Statistics 300
Instructor: L.C.Larsen

Name: Solution

Summer 2011
M/Tu/W/Th 5:3-8:15 pm

(10 points; 10 minutes)

1. **The data below are for 589 randomly selected people with autism. Use the data to make a 92% confidence interval for the difference between the proportion that had health insurance in the "<1" age group and the proportion that had health insurance in the ">4" age group. Then answer the question at the bottom of the page.**

$$92\%\ CI(p_{>4} - p_{<1}) = (\hat{p}_{>4} - \hat{p}_{<1}) \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_{>4}\hat{q}_{>4}}{n_{>4}} + \frac{\hat{p}_{<1}\hat{q}_{<1}}{n_{<1}}}$$

Age in Years when autism diagnosed	Family had Health Insurance		N
	Yes	No	
< 1	22	31	53
1	34	33	
2	47	46	
3	60	57	
4	67	64	
> 4	68	60	128

$$\boxed{\hat{p}_{<1} = \frac{22}{53} = 0.415 \quad \hat{q}_{<1} = 0.585}$$

$$\boxed{\hat{p}_{>4} = \frac{68}{128} = 0.531 \quad \hat{q}_{>4} = 0.469}$$

confidence $= 0.92$ $\quad \alpha = 0.08$ $\quad \alpha/2 = 0.04$

$z_{\alpha/2} = 1.75$

	.05
-1.7	$\boxed{.0401}$

$$92\%\ CI(p_{>4} - p_{<1}) = (0.531 - 0.415) \pm 1.75\sqrt{\frac{(0.531)(.469)}{128} + \frac{(.415)(.585)}{53}}$$

$$= \cancel{0.062}\ 11.6 \pm (1.75)(0.0808)$$

$$= \cancel{0.062}\ 11.6 \pm .1414 = [\cancel{-0.079}\ -0.0254 < (p_{>4} - p_{<1}) < \cancel{0.2034}\ 0.2574]$$

Based on your confidence interval, is it reasonable to claim that the proportion that had health insurance in the "<1" population is greater than the proportion that had health insurance in the ">4" population?

Yes   **No**   **Why?** There are negative values in the CI for $(p_{>4} - p_{<1})$

if $(p_{<1} > p_{>4})$ then $(p_{>4} - p_{<1})$ will be negative.

Statistics 300
Instructor: L.C.Larsen

Name: Solution

Summer 2011
M/Tu/W/Th 5:3-8:15 pm

DO PROBLEM 2 OR PROBLEM 4 ON THE NEXT PAGE.

(9 points; 9 minutes)

2. **Use the data below to test the claim that "the higher the wind speed the cooler the temperature." That is, test whether the data imply a negative population correlation between temperature and wind speed.** (Use a 0.025 significance level for this test.)

Day	Wind Speed	Temperature
1	6	87
2	6	96
3	6	87
4	8	92
5	3	98
6	19	79
* In 1000's of dollars		

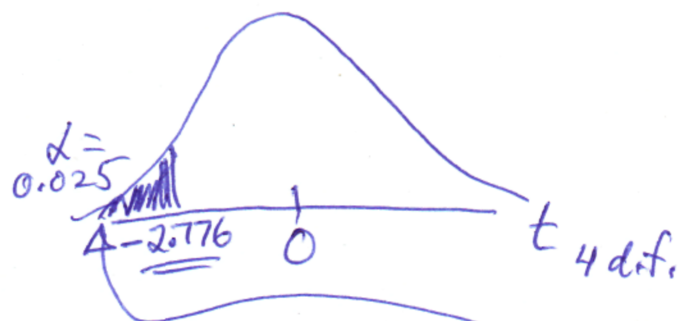
Claim: $\rho < 0$

H_0: $\rho \geq 0$

H_1: $\rho < 0$

$\alpha = 0.025$ left tail

critical region



$r = -0.822$

$n = 6$

$df = 6-2 = 4$

Test Statistic

$$\frac{r}{\sqrt{\frac{1-r^2}{n-2}}} = \frac{-0.822}{\sqrt{\frac{1-(-0.822)^2}{6-2}}} = \frac{-0.822}{0.2847} = -2.887$$

