled variance" = 141.01 **?**

= MS(error)

10 producers

Grams of Butter Fat per 10 liters of Milk

	A	B	C	D	E	F	G	H	I	J
	95	118	99	99	87	108	107	92	82	92
	87	95	103	110	81	112	87	98	119	92
	90	105	97	120	105	97	91	113	91	100
	113	83	101	112	109	117	110	112	97	116
	118	84	100	81	112	101	104	111	103	118
	102	110	105	88	85	108	100	107	104	116
	95	81	86	96	119	97	98	91	116	115
	91	100	116	88	120	89	92	100	109	113
	92	93	83	112	107	99	101	98	89	105
	89	103	114	87	95	109	110	100	100	94
	84	109	98	94	85	100	112	101	81	110
	120	86	102	84	116	99	95	82	80	109
	95	89	86	114	106	95	109	83	82	116
	114	109	81	105	102	88	101	85	90	118
	120	80	87	93	118	116	93	119	96	101
	113	106	100	86	89	116	116	106	82	117
	110	83	83	112	100	87	86	113	115	112
	115	109	98	83	107	97	85	86	115	105
	90	83	116	96	86	106	97	99	83	99
	109	104	84	86	101	95	103	108	93	111
	81	102	88	91	91	108	111	111	118	85
	112	106	92	120	89	112	83	92	85	101
	102	114	111	119	116	100	95	83	108	111
	96	85	108	109	112	111	87	81	80	83
	119	113	109	90	84	102	106	118	116	104
	100	110	103	104	83	89	82	93	107	92
	85	90	105	113	80	100	86	94	82	
	106	103	95	99	94	99	105	100	114	
	117	84	120	83		99	118	101	85	
	116	115	89	99		104	80	109	119	
	109	99	103	116			118	81	96	
	82	116	90	98			112	103	83	
	114	118	120	110			82	101		
	111	86					83	81		
		92					87			
n =	34	35	33	33	28	30	35	34	32	26
Mean =	102.7	98.9	99.2	99.9	99.3	102.0	98.1	98.6	97.5	105.2
StDev =	12.6	12.3	11.3	12.4	13.1	8.4	11.6	11.4	14.0	10.5

Statistics 300
Exam #3

Name: Solution

Fall 2012
Saturdays 9:00 am - 1:05 pm

(9 points; 10 minutes)

11. Two programs for encouraging school attendance were studied at some schools. Use the results to test the claim that the average number of attendance days (per 100 students) at all schools would be at least 500 days greater if all schools used Method B instead of Method A. Variability in the number of attendance days is clearly greater with Method B than it is with Method A. Use a 5% significance level for this test.

$\mu_B \geq \mu_A + 500$

$\sigma_A^2 \neq \sigma_B^2$ Do not pool variances

Attendance Results During Study Days per 100 students per school		
Sample Statistic	Method A	Method B
n =	17	12
mean =	17020.1	17575.6
st. dev. =	425.5	702.6

df = 16 11

use smaller

Claim: $(\mu_B - \mu_A) \geq 500$

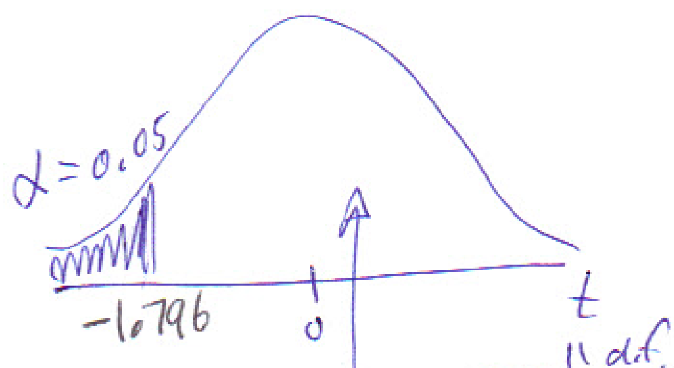
H_1: $(\mu_B - \mu_A) \geq 500$

H_1: $(\mu_B - \mu_A) < 500$

$\alpha = 0.05$ left tail

Test Statistic

$$\frac{(\bar{x}_B - \bar{x}_A) - (\mu_B - \mu_A)_0}{\sqrt{\dfrac{S_A^2}{N_A} + \dfrac{S_B^2}{N_B}}}$$



$$= \frac{(17575.6 - 17020.1) - 500}{\sqrt{\dfrac{(425.5)^2}{17} + \dfrac{(702.6)^2}{12}}} = \frac{55.5}{227.6} = 0.244$$

Do not reject Ho!

Statistics 300
Exam #3

Name: Solution

Fall 2012
Saturdays 9:00 am - 1:05 pm

(8 points; 8 minutes)

12. Facing serious budget problems, the city manager and the Police Chief want to use the police officers in the most helpful way possible. They believe more officers are needed on duty on Friday and Saturday nights than on other nights of the week because people get drunk more often on those nights. But others think their assumption is wrong. Use the data on arrests that involve alcohol by day of week to test the claim that such arrests occur on all days of the week with equal frequency. Let $\alpha = 0.05$ for this test.

Data for Year = 2008			
Day of the Week	Number of Arrests Involving Alcohol (OBS)	EXP	$\frac{(Obs-Exp)^2}{Exp}$
Sun	360	482.3	31.01
Mon	418	482.3	8.57
Tue	513		1.95
Wed	465		0.62
Thu	378		22.56
Fri	601		29.21
Sat	641	482.3	52.22
Total	3376		$\Sigma = 146.14$

1/7 of obs. are expected for each day of week if Ho: is true.

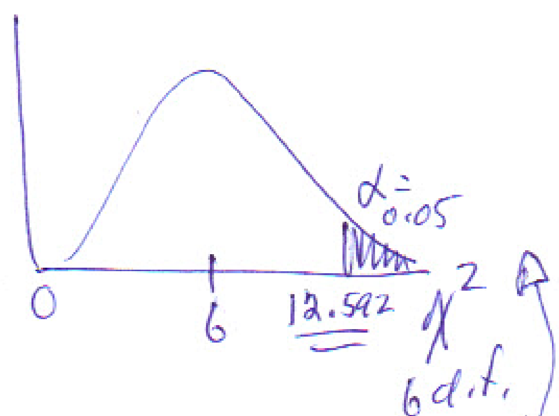
$(1/7)\,3376 = 482.3$

H_0: Alcohol related arrests occur w/ equal freq on all days

H_1: Not Ho:

$\alpha = 0.05$ right tail

$df = 7-1 = 6$



$$\sum\left[\frac{(OBS-EXP)^2}{EXP}\right] = 146.14$$

REJECT Ho: