Final Exam

Statistics 300:
Introduction to Probability and Statistics

Summer Semester 2011
Cosumnes College

Instructor: L.C. Larsen

Instructions

Time: 2 hours and 10 minutes

Materials: Open book, notes, homework, etc.

Instruments: Calculator/Laptop of student's choice

No phones or consultants

Answers to confidence interval problems must include the expression (the formula) in symbolic form and the expression with all of the values inserted in the proper places. Then, the final answer can be calculated by any method or device.

Unless a p-value is given in the problem, each hypothesis test problem must include all four parts of the traditional approach to hypothesis tests, including the expression (the formula) for the test statistic in symbolic form and the expression with the values in the right places. The result can then be calculated by whatever method you like (TI-83, laptop computer, etc.).

If more space is needed for a problem, continue your work on the back of the page.

(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.**

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

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?** _____

_____

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

Claim: _____

H_0: _____

H_1: _____

(4 points; 2 minutes)

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

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

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

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

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 58 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 =									Data Hidden
St.Dev. =									
n =	9	10	6	8	6	7	5	7	

Complete the Analysis of Variance Table and test the claim.

Source	Sum of Squares	df	Mean Square	F
Recipes	89.2271			
Error	128.728			
Total				

Claim: _____

H_0: _____

H_1: _____

(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: _____

H_0: _____

H_1: _____

(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?

(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?

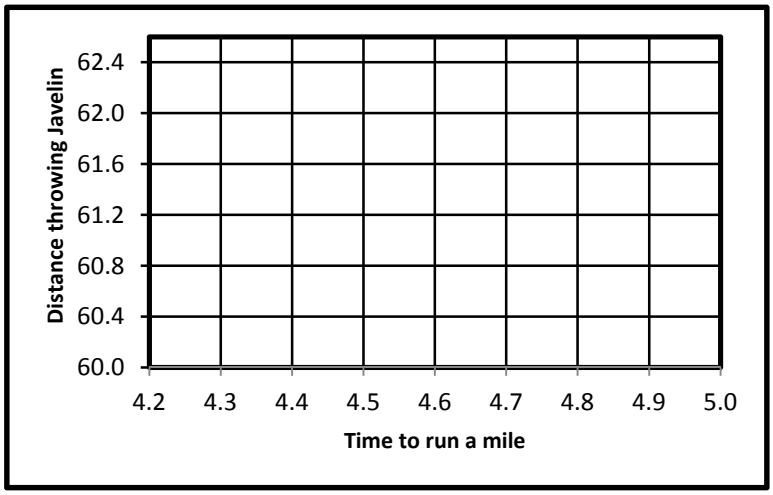
(c) What <u>percentile</u> did the test statistic in part (a) represent? Test statistic = $P_{(\ \ )}$
It may help to draw a critical region picture that includes the p-value.

Statistics 300 **Summer 2011**
Instructor: L.C.Larsen **Name:** _____ **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**



(a) Plot the data on the graph.

(b) Write the equation of the regression "best fit" line. _____

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

(c) Predict "Javelin Distance" if "Time for Mile" is 4.5 . _____

(d) What is the sample correlation between mile time and Javelin distance? _____

(e) What fraction of the variation in "distance" does "mile time" explain? _____

(f) What is the expression for the "total variation in Y"? _____

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

(h) What is the expression for the "explained variation in Y"? _____

(i) What is the value of the "explained variation in Y"? _____

(j) What is the expression for the "unexplained variation in Y"? _____

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

(l) What is the expression for the "Standard Error of Estimate"? _____

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

(9 points; 10 minutes)

7. **Random samples of 200 women and 200 men were asked what size of car drove -- small, medium, or large. 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.)

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

Claim: _____

H_0: _____

_____

H_1: _____

_____

(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%**

Claim: _____

H_0: _____

H_1: _____

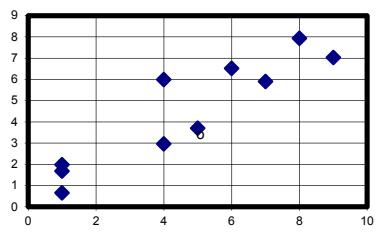
Statistics 300 **Summer 2011**
Instructor: L.C.Larsen **Name:** _____ **M/Tu/W/Th 5:3-8:15 pm**

(10 points; 6 minutes)

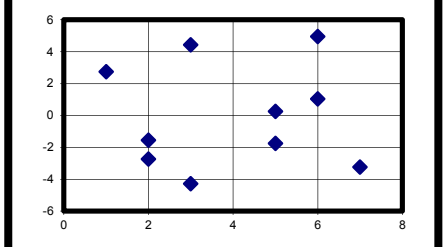
10. Assign the letters of the appropriate figures to each of the "sample correlation" values offered below. If an "r" value has no appropriate figure, write "none" next to it.