Reject H_0:

(4 points; 2 minutes)

3. **(a) What value is at the center (mean) of the standard normal (Z) distribution?** 0

   (b) What value is at the center (mean) of the "t" distribution with 12 d.f.? 0

   (c) What value is at the center (mean) of the"chi-square" distribution with 12 d.f.? 12

   (d) What value is at the center (mean) of the "F" distribution with 12 numerator d.f. and 12 denominator d.f.? 1

Statistics 300
Instructor: L.C.Larsen

Name: Solution

Summer 2011
M/Tu/W/Th 5:3-8:15 pm

DO PROBLEM 4 OR PROBLEM 2 ON THE PREVIOUS PAGE.

(9 points; 8 minutes)

4. **The General Mills company is experimenting with different amounts of flour, sugar, salt, and shortening for a new cake mix. Their expert cooks have eight different experimental recipes. The cooks use these recipes to make a total of 61 cakes. After baking, the "firmness" of each cake is measured. Use the information to complete the Table and test the claim that the populations of all cakes made with each of the 8 recipes have the same average firmness.** (Use $\alpha = 0.05$.)

	Firmness of Cakes Made with Recipe #							
	1	2	3	4	5	6	7	8
Mean =								
St.Dev. =								
n =	12	10	6	8	6	7	5	7

Data Hidden

Complete the Analysis of Variance Table and test the claim.

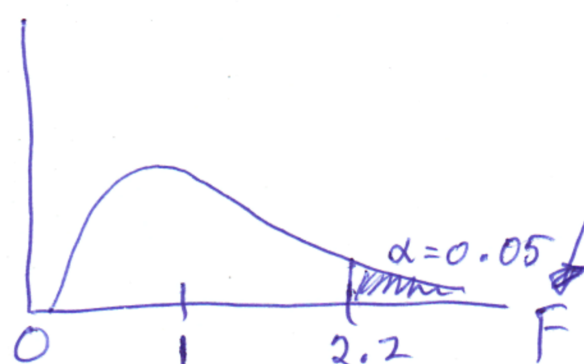
Source	Sum of Squares	df	Mean Square	F
Recipes	89.7486	7	12.82	5.2029
Error	130.61	53	2.464	
Total	220.3586	60		

Claim: all 8 population means are equal

H_0: $\mu_1 = \mu_2 = \cdots = \mu_8$

H_1: Not Ho!

$\alpha = 0.05$



7 d.f. numer.
53 d.f. denom.

Reject Ho!

Statistics 300
Instructor: L.C.Larsen

Name: Solution

Summer 2011
M/Tu/W/Th 5:3-8:15 pm

(9 points; 10 minutes)

5. A study compared the average performance of "A" students at Community Colleges (CC) on a standardized test to the average performance of "A" students at University of California (UC) cumpuses on the same test. Test the claim of UC supporters that the population of UC "A" students average at least 5 points higher compared to the population of CC "A" students.
(Use α = 0.05, and assume that variability in test scores of students at CC and UC is the same.)

Scores of "A" Students	CC	UC
$\bar{x}$ =	168	170
s =	14	12
n =	17	13

Claim: $\mu_{uc} \geq \mu_{cc} + 5$

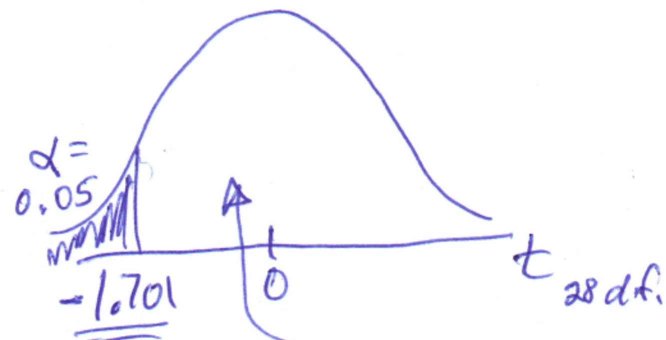
H_0: $(\mu_{uc} - \mu_{cc}) \geq 5$

H_1: $(\mu_{uc} - \mu_{cc}) < 5$

$\alpha = 0.05$ left tail

df = 16 + 12 = 28

$$S^2_{pool} = \frac{(n_{cc}-1)S^2_{cc} + (n_{uc}-1)S^2_{uc}}{(n_{cc}-1) + (n_{uc}-1)} = \frac{(16)(14)^2 + (12)(12)^2}{16 + 12} = 173.7$$



