Stat 104 – Final Exam Name: _____
December 13, 2011

INSTRUCTIONS: Read the questions carefully and completely. Answer each question and show work in the space provided. Partial credit will not be given if work is not shown. When asked to explain, describe, or comment, do so within the context of the problem. Be sure to include units when dealing with quantitative variables.

1. **[14 pts]** Short answer.

   a) [3] Statistics is about … _____. (Fill in the blank with one word.)

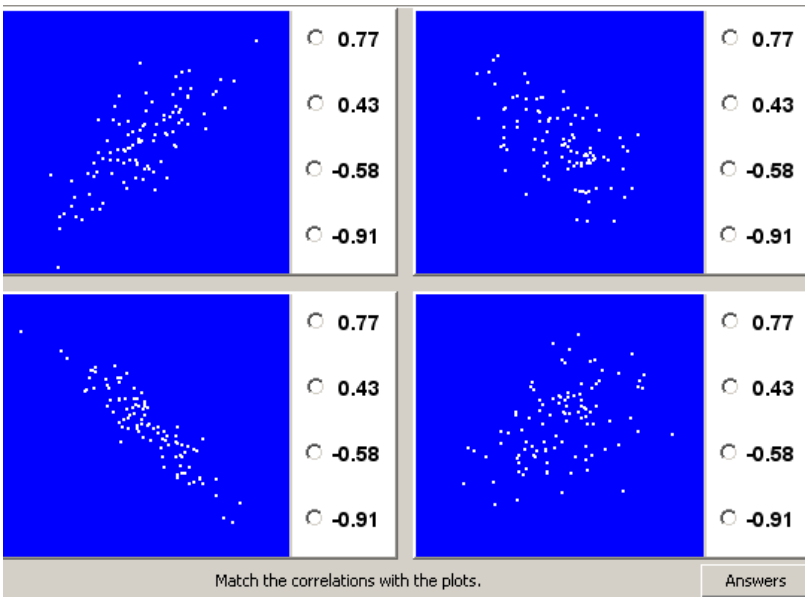
   b) [2] A _____ is a numerical summary of a population,

      while a _____ is a numerical summary of a sample.

   c) [3] The sample median of 11 numbers is 20. If one of those numbers is changed from 43 to 21, what is the value of the new sample median?

   d) [3]The sample mean of 11 numbers is 20. If one of those numbers is changed from 43 to 21, what is the value of the new sample mean?

   e) [3] Match the correlation with the plot.



2. **[21 pts]** Multiple Choice:

a) ___ For testing the hypothesis $\boldsymbol{H_0 : p = 0.2}$ against $\boldsymbol{H_0 : p \neq 0.2}$, the value of the test statistic, based on a random sample of 100, is z = 1.78. The P-value for the test is:
A: 1.9250 B: 0.9625 C: 0.0375 D: 0.0750 E: 0.0500

b) ___ The correct interpretation of a 95% confidence is?
A: I am 95% sure I don't know the correct answer.
B: 95% of the sample values are in the confidence interval.
C: 95% of the population values are in the confidence interval.
D: If many confidence intervals are constructed, using random samples, approximately 95% of the intervals will contain the population parameter.

c) ___ The P-value is …?
A: The probability that the null hypothesis is rejected.
B: The probability that the null hypothesis is true.
C: The probability of observing a value of the test statistic more extreme than the one observed when the null hypothesis is true.
D: The probability that the null hypothesis is not rejected.

d) ___ You have calculated the correlation coefficient between two numerical variables to be $r = +1.34$. This would indicate?
A: A strong linear relationship.
B: A weak linear relationship.
C: No linear relationship.
D: You have calculated the correlation coefficient incorrectly.

e) ___ Which of the following is not a measure of center?
A: Midrange.
B: Median.
C: Interquartile range.
D: Median.

f) ___ Holding all other things constant, if you increase the sample size the confidence interval will
A: get wider.
B: not change.
C: get narrower.
D: None of the above.

