STATISTICS 151 SAMPLE FINAL EXAMINATION

Instructions:

A. This is a closed book examination. You are permitted to use a non-programmable calculator. No devices with infrared ports or any other communication options are permitted. Calculators brought into the examination room are subject to inspection and, in case of doubt, may be taken away for inspection.

B. This sample exam is to assist you in reviewing the concepts covered in Stat 151. However, the actual exam may cover also other topics covered in the course. This exam consists of two parts.

In the first part there are 25 multiple-choice questions. For each question exactly one answer is correct. Circle the correct answer to each multiple-choice question. In this part all answers will be graded right or wrong (no partial credit).

In the second part there are two long-answer problems. Show all your work because partial credit may be given.

C. You will need the tables of the standard normal distribution, critical values of the t distributions and F distributions to answer some of the questions.Calculations may be made on the reverse side of the preceding page.

PART 1 (MULTIPLE-CHOICE QUESTIONS, 1 POINT EACH)

1. A market research company wishes to find out whether the population of students at a university prefers brand A or brand B of instant coffee. A random sample of students is selected, and each one is asked to try brand A first and then brand B (or vice versa, with the order determined at random. They then indicate which brand they prefer. This is an example of
(a) an observational study (b) a randomized experiment **(c) a matched pairs design**
(d) a double-blind study (e) none of these

2. In a dice game, the player independently rolls a fair red die and a fair green die. The player wins if and only if the red die shows a 1,2, or 3, or if the total on the 2 dice is 11. What is the probability that the player will win?

(a) 7/36 (b) 4/9 (c) 19/36 **(d) 5/9** (e) 29/36

3. A city emergency station uses two ambulances. The probability that a particular ambulance will be available in case of emergency is 0.99. Assuming that the availability of any of the two ambulances is independent of the other, what is the probability that at least one ambulance will be available in emergency?

(a) 0.0001 (b) 0.9801 (c) 0.9900 (d) 0.9990 **(e) 0.9999**

4. A soft-drink machine is set so the amount of drink dispensed is a random variable having a normal distribution with the mean of 16 ounces and the standard deviation of 2 ounces. What is the fraction of bottles with the weight exceeding 18 ounces?

(a) 0.1587 (b) 0.1789 (c) 0.3174 (d) 0.3865 (e) 0.8413

5. Refer to the previous problem. What is the probability that the total weight of 4 randomly selected bottles is less than 68 ounces?

(a) 0.1345 (b) 0.1915 (c) 0.4772 (d) 0.6862 **(e) 0.8413**

6. A manufacturer has observed that the time that elapses between the placement of an order with a just-in-time supplier and the delivery of the parts is uniformly distributed between 100 and 180 minutes. What proportion of orders takes

between 2 and 2.5 hours to be delivered?

(a) 0.125 (b) 0.250 (c) 0.325 **(d) 0.375** (d) 0.425

7. Suppose you have generated 100 random samples of a given size n =20 from a standard normal population. For each sample you test H_o: $\mu = 0$ against H_A: $\mu > 0$ at the 0.05 level of significance. On the average how many times out of 100 would you expect to make the correct decision (that is not to reject H_0)?

(a) 5 (b) 20 (c) 80 (d) 85 **(d) 95**

8. Suppose you have generated 1,000 random samples of a given size n=100 from a standard normal population. For each sample you test H_o: $\mu = 0$ against H_a: $\mu > 0$. We reject H_o if the sample mean exceeds 0.212. On the average how many times out of 1000 would you expect to make the wrong decision about H_o (that is rejecting H_o):

(a) 17 (b) 19 (c) 21 (d) 23 (e) 25

9. In the drought year 1988, statements were made that over half of Indiana corn producers did not get back from their corn crop the money they put into seed, fertilizer, etc. To check this a random sample of 800 farms is chosen and a brief audit is made on each of these farms. Of these farms, 405 did not recover their costs from their corn crops. Is this good evidence for the claim in the first sentence? Formulate the null and alternative hypotheses with p- the true proportion of farms that did not recover their costs.

(a) H_0: p =405/800 vs. H_A: p > 405/800 (b) H_0: p = 395/800 vs. H_A: p < 395/800 **(c) H_0: p = 0.5 vs. H_A: p > 0.5**
(d) H_0: p = 0.5 vs. H_A: p < 0.5 (e) H_0: p = 405/800 vs. H_A: p < 405/800

10. Refer to Question 9. What is the value of the appropriate test statistic for the problem?

(a) 0.3537 (b) 1.845 (c) 2.148 (d) 2.645 (e) 3.124

11. Using the same set of data, you compute a 95% confidence interval and a 99% confidence interval for a population mean. Which of the following statements is correct?

(a) the intervals have the same width
(b) the 99% confidence interval is wider
(c) the 95% confidence interval is wider
(d) the 99% confidence interval is wider only if the sample size is sufficiently large
(e) you cannot determine which interval is wider unless you know the sample size n and s

12. In order to test H_0: $\mu = 0$ vs. H_a: $\mu > 0$ a simple random sample of n=25 observations from a normal population with standard deviation 1 was obtained. The value of the test statistic is 2. What is the p-value of the test?

(a) 0.0114 **(b) 0.0228** (c) 2 (d) (e) 0.9772

13. We test H_0: $\mu = 0$ vs. H_a: $\mu > 0$ for a standard normal distribution using a random sample of n=100.Suppose the sample mean is 0.1. What is the value of the appropriate test statistic?

(a