Test Statistic

$$\frac{(\bar{x}_{uc} - \bar{x}_{cc}) - (\mu_{uc} - \mu_{cc})_0}{\sqrt{S^2_{pool}/n_{uc} + S^2_{pool}/n_{cc}}} = \frac{(170-168) - 5}{\sqrt{\frac{173.7}{13} + \frac{173.7}{17}}} = \frac{-3}{4.856} = \boxed{-0.618}$$

Do not reject H_0:

(3 points; 3 minutes)

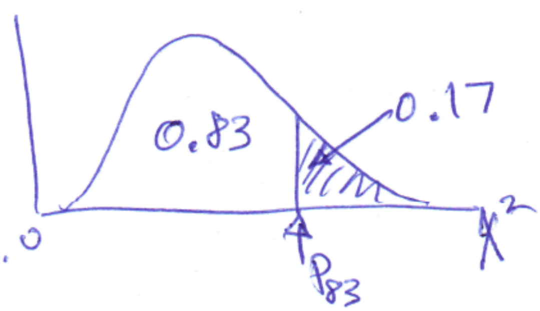
6. Answer questions (a) through (c):

(a) A contingency table (5 rows, 3 columns) is tested for independence.
If Alpha = 0.05 and P-value = 0.17, what do you conclude? Reject or Do not reject? p-val > α

(b) A correlation coefficient (23 {x,y} pairs) is tested: H0: r > 0.
If Alpha = 0.025 and P-value = 0.017, what do you conclude? Reject or Do not reject? p-val < α

(c) What percentile did the test statistic in part (a) represent? Test statistic = P(83)
It may help to draw a critical region picture that includes the p-value.

right tail test, so p-value is in right tail.

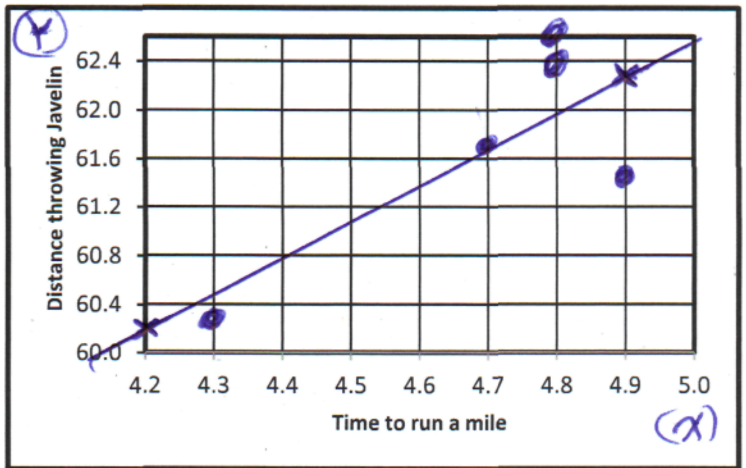


Statistics 300 | Name: Solution | Summer 2011
Instructor: L.C.Larsen | | M/Tu/W/Th 5:3-8:15 pm

(14 points, 18 minutes)

1. **Use the "distance" data to complete questions 'a' through 'm'.**

Athletic Performance		
Athlete	Time for Mile	Distance Javelin
1	4.3	60.3
2	4.8	62.4
3	4.7	61.7
4	4.9	61.4
5	4.8	62.5
	X	Y



(a) Plot the data on the graph.

(b) Write the equation of the regression "best fit" line. $\hat{y} = 47.8 + 2.95(x)$

Plot the "best fit" line on the graph..

