22S:30/105, Statistical Methods and Computing
Spring 2012, Instructor: Cowles
Final Exam

Name: _____ Course no. (30 or 105) _____

1. Identify the variable type (binary, nominal, ordinal, quantitative discrete, quantitative continuous) of each of the following variables:

   (a) student's grades on a history test (A,B,C,D,F)

   (b) breeds of dogs (German shepherd, chihuahua, collie, etc.)

   (c) whether adults have completed high school (yes or no)

2. A study examined the relationship between the percent of adult birds in a colony who return from the previous year and the number of new adults who join the colony. The data were taken from a graph in Saether, B-E., Engen, S., and Mattysen, E. 2002. Demographic characteristcs and population dynamical patterns of solitary birds. Science, 295, pp. 2070-2073 and are used in Moore's Basic Practice of Statistics.

   For each of 13 colonies of sparrowhawks, the variables are:

   - `percent`:percent return
   - `new`: new adults

   (a) From the SAS output, find the point estimate and 95% confidence interval for the slope. Give the numbers here.

   (b) Explain briefly what this slope means about returning adult birds and new adults. (If you couldn't do the previous question, pretend that the point estimate was -1.0 and the 95% c.i. was (-1.5,-0.5).

   (c) Note that the value of the predictor variable in the first observation is 74. Give a 95% prediction interval for the number of new adults joining the colony in a future year in which the percent returning is 74. (Numeric answer.)

(d) We discussed several important things to look for in a residual plot. Identify two of them, and state whether the residual plot shown in the output indicates any problems with the linear regression.

(e) Using conventional statistical symbols, state the null hypothesis that there is no linear relationship between `percent` and `new`.

(f) Can we reject that null hypothesis at significance level $\alpha = 0.05$? Justify your answer using two different parts of the SAS output.

(g) What is the sample correlation coefficient r between `percent` and `new`? (Numeric answer)

3. For each of the following statistical analyses, tell whether ANOVA, a Chi-square test, a paired t-test, or a two-independent-sample t-test would be most appropriate. Briefly justify your answers.

   (a) Researchers wish to investigate whether the mean tail length is is different in two different species of monkeys. They obtain 50 monkeys of each species and measure the tail of each animal.

   (b) Researchers wish to investigate whether the proportion of infertile females is different in four different species of monkeys. They identify 20 female monkeys of each species and determine whether each is infertile.

   (c) Researchers wish to investigate whether people's systolic blood pressure changes

during exercise. They recruit 25 joggers and measure each person's systolic blood pressure immediately before and immediately after a 15-minute jog.

4. Circle all of the statements below that are true.

   (a) The power of a statistical test is the probability of rejecting the null hypothesis if it is false.

   (b) Increasing the sample size increases the power of a test (everything else held constant).

   (c) We prefer to have small power in a statistical test.

   (d) Power is the probability of type I error.

5. Researchers wish to test whether the population mean height of U.S. 6-year-old boys is less than 40 inches. Their hypotheses are:

