Stat 2300 International, Fall 2006 – Final

Friday, December 8, 2006

Your Name: _____

A Number: _____

The Final consists of 40 questions: 20 multiple-choice questions (with exactly 1 correct answer) and 20 text-based questions where you have to provide a verbal explanation or calculate one or multiple numerical values. Some of the questions require you to use Excel. For other question, you can decide yourself whether you want to use Excel, a calculator, or do the calculations by hand.

The exam is worth a total of **500 points**. The number of points for each question is indicated in parentheses at the beginning of each question. You have approximately **180 minutes** to complete the exam. Try to correctly answer as many questions as possible during this time period. **You are allowed to answer questions in any order.** Start with a question that seems the easiest for you. If you cannot answer a question within a short time, move to another question, and come back to the previously unanswered questions toward the end of the exam.

Mark your answers to the multiple choice questions and fill in the spaces for the text-based questions on the **answer sheets**. Do not write your answers on the pages with the questions. However, you can use those pages for personal comments and calculations. Make sure to write your name and your A number on the pages with the questions **and** on the answer sheets. You have to turn in the pages with the questions **and** the answer sheets.

For multiple choice questions, mark exactly one of the choices (representing options a, b, c, or d) with a solid dot. Only 1 answer is correct. If you mark more than 1 of the options, this will automatically be an incorrect answer (even if one of the options you have marked is correct). For the text-based questions, indicate the formula you are using, the numerical values you have to fill in, and the final result (e.g., $n*p = 10*0.2 = 2$). Just the (correct) final result will not provide you with the full points for this question. If you use Excel to calculate a result, indicate which feature you have used (e.g., Summary Statistics) and the result from within Excel. Please do not write outside the text boxes.

In case of any problems with Excel, please inform your local instructor as quickly as possible and try to answer the questions as far as possible. Return to these questions later on. In case Excel is unavailable for an extended time period, your exam score will be adjusted accordingly.

1. **(10 Points)** When carrying out a sample test (with σ known) of H_0: $\mu = 10$ *vs*. H_a: $\mu \neq 10$ by using a *p*-value, we reject H_0 at level of significance α if and only if the *p*-value is

    a. Greater than $\alpha/_2$
    b. Greater than α
    c. Less than α
    d. Less than $\alpha/_2$

2. **(10 Points)** Which of the following is a categorical variable?

    a. Air Temperature
    b. Bank Account Balance
    c. Daily Sales in a Store
    d. Whether a Person Has a Traffic Violation

3. **(10 Points)** Type II error is defined as the probability of _____ H_0 , when it should _____

    a. failing to reject, be rejected.
    b. failing to reject, not be rejected.
    c. rejecting, not be rejected.
    d. rejecting, be rejected.

4. **(10 Points)** A professional basketball player is averaging 21 points per game. He will be retiring at the end of this season. The team has multiple options to replace him. However, the owner feels that signing a replacement is only justified, if he can average more than 22 points per game. Which of the following are the appropriate hypotheses for this problem?

    a. H_0: $\mu \leq 21$ *vs*. H_a: $\mu > 21$
    b. H_0: $\mu \leq 22$ *vs*. H_a: $\mu > 22$
    c. H_0: $\mu \geq 21$ *vs*. H_a: $\mu < 21$
    d. H_0: $\mu \geq 22$ *vs*. H_a: $\mu < 22$

5. **(10 Points)** The _____ of two events *X* and *Y* is another event that consists of the sample space outcomes belonging to either event *X* or event *Y* or both event *X* and *Y*.

    a. Complement
    b. Union
    c. Intersection
    d. Conditional probability

6. **(10 Points)** A financial analyst working for a financial consulting company wishes to find evidence that the average price-to-earnings ratio in the consumer industry is higher than average price-to-earnings ratio in the banking industry. The null hypothesis is

   a. $\mu_{consumer} - \mu_{banking.} \leq 0$
   b. $\mu_{consumer} - \mu_{banking.} > 1$
   c. $\mu_{consumer} - \mu_{banking.} < 1$
   d. $\mu_{consumer} - \mu_{banking} \neq 0.$

7. **(10 Points)** In testing the difference between the means of two independent samples with unknown variances, the correct test statistic to use is:

   a. *Z* statistic.
   b. *t* statistic.
   c. *F* statistic.
   d. None of the above.