(c) Predict "Javelin Distance" if "Time for Mile" is 4.5 . $y'(4.5) = 47.8 + 2.95(4.5)$ 61.07

(d) What is the sample correlation between mile time and Javelin distance? $r = 0.7781$

(e) What fraction of the variation "distance" does "mile time" explain? $r^2 = 0.6054$

percentage, or proportion

(f) What is the expression for the "total variation in Y"? $\sum (y - \bar{y})^2$

(g) What is the value of the "total variation in Y"? 3.172

(h) What is the expression for the "explained variation in Y"? $\sum (\hat{y} - \bar{y})^2$

(i) What is the value of the "explained variation in Y"? $= r^2(\text{total}) = 1.9203$

(j) What is the expression for the "unexplained variation in Y"? $\sum (y - \hat{y})^2$

total − explained =

(k) What is the value of the "unexplained variation in Y"? 1.2517

(l) What is the expression for the "Standard Error of Estimate"? $S_e = \sqrt{\frac{\sum (y - \hat{y})^2}{n-2}}$

(m) What is the value of the "Standard Error of Estimate"? 0.646

$$\sqrt{\frac{\sum (y - \hat{y})^2}{n-2}}$$

(9 points; 10 minutes)

7. **~~A thousand randomly selected adults 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 choose different sizes of cars in the same proportions.** (Use α = 10% for this test.)

400 adults were asked what size car they have.

Size of Automobile	Gender		Row Total
	Female	Male	
Large	62 / 50	38 / 50	100
Medium	57 / 56.5	56 / 56.5	113
Small	81 / 93.5	106 / 93.5	187
Col. Total	200	200	400

Claim: men and woman choose sizes of cars in different proportions

H_0: men and women choose sizes of cars in the same props.

H_1: men ≠ women choose sizes of cars in different proportions

$\alpha = 0.10$ right tail

$df = (r-1)(c-1)$
$= (3-1)(2-1) = (2)(1) = 2$

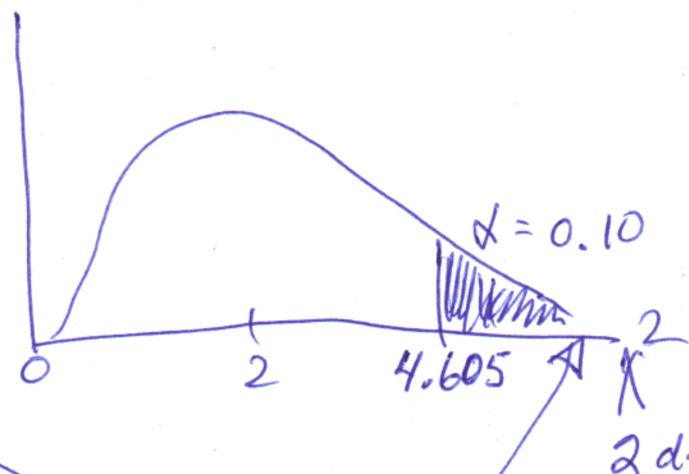
Expected counts = $\dfrac{(\text{row total})(\text{column total})}{\text{grand total}}$

$(O-E)^2/E$

2.88	2.88

total = 5.76

$\sum\left[\frac{(O-E)^2}{E}\right] \geq 5.76$



reject Ho!

Statistics 300
Instructor: L.C.Larsen

Name: Solution

Summer 2011
M/Tu/W/Th 5:3-8:15 pm

(8 points; 8 minutes)

8. **The ethnicity and gender of a random sample of 1000 CRC students is to be compared to the percentages of these categories in the Area from which CRC draws students. Test the claim the CRC student body does not match the proportions in the surrounding area.**
(Use a Type I error rate of 0.05 to make your decision.)

Data on Ethnicity and Gender Monday 6 p.m. Audience

Ethnicity / Gender	CRC Sample	CRC Area
Pale Males	300	28%
Pale Females	330	29%
Non-Pale Males	80	20%
Non-Pale Females	290	23%
Total =	1000	100%

(% * 1000)

Expected	$(O-E)^2/E$
280	1.43
290	5.52
200	72.00 enough
230	15.65 enough

$\Sigma = 94.6$

Claim: CRC proportion ≠ CRC area props.