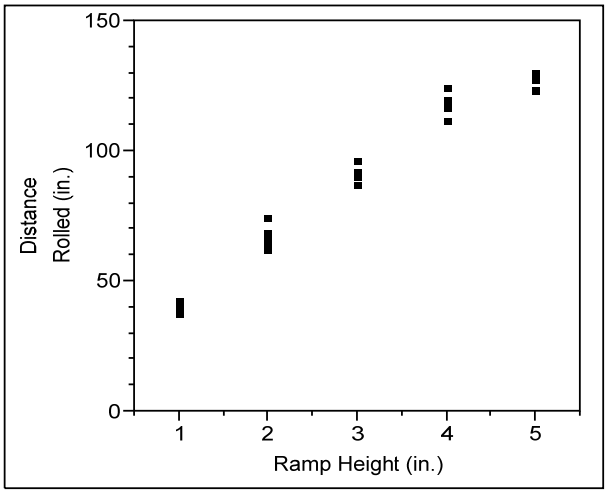
g) ___ Holding all other things constant, if you increase the level of confidence the confidence interval will
A: get wider.
B: not change.
C: get narrower.
D: None of the above.

3. [**20 pts**] For their project in an introductory statistics class a group of students looked at the relationship between the height of a ramp (inches) and the distance a golf balled rolled (inches) when released from the given height.

   a) [2] What is the explanatory variable? What type of variable, categorical or numerical, is it?

   b) [2] What is the response variable? What type of variable, categorical or numerical, is it?

   c) [3] Below is a plot of distance rolled versus ramp height.



   Describe the relationship between ramp height and distance rolled.

   Below is partial JMP output for the least squares regression line.

Linear Fit

Distance Rolled (in.) = 22.43 + 22.492*Ramp Height (in.)

Summary of Fit

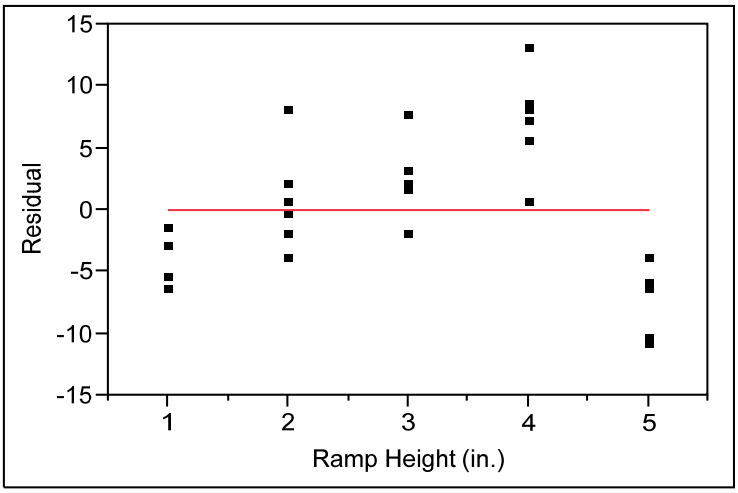
RSquare	0.96329
RSquare Adj	0.96198
Root Mean Square Error	6.42733
Mean of Response	89.9
Observations (or Sum Wgts)	30

d) [4] Give an interpretation of the estimated slope within the context of the problem.

e) [3] Use the least squares regression line to predict the distance rolled for a golf ball rolled from a height of 2.5 inches.

f) [2] What is the value of the correlation coefficient?

g) [4] Below is a plot of residuals versus the ramp height. Describe what you see in the plot and what this tells you about the straight line fit?



4. [**10 pts**] A study was done on the effectiveness of antidepressants. Three hundred individuals who had suffered depression in the past but were currently not depressed were assigned to one of three groups at random. The individuals in one group got Desipramine, the individuals in another group got Lithium and individuals in the third group got a Placebo. After a month, individuals were assessed to see if they had relapsed into depression. Below are the data.

	Relapse?		
Drug	No	Yes	Total
Desipramine	58	42	100
Lithium	25	75	100
Placebo	17	83	100
Total	100	200	300

a) [3] Why is this study an experiment?

