Final Exam

Statistics 300:
Introduction to Probability and Statistics

Fall 2012
Cosumnes College

Instructor: L.C. Larsen

Instructions

Time: 2 hours and 5 minutes

Materials: Open book, notes, homework, etc.

Instruments: Calculator/Laptop of student's choice

No phones or consultants
Except to call the instructor : 346-6324.

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 (for AOV the Table is the formula 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.

Statistics 300 **Fall 2012**
Exam #3 **Name:** _____ **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			Row Total
	AZ	CA	WA	
Phoenix Suns	129	29	26	184
Sacramento Kings	40	129	16	185
Seattle Sonics	33	21	177	231
	202	179	219	600

H_0: _____

H_1: _____

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

(8 points; 8 minutes)

2. Engineering students prepared a car 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 speed and fuel use are negatively correlated. Use a Type I Error Rate of 0.025 for this test.

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: _____

_____

H_0: _____

H_1: _____

Statistics 300 **Fall 2012**
Exam #3 **Name:** _____ **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.

Claim: _____

H_0: _____

H_1: _____

Gallons of Milk per Day		
Cow	**January**	**July**
1	**5.3**	**5.1**
2	**5.4**	**6.8**
3	**6.0**	**6.9**
4	**5.9**	**6.9**
5	**6.8**	**7.3**
6	**4.8**	**5.8**
7	**5.5**	**6.3**
8	**6.3**	**6.9**
mean =	**5.75**	**6.50**
stdev =	**0.63**	**0.73**
n =	**8**	**8**

Statistics 300 **Fall 2012**
Exam #3 **Name:** _____ **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.

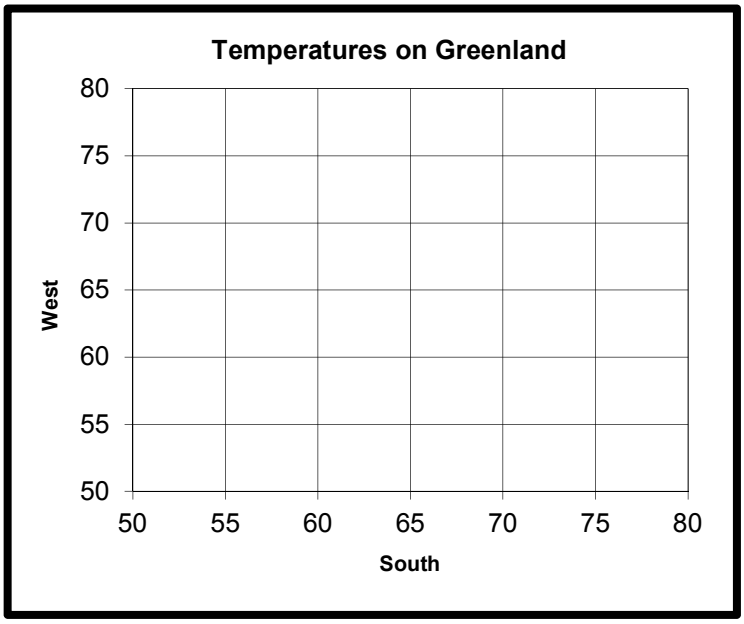
Age in Years when autism diagnosed	Family had Health Insurance		
	Yes	No	Total
< 1	71	29	100
1	69	31	100
2	69	31	100
3	77	23	100
4	70	30	100
> 4	77	23	100
Total	433	167	600

Statistics 300 **Fall 2012**
Exam #3 **Name:** _____ **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	North	South
53	59	58	61
78	80	79	78
53	59	51	66
74	79	78	79
79	63	66	68
53	54	50	53
80	69	76	72
53	74	58	64
58	65	61	57
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 : _____

(c) Plot your line on the graph.

(d) What is the linear correlation for the given North and East data? _____

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

_____ = _____ + _____
Total Variation Explained Variation Unexplained Variation

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

_____ = _____ + _____
Total Variation Explained Variation Unexplained Variation

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

_____ _____
Symbolic Expression Value

(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:

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

- O The p-value was less than 0.10
- O The critical value was from the Z table
- O The critical value was for 0.025 in the right tail
- O 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:

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

(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.

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

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?** _____

_____

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?** _____

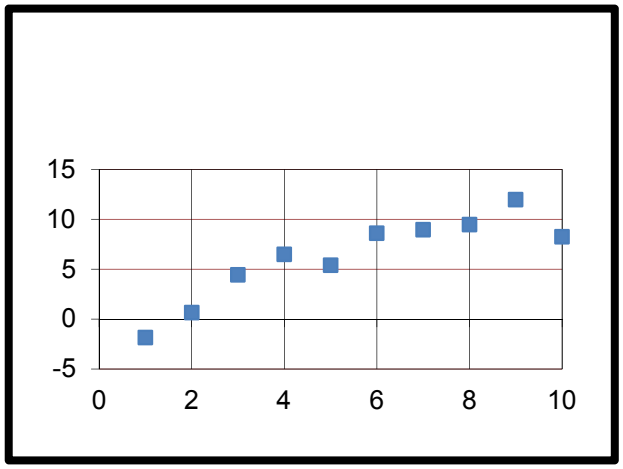
_____

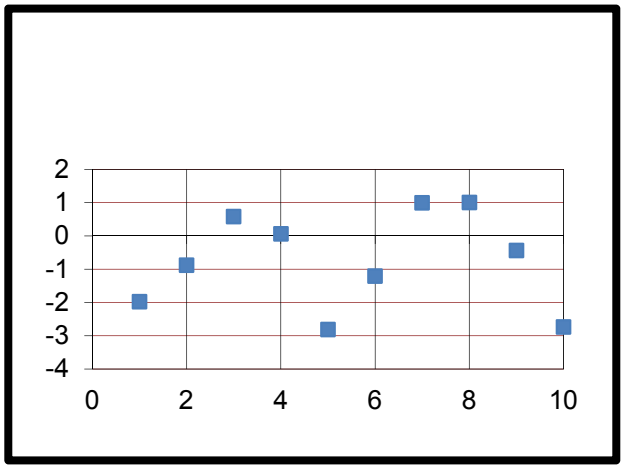
Statistics 300 **Fall 2012**
Exam #3 **Name:** _____ **Saturdays 9:00 am - 1:05 pm**