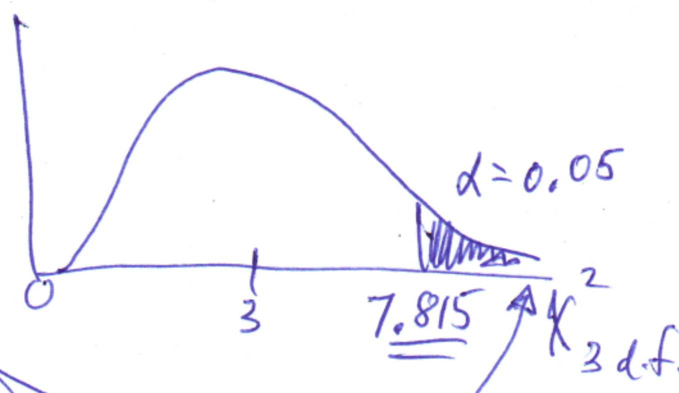
H_0: CRC proportions = CRC area proportions

H_1: " ≠ "

$\alpha = 0.05$ right tail

$df = k - 1 = 4 - 1 = 3$

$$\Sigma\left[\frac{(O-E)^2}{E}\right] = 94.6$$



Reject Ho!

Statistics 300
Instructor: L.C.Larsen

Name: Solution

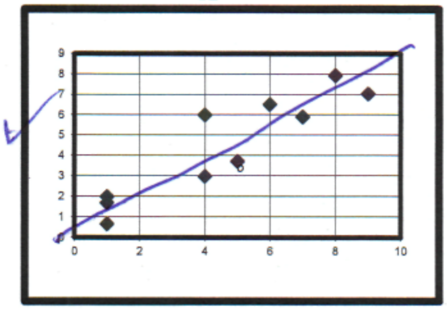
Summer 2011
M/Tu/W/Th 5:3-8:15 pm

(10 points; 6 minutes)

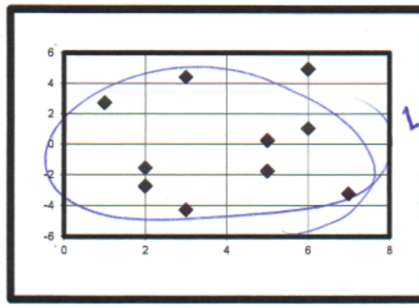
10. Assign the letters of the appropriate figures to each of the "sample correlation" values offered below. If an "r" value has no appropriate figure, write "none" next to it.

value of "r"	Figure Letter(s)
1.50	None
1.00	D, E
0.90	A
0.70	F
0.00	B, G
- 0.50	J
- 0.70	H
- 0.90	C
- 1.00	I