8. **(10 Points)** If the random variable *X* has a mean of μ and a standard deviation σ, then $(X-\mu)/\sigma$ has the following mean and standard deviation (in this order):

   a. μ and σ
   b. $\overline{X}$ and *s*
   c. 1 and 0
   d. 0 and 1

9. **(10 Points)** The degrees of freedom for a *t*-test about the differences between two population means when the variances are unknown and the sample sizes are 20 and 22 respectively are:

   a. 42
   b. 41
   c. 40
   d. 22

10. **(10 Points)** The advantage of the randomized block design over the completely randomized design is that we are comparing the treatments by using _____ experimental units.

    a. randomly selected
    b. the same
    c. different
    d. representative

11. (10 Points) Which is not an assumption of a multiple regression model?

- **a.** Positive autocorrelation of error terms
- **b.** Normality of error terms
- **c.** Independence of error terms
- **d.** Constant variation of error terms

12. (10 Points) In a one-way analysis of variance with three treatments, each with five measurements, in which a completely randomized design is used, what is the degrees of freedom for treatments?

- **a.** 5
- **b.** 2
- **c.** 3
- **d.** 15

13. (10 Points) If the sampled population has a mean 48 and standard deviation 16, then the mean and the standard deviation for the sampling distribution of $\bar{X}$ for $n = 4$ are:

- **a.** 4 and 1
- **b.** 12 and 4
- **c.** 48 and 4
- **d.** 48 and 8

14. (10 Points) Consider the one-way ANOVA table below.

Source	d.f.	Sum of Squares
Model	3	213.88125
Error	20	11.208333
Total	23	225.0895

What is the mean square error?

- **a.** 71.297
- **b.** .5604
- **c.** 1.297
- **d.** 213.8810

15. (10 Points) The reaction time in seconds to a stop light of a group of adult men were found to be

0.74, 0.71, 0.41, 0.82, 0.74, 0.85, 0.99, 0.71, 0.57, 0.85, 0.57, 0.55 (mean = .709)

What is the range?

- **a.** 0.190
- **b.** 0.520
- **c.** 0.580
- **d.** 0.1613

16. (10 Points) For the same set of observations on a specified dependent variable, two different independent variables were used to develop two simple linear regression models. The results are summarized as follows:

Model I	Model II
$r^2 = .92$	$r^2 = .85$
$s = 1.65$	$s = 1.91$

Based on the above results, we can conclude that:

- **a.** A prediction based on Model I is likely better than a prediction based on Model II.
- **b.** A prediction based on Model II is likely better than a prediction based on Model I.
- **c.** The SSE for Model II is smaller than the SSE for Model I.
- **d.** The total variation is different for Model I and Model II

17. (10 Points) The following results were obtained from a simple regression analysis:

$Y = 37.2895 - 1.2024 * X$
$r^2 = .6744$
$s^2 = .2934$

For each unit change in X (independent variable), the estimated change in the average value of Y (dependent variable) is equal to:

- **a.** –1.2024
- **b.** .6774
- **c.** 37.2895
- **d.** 1.2024

18. (10 Points) In order to test the significance of a single independent variable of a multiple regression model, we use:

- **a.** the partial *F* test
- **b.** *t* test
- **c.** the overall *F* test
- **d.** Durbin Watson test

19. (10 Points) The number of ways to arrange *x* successes among *n* trials is equal to

- **a.** $\frac{n!}{x!(n-x)!}$
- **b.** $\frac{n!}{(n-x)!}$
- **c.** n/x
- **d.** $n!/x!$

20. (10 Points) Which of the following is correct? The width of a confidence interval will be:

- **a.** Narrower for 99% confidence than for 95% confidence.
- **b.** Wider for a sample size of 100 than for a sample size of 50.
- **c.** Narrower for 90% confidence than for 95% confidence.
- **d.** Wider when the sample standard deviation (*s*) is small than when *s* is large.

21. (15 Points) Suppose that for events A, B, C, and E, we know that P(A) = .45, P(B) = .20, P(C) = .35, P(E|A) = .10, P(E|B) = .05, and P(E|C) = 0. Determine P(B|E).

22. (15 Points) An insurance company will insure a $20,000 car for its full value against theft at a premium of $500 per year. Suppose that the probability that the car will be stolen is 0.01, and let *X* denote the insurance company's profit. Keep in mind that the premium has to be paid no matter whether the car gets stolen or not! Calculate the insurance company's expected profit.

23. (15 Points) Suppose that *X* has a binomial distribution with $n = 500$ and $p = .6$. Make a continuity correction and then use the normal approximation to the binomial to find the probability $P(X = 305)$.

24. (15 Points) Download the data set http://www.math.usu.edu/~symanzik/teaching/2006_stat2300/data/CellUse.xls from the Web and use Excel to construct a 90% confidence interval for the average number of minutes of monthly cell phone usage of all USU students. You should assume that the numbers reported are based on a random sample of 100 USU students.

