Statistics 31, Section 1, Final Examination
Tuesday, May 10, 2005

Name: _____

Pledge: I have neither given nor received aid on this examination.

Signature: _____

Instructions: Do not do any actual numerical calculations. Answers in a form that you would type into an Excel field, such as "=28*SQRT(82)^2", with a *working* answer, are expected).

1. State whether each boldface number is a parameter or a statistic:

a. A carload of ball bearings has a mean diameter of **2.496** cm. This is within specifications for acceptance of the lot by the purchaser. The inspector happens to inspect **100** bearings from the lot with a mean diameter of **2.483 cm**. This is outside the specification limits, so the lot is mistakenly rejected.

b. The Carolina Abecedarian Project investigated the effect of high quality preschool programs on children from poor families. Children were randomly assigned to two groups. One group participated in a year round preschool program from age three months. The control group received social services, but no preschool. At age 21, **35%** of the treatment group, and **14%** of the control group were attending a 4 year college, or had already graduated from college.

2. A company makes 30% of its cars at factory A, 20% at Factory B, and the rest at Factory C. Factory A produces 20% lemons, Factory B produces 30% lemons, and Factory C produces 10% lemons. A car is chosen at random. What is the probability that:

a. It came from Factory A or from Factory B?

b. It came from Factory C?

c. It is a lemon if it came from Factory C?

d. It is a lemon from Factory C?

e. It is a lemon?

f. It came from Factory C, if it is a lemon?

3. Some measurements of oxygen uptake (VO2) as a function of Heart Rate (HR), for a single individual, resulted in the following EXCEL Regression analysis:

	C	D	E	F	G	H	I	J	K
16		*Coefficients*	*andard Err*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*ower 99.0%*	*pper 99.0%*
17	Intercept	-8.07	2.526866	-3.19368	0.085637	-18.9422	2.802233	-33.1491	17.00911
18	X Variable	0.09	0.026458	3.40168	0.076619	-0.02384	0.203838	-0.17259	0.352591

a. What are the least squares estimates of the slope and intercept?

b. Give a 99% confidence interval for the slope of the regression line.

c. Test the null hypothesis that the slope of the regression line is 0. Interpret the result from the gray level point of view.

d. Write an EXCEL command to predict the oxygen uptake for this individual, when the Heart Rate is 98.

e. Suppose that there were 10 measurements made, that the answer to part d is in the cell B33, that the EXCEL standard error is in the cell C15, and that… → $\frac{(98-\bar{x})^2}{\sum_i (x_i-\bar{x})^2} = 1.7$
Write an EXCEL command to calculate a 95% Prediction Interval for a new measurement of VO2, when the Heart Rate is 92.

f. Under the assumption of part e, write an EXCEL command for a 90% Confidence Interval for the mean value of VO2, when the Heart Rate is 98.

4. Researchers planning a study of 3rd grade children are studying the mean performance on a reading test. A small pilot study finds a sample standard deviation of s = 10 points, so in preliminary studies they take the population s. d. to be 10.

a. Available funding allows 25 students. Calculate the margin of error of the 95% confidence interval for the population mean based on a sample size of 25.

b. There are many other demands on the research budget, some of which require a downward revision of the budget. What is the margin of error if only 9 students can participate because of this budgetary limitation?

c. Find the smallest value of the sample size that would satisfy the goal of a 95% margin of error of 3 or less.

5. The lengths of horse pregnancies are approximately normally distributed with mean 200 days, and standard deviation 10 days. Use the 68-95-99.7 rule (and *not* EXCEL), to:

a. Give values between which the middle 99.7% of pregnancies fall.

b. How long do the longest 2.5 % of pregnancies last?

6. A basketball player made only 30% of free throws last season. After practicing free throw shooting all summer, in the first 8 games this season, she made 20 free throws in 40 attempts.

a. Formulate H_0 and H_A (in terms of population parameters and numbers) for testing that her free throw rate is indeed different from what it was last season.

b. Calculate a P-value for the hypotheses in part a.

c. If the numerical value of the answer to part b had been 0.053, interpret the result from the gray-level viewpoint.

d. Give a 90% best guess Confidence Interval for her free throw percentage for the new season.

e. Give a 90% conservative Confidence Interval for her free throw percentage for the new season.

f. What assumptions are needed for these calculations to be correct?

7. A study of the survey response rates between letters and phone calls gave the following results (together with some preliminary calculations):

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Reponse	Letter	Call			Prop'ns				E prop'ns				E counts		
2	Yes	30	50	80		0.167	0.28	0.44		0.22	0.22	0.44		40	40	80
3	No	60	40	100		0.333	0.22	0.56		0.28	0.28	0.56		50	50	100
4		90	90	180		0.5	0.5	1		0.5	0.5	1		90	90	180

a. Formulate hypotheses for whether or not the survey type has an effect on response rate.

b. Write an EXCEL command for computing the value in the cell F2, based upon the values in cells to the left of that.

c. Write an EXCEL command for computing the value in the cell J2, based upon the values in cells to the left of that.

d. Write and EXCEL command for computing the value in the cell N2, based upon values in cells to the left of that.

e. Write an EXCEL command to compute the P-value, based on the chi-square test of the hypotheses in part a.

f. If the P-value in part f is 0.006, interpret the result from the yes-no viewpoint.

g. What are the degrees of freedom for the Chi Square distribution used in part f?

8. An investigator had 5 right handed students turn a knob (with their right hands). Once clockwise, and again counter clockwise. The table below gives the times required (in seconds) to move an indicator a fixed distance.

	A	B	C	D	E	F
1	Subject	1	2	3	4	5
2	Right Thread	112	116	108	110	113
3	Left Thread	121	128	118	121	124
4	Difference	9	12	10	11	11

a. How should randomization have been used in this experiment?

b. Formulate (in terms of numbers) appropriate null and alternative hypotheses to test that there is a significant difference between the times required for the left and right threaded knobs.

c. Write an EXCEL command for the t-statistic to assess the strength of the data for the hypotheses in part b.

d. If the answer to part c is in the cell B6, write an EXCEL command to calculate the P-value for testing the hypotheses in part b.

e. If the answer to part d is 0.49, what do you conclude from the gray level point of view?