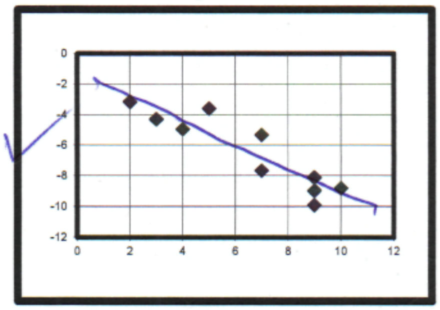
A



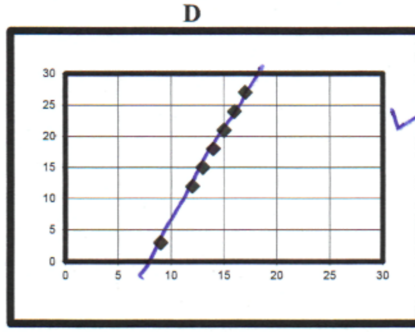
B



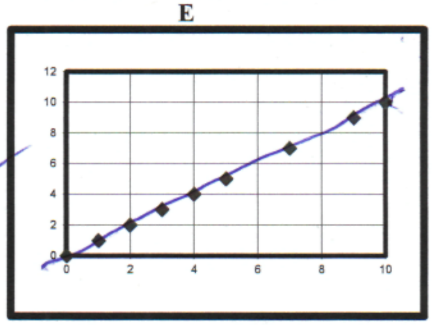
C



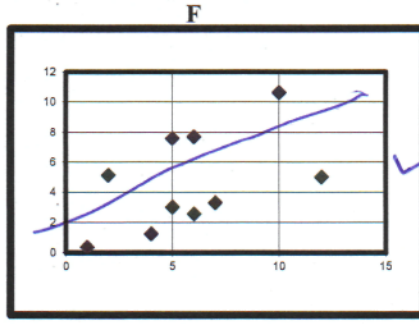
D



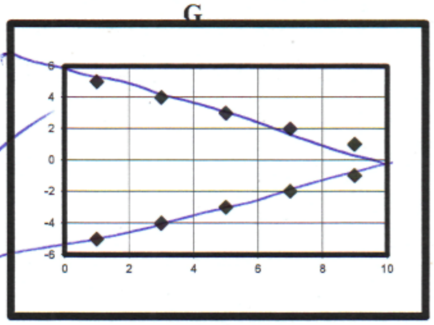
E



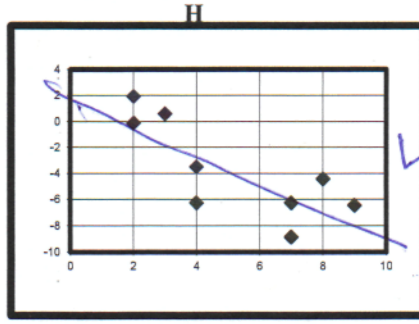
F



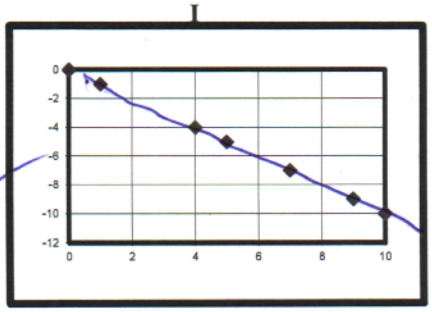
G



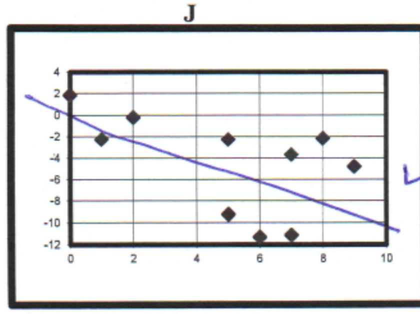
H



I



J



(9 points; 9 minutes)

11. Some people think there should be laws that say children cannot bring ~~the~~ lunch to school from home because schools know better than parents how to feed the kids. A random selection of lunches from ~~the~~ schools is compared to lunches from homes. Use the data below to prepare a 98% confidence interval for the difference in average calories for the populations of all lunches preovided by schools and all lunches provided by parents. There is good reason to believe that lunches from homes are more variable than are lunches from schools.

- do not pool variances
- use smaller d.f.

Data on Calories in Lunches		
	From Homes	From Schools
$\bar{x}$ =	1285	1334
S =	146	112
n =	25	16

df = 24 15

confidence = 0.98

$\alpha = 0.02$ in 2 tails

$t = 2.602$

$$98\%\ CI(\mu_S - \mu_H) = (\bar{x}_S - \bar{x}_H) \pm t\sqrt{\frac{S_S^2}{n_S} + \frac{S_H^2}{n_H}}$$

$$= (1334 - 1285) \pm 2.602\sqrt{\frac{(112)^2}{16} + \frac{(146)^2}{25}}$$

$$= 49 \pm (2.602)(40.5)$$

$$= 49 \pm 105.4$$

$$\boxed{-56.4 < (\mu_S - \mu_H) < 154.4}$$

Based on your confidence interval, is it reasonable to claim that the average calories in all lunches provided by schools is 400 more than the average calories in all lunches that kids bring from home?

Yes **No**

Why? Differences of $(\mu_S - \mu_H) = 400$ are Not in the CI.

If $\mu_S = \mu_H + 400$
then $(\mu_S - \mu_H) = \boxed{400}$

Statistics 300
Instructor: L.C.Larsen

Name: Solution

Summer 2011
M/Tu/W/Th 5:3-8:15 pm

DO THIS PROBLEM OR PROBLEM 13

proportion, %

(10 points; 10 minutes)

12. Use the data below for two randomly selected samples to test this claim: "Cars that run on just gasoline have the same relative frequency of major engine repairs when compared to cars that run on gasoline plus the additive called GoMore." DO NOT DO THIS PROBLEM AS A CONTINGENCY TABLE! (Let α = 0.02 for this test.)