value of "r"	Figure Letter(s)
1.50	
1.00	
0.90	
0.70	
0.00	
- 0.50	
- 0.70	
- 0.90	
- 1.00	

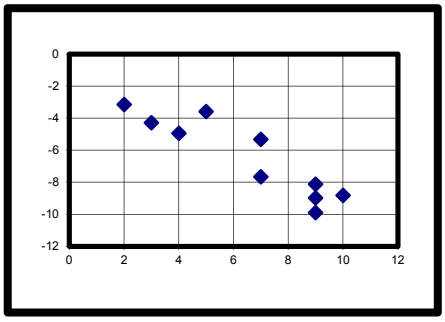
A



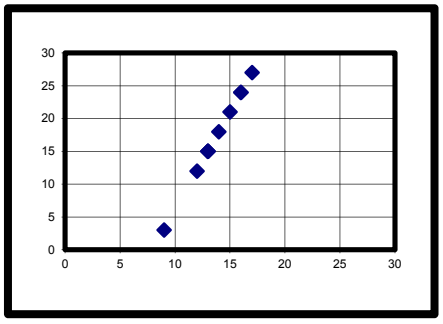
B



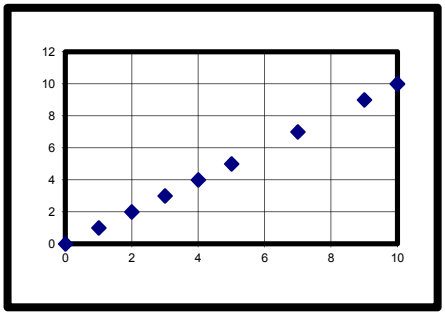
C



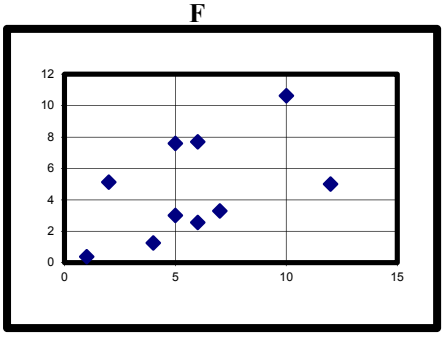
D



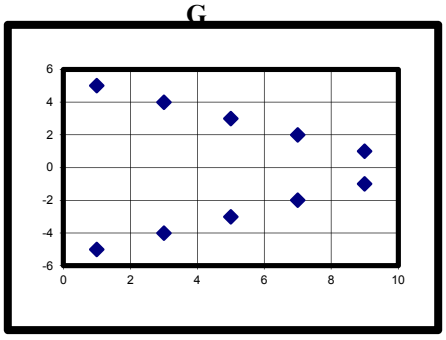
E



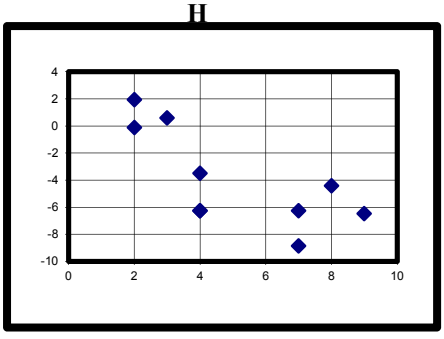
F



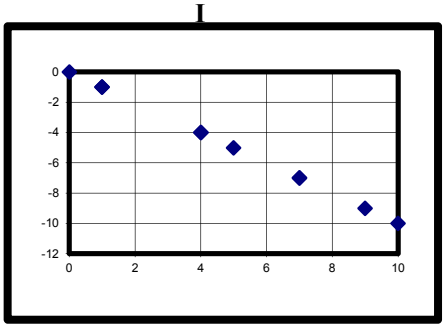
G



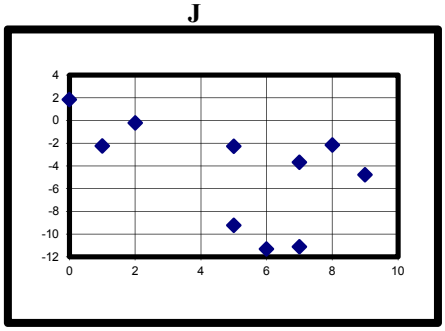
H



I



J



(9 points; 9 minutes)

11. Some people think there should be laws that say children cannot bring lunch to school from home because schools know better than parents how to feed the kids. A random selection of lunches from 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 provided 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.

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

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?** _____

_____

DO THIS PROBLEM OR PROBLEM 13

(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 $\alpha = 0.02$ for this test.)

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

Claim: _____

H_0: _____

H_1: _____

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.) **Only 6 experimental units were used in the study.**

Experimental Unit (car/driver)	miles per gallon	
	Just Gas	Gas + GoMore
1	26.1	29.7
2	31.2	34.9
3	29.1	32.5
4	31.8	31.7
5	32.9	35.1
6	24.7	26.8

Claim: _____

H_0: _____

H_1: _____