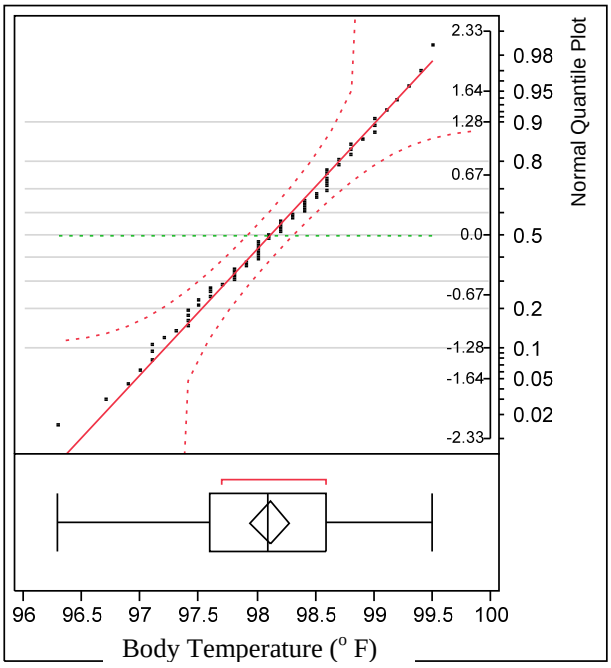
b) [2] What proportion of the 100 individuals given Desipramine did not have a relapse?

c) [5] Construct an 80% confidence interval for the population proportion of individuals who would not suffer a relapse if they were given Desipramine.

5. **[5 pts]** For the Spooky Loot® scratch ticket game, the proportion of winning tickets (tickets where you get a prize of at least $2) is 0.274. Consider purchasing a random sample of 100 tickets from the population of 1,080,000 tickets available. Describe the **sampling distribution** of the **sample proportion** of winning tickets?

- Shape
- Center (Mean)
- Variability (Std Dev)

6. [**10 pts**] A random sample of 65 men had their temperature measured in ^{o}F. Graphical and numerical summaries for the 65 men are given below.



100.0%	maximum	99.5
75.0%	quartile	98.6
50.0%	median	98.1
25.0%	quartile	97.6
0.0%	minimum	96.3
	Mean	98.104615
	Std Dev	0.6987558
	Std Err Mean	0.08667
	N	65

a) [3] Looking at the Normal Quantile Plot describe what you see and how this supports the assumption that the population distribution of men's body temperature is normally distributed.

b) [3] Looking at the Box Plot describe what you see and how this supports the assumption that the population distribution of men's body temperature is normally distributed.

c) [4] Construct a 98% confidence interval for the population mean body temperature of all men.

7. **[15 pts]** Wearing a seat belt can save your life if the car you are driving is involved in an accident. In Massachusetts, 67 percent of drivers wear seat belts. The proportion of drivers that use seat belts depends on things like age, gender, ethnicity, and local law. For a random sample of 117 Hispanic drivers in Boston, Massachusetts, 68 were wearing seat belts. Does this sample provide convincing evidence that Hispanic drivers in Boston are less likely to wear seat belts than Massachusetts drivers in general? Fill in the missing steps of the following hypothesis test.

   a) [4] Step 1: Hypotheses – set up an appropriate null and alternative hypothesis. Be sure to clearly define, in words, what the population parameter is.

   b) Step 2: Conditions

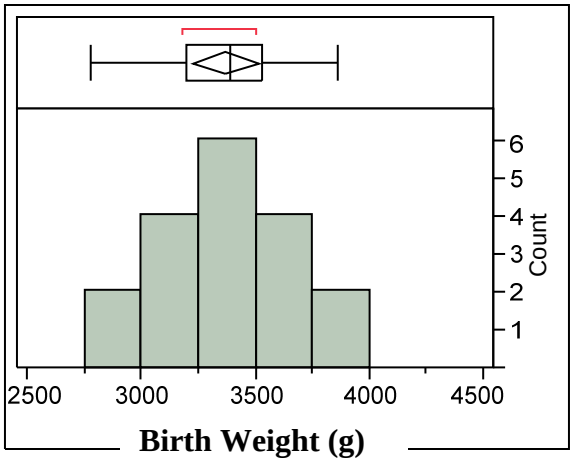
      The conditions necessary for conducting the test of hypotheses are satisfied.

   c) [6] Step 3: Sample Evidence – calculate the sample proportion, the value of the test statistic, z, and use the appropriate table to find the P-value.

   d) [2] Step 4: Decision – Use the P-value to reach a decision.

   e) [3] Step 5: Conclusion – State a conclusion within the context of the problem.