Car needed major engine repairs	(G) Just Gasoline	(GM) Gasoline + GoMore
No	75	42
Yes	8	8
Sample Size	83	50

$\hat{p}_G = 8/83 = 0.0964$ $\quad \hat{p}_{GM} = \frac{8}{50} = 0.1600$

claim: $p_G = p_{GM}$

$(p_G - p_{GM}) = 0$

need $\bar{p}$ & $\bar{q}$

$\bar{p} = \frac{\text{all "successes"}}{\text{all trials}} = \frac{8+8}{83+50} = 0.1203$

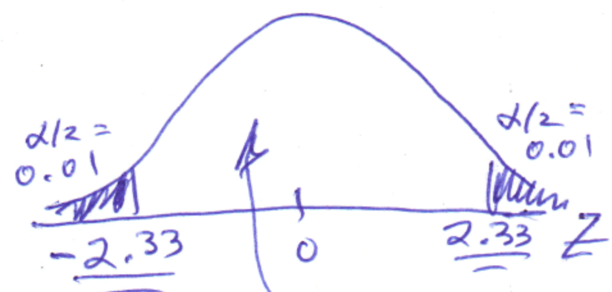
$\bar{q} = 0.8797$

Claim: ______

H_0: $(p_G - p_{GM}) = 0$

H_1: $(p_G - p_{GM}) \neq 0$

$\alpha = 0.02$ in 2 tails



Test Statistic

$$\frac{(\hat{p}_G - \hat{p}_{GM}) - (p_G - p_{GM})_0}{\sqrt{\frac{\bar{p}\bar{q}}{n_G} + \frac{\bar{p}\bar{q}}{n_{GM}}}} = \frac{(.0964 - .1600) - 0}{\sqrt{\frac{(.1203)(.8797)}{83} + \frac{(.1203)(.8797)}{50}}} = \frac{-0.0636}{0.0582} = -1.09$$

~~Do~~ not reject H_0!

Statistics 300
Instructor: L.C.Larsen

Name: Solution

Summer 2011
M/Tu/W/Th 5:3-8:15 pm

DO THIS PROBLEM OR PROBLEM 12

(10 points; 10 minutes)

13. An advertisement says "add GoMore to your gasoline and get at least 2 miles per gallon more than you would get with just gasoline." Use the data from a small test below to test whether the advertised claim is true on average. (Use a Type I error probability or 0.025.)

matched pairs

Experimental Unit (car/driver)	miles per gallon Just Gas	Gas + GoMore	diff GM − G
1	26.1	29.7	3.6
2	31.2	34.9	3.7
3	29.1	32.5	3.4
4	31.8	31.7	−0.1
5	32.9	35.1	2.2
6	24.7	26.8	2.1

Claim: $\mu_{GM} \geq \mu_G + 2$

H_0: $(\mu_{GM} - \mu_G) \geq 2$

H_1: $(\mu_{GM} - \mu_G) < 2$

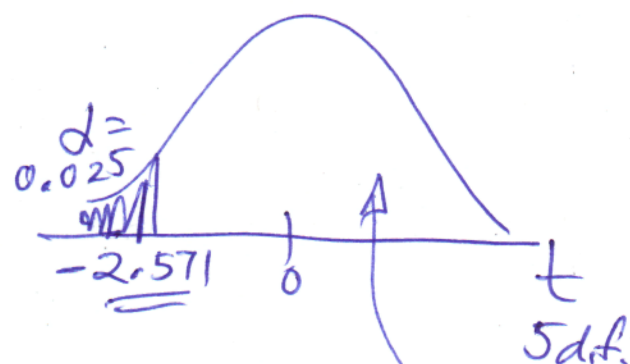
$\alpha = 0.025$ left tail

$\bar{d} = 2.483$

$s_d = 1.447$

$n = 6$

$df = 5$



Test statistic

$$\frac{\bar{d} - \mu_d}{s_d/\sqrt{n}} = \frac{2.483 - 2}{1.447/\sqrt{6}} = \frac{0.483}{0.591} = 0.817$$

Do not reject H_0.