$$H_0: \quad \mu \geq 40$$
$$H_A: \quad \mu < 40$$

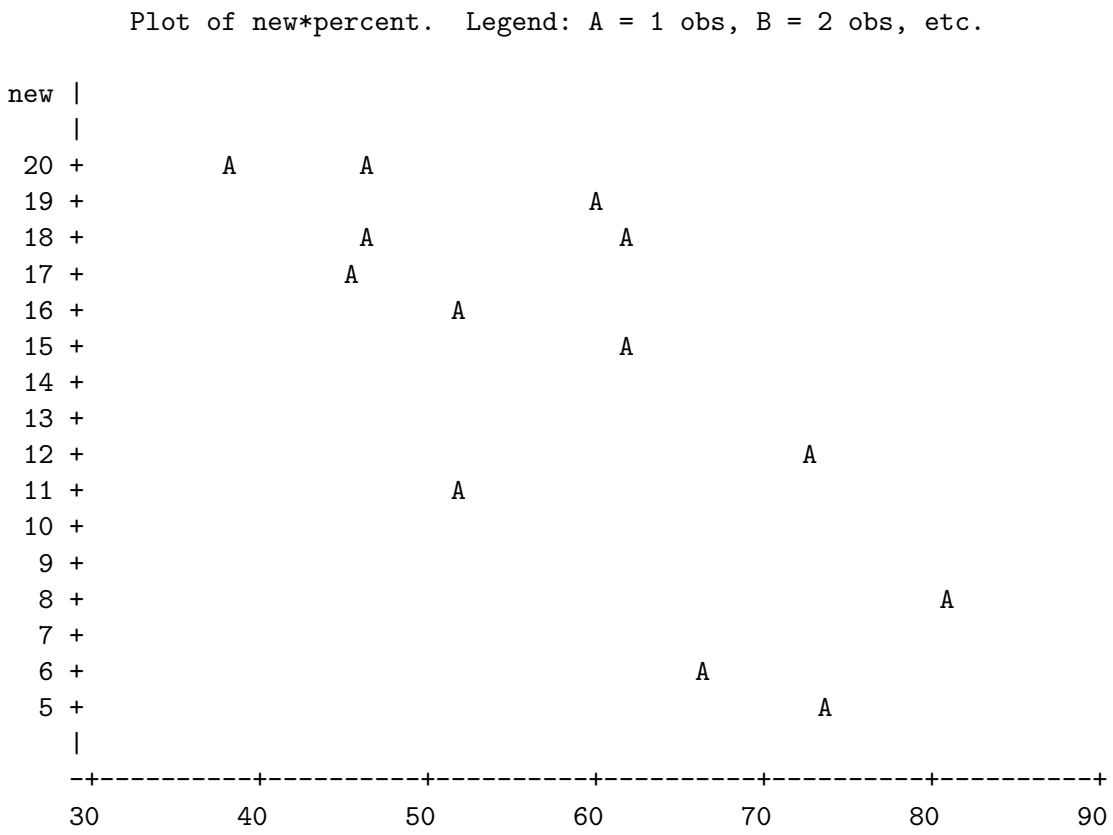
   They will measure a random sample of 64 boys. They are convinced that heights in this population follow a normal distribution with $\sigma = 4$.

   (a) Is this a one-sided or a two-sided test? Explain.

   (b) Will large or small values of $\bar{x}$ lead to rejecting H_0?

   (c) Determine the critical value–the value of $\bar{x}$ such that anything equal to or more extreme than it will lead to rejecting H_0. Numeric answer; show your work.

   (d) What is the power of this test against the alternative that $\mu = 39$?



The REG Procedure
Model: MODEL1
Dependent Variable: new

Number of Observations Read	13
Number of Observations Used	13

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	188.40068	188.40068	14.01	0.0032
Error	11	147.90701	13.44609		
Corrected Total	12	336.30769			

Root MSE	3.66689	R-Square	0.5602
Dependent Mean	14.23077	Adj R-Sq	0.5202
Coeff Var	25.76734		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	31.93426	4.83762	6.60	<.0001
percent	1	-0.30402	0.08122	-3.74	0.0032

Parameter Estimates

Variable	DF	95% Confidence Limits	
Intercept	1	21.28674	42.58178
percent	1	-0.48279	-0.12526

The REG Procedure
Model: MODEL1
Dependent Variable: new

Output Statistics

Obs	percent	Dependent Variable	Predicted Value	Std Error Mean Predict	95% CL Mean	
1	74	5.0000	9.4366	1.6355	5.8370	13.0362
2	66	6.0000	11.8687	1.1969	9.2345	14.5030
3	81	8.0000	7.3084	2.1105	2.6632	11.9536

Output Statistics

Obs	percent	95% CL Predict		Residual
1	74	0.5994	18.2737	-4.4366
2	66	3.3789	20.3586	-5.8687
3	81	-2.0037	16.6205	0.6916

The REG Procedure
Model: MODEL1
Dependent Variable: new