25. (15 Points) Download the data set http://www.math.usu.edu/~symanzik/teaching/2006_stat2300/data/WebVisit.xls from the Web and use Excel to calculate the (i) mean, (ii) median, and (iii) population variance for the number of visitors (in millions) of the top 10 Websites in April 2004, i.e., these data represent the entire population (which is the top 10 Websites).

Use the following to answer questions 26-29: The manager of a local specialty store is concerned with a possible slowdown in payments by her customers. She measures the rate of payment in terms of the average number of days receivables are outstanding. Generally, the store has maintained an average of 50 days with a standard deviation of 10 days. A random sample of 25 accounts gives an average of 54 days outstanding with a standard deviation of 8 days.

26. (15 Points) Set up the null and alternative hypotheses needed to show that there has been a slowdown in payments by the company's customers.

27. (15 Points) Calculate the appropriate test statistic to test the hypotheses.

28. (15 Points) What is the rejection point for testing these hypotheses at $\alpha = .01$ and on which distribution is this based? Indicate the degrees of freedom if appropriate.

29. (15 Points) Indicate a 99% confidence interval for the population mean time (in days) all customer accounts for this particular store are outstanding.

What is the effect of the number of households (x – in 1,000) in a city on the total electronics sales (y – in \$1,000) for that city for an electronics store chain? To answer this question, the chairman of this store chain undertakes a study to relate the number of households to sales. To do this, sales records of 15 stores of this chain and the corresponding number of households are randomly selected. Download the data set http://www.math.usu.edu/~symanzik/teaching/2006_stat2300/data/Electronics.xls from the Web and use Excel to answer questions 30-35.

30. (15 Points) Obtain the least squares point estimates for the intercept (b_0) and the slope (b_1).

31. (15 Points) Assuming that the store chains wants to open a new store in a city that has 200,000 households, use the least squares line to obtain a point prediction of the total expected electronics sales (in \$1,000) for this city.

32. (15 Points) Indicate the residual sum of squares (SSE) and use this to calculate s^2 and s.

33. (15 Points) Identify s_{b_1} and the t statistic for testing the significance of the slope. Show how t has been calculated by using b_1 and s_{b_1}.

34. (15 Points) Identify the p-value for testing H_0: $\beta_1 = 0$ versus H_a: $\beta_1 \neq 0$. Using the p-value, determine whether we can reject H_0 by setting α equal to .10, .05, .01, and .001. What do you conclude about the relationship between GPA and starting salary?

35. (15 Points) Indicate the 90% confidence interval for β_1.

A consumer preference study involving three different bottle designs (A, B, and C) for the jumbo size of a new liquid laundry detergent was carried out using a randomized block experimental design, with supermarkets as blocks. Specifically, four supermarkets were supplied with all three bottle designs, which were priced the same and carried the same amount of detergent. Data collected consists of the number of bottles of each design sold in a 24-hour period at each supermarket. Download the data set http://www.math.usu.edu/~symanzik/teaching/2006_stat2300/data/BottleDes2.xls from the Web and use Excel to answer questions 36-40.

36. (15 Points) For supermarkets 1, 2, 3, and 4, indicate the average number (per design) of laundry detergent bottles sold.

37. (15 Points) Test the null hypothesis H_0 that no differences exist between the effects of the bottle designs on mean daily sales. Set $\alpha = .05$. Report the corresponding (i) F statistic, (ii) degrees of freedom, and (iii) p-value for this test, (iv) state whether we reject the null hypothesis at this significance level, and (v) clearly conclude whether the different bottle designs have / do not have different effects on mean sales.

38. (15 Points) Test the null hypothesis H_0 that no differences exist between the effects of the supermarkets on mean daily sales. Set $\alpha = .05$. Report the corresponding (i) F statistic, (ii) degrees of freedom, and (iii) p-value for this test, (iv) state whether we reject the null hypothesis at this significance level, and (v) clearly conclude whether the different supermarkets have / do not have different effects on mean sales.

39. (15 Points) To calculate Tukey simultaneous 95% confidence intervals to make pairwise comparisons of the bottle design effects on mean daily sales, we need the value q_α from Table A.9 in the appendix of the textbook. Specify which numbers we have to use for (i) r and (ii) v, (iii) whether we have to look up $q_{.10}$, $q_{.05}$, or $q_{.01}$, and (iv) indicate the resulting value.

40. (15 Points) Use Tukey simultaneous 95% confidence intervals to make pairwise comparisons of the bottle design effects on mean daily sales. If you could not determine the appropriate value of q_α in question 39, indicate so and use 4.0 instead in this question. Which bottle design(s) maximize mean sales?