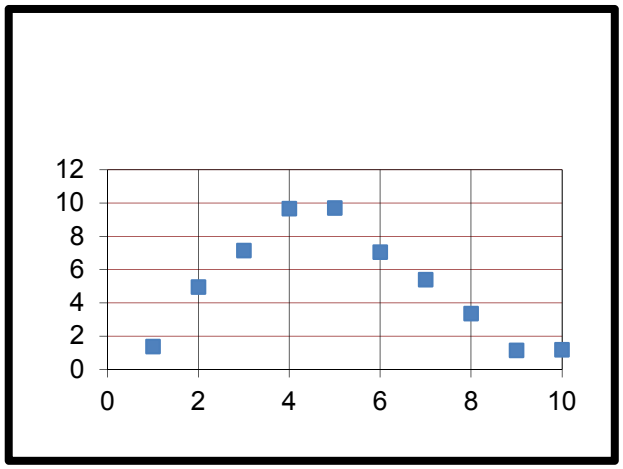
(6 points; 6 minutes)

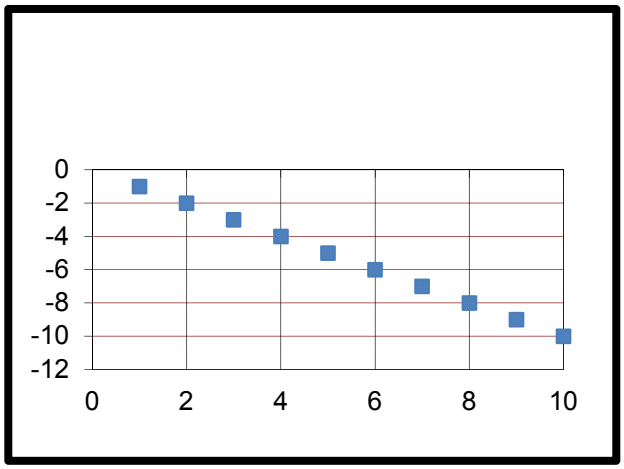
8. Connect each picture with one of the candidate "r" values by writing the appropriate candidate "r" value in the space at the top of each graph.

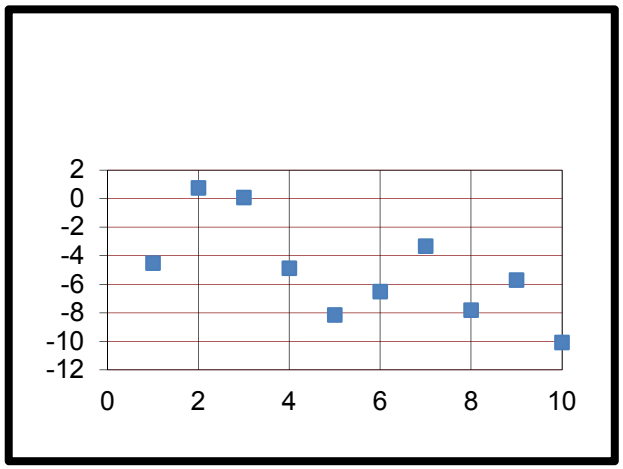
Candidate values of "r", the sample correlation coefficient.
0.00 -0.70 -0.90 -1.00 0.70 0.90 1.00

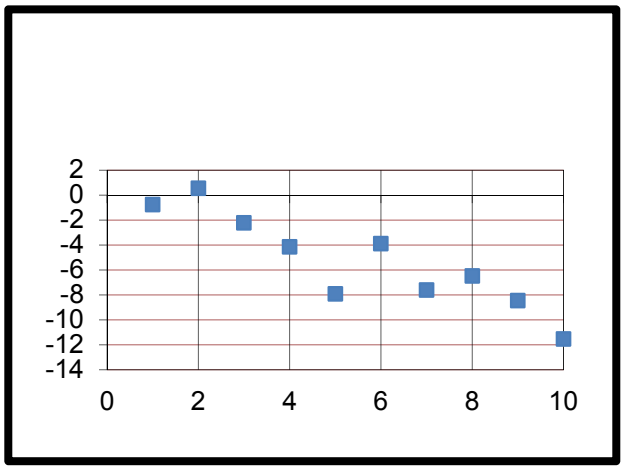












(9 points; 9 minutes)

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**
Meat-eaters	**53**	**347**
Vegans	**48**	**352**

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?** _____

_____

_____

(9 points; 7 minutes)

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.

AOV Table

Source	SS	df	MS	F	p-value
Producer			172.111		0.282018
Error					
Total	45292				

H_0: _____

H_1: _____

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

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

(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.

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

Claim: _____

H_1: _____

H_1: _____

(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
Sun	360
Mon	418
Tue	513
Wed	465
Thu	378
Fri	601
Sat	641
Total	3376

H_0: _____

_____

H_1: _____

_____