8. **[15 pts]** A random sample of births at a local hospital included 18 baby girls. The birth weight (grams) of these 18 baby girls is summarized below. We want to use these data to see if the population mean birth weight of females is 3175 g (7 pounds) against an alternative that it is greater than 3175 g.



				t Test	
Mean	3368.44	Hypothesized Value	3175	Test Statistic	2.7652
Std Dev	296.80	Actual Estimate	3368.44	Prob > |t|	0.0132*
Std Err Mean	69.96	DF	17	Prob > t	0.0066*
Upper 95% Mean	3516.04	Std Dev	296.8	Prob < t	0.9934
Lower 95% Mean	3220.84				
N	18				

a) [3] Describe the distribution of birth weight. Does this support the condition that the population distribution of female birth weights is normal?

b) [2] Set up the null and alternative hypotheses using appropriate statistical notation.

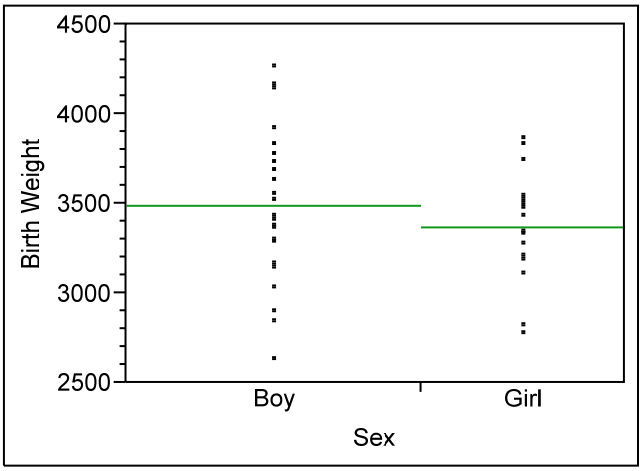
c) [3] Give the value of the test statistic and the P-value.

d) [2] Use the P-value to reach a decision.

e) [3] State a conclusion within the context of the problem.

f) [2] Give the 95% confidence interval for the population mean birth weight of females. Hint: It is given in the JMP output.

9. **[10 pts]** The random sample of births in problem 8 included 26 baby boys. Data on birth weights (grams) for both girls and boys is given below.



Level	Number	Mean	Std Dev	Std Err Mean	Lower 95%	Upper 95%
Boy	26	3488.77	403.304	79.094	3325.9	3651.7
Girl	18	3368.44	296.800	69.956	3220.8	3516.0

a) [2] Which group, boys or girls, tend to have the greater birth weights? Support your answer using the data.

b) [2] Which group, boys or girls, tend to have more variability in birth weights? Support your answer using the data.

c) [4] If we wanted to test the hypothesis that there is no difference in population mean birth weight between girls and boys against the alternative that boys tend to be heavier than girls at birth (on average), the test statistic turns out to be –1.14 and a P-value of 0.1305. What does this tell you about the difference between girls and boys birth weights? Support your answer statistically.

d) [2] At birth Dr. Bob weighed 5329.7 grams. What is Dr. Bob's standardized birth weight?

10. **[5]** The population distribution for the age of male students in introductory statistics classes has a population mean of 20.7 years and a population standard deviation of 1.73 years. A random sample of n = 100 male students is selected from introductory statistics classes. Describe the **sampling distribution** of the **sample mean age** of female students.

    - Shape
    - Center (Mean)
    - Variability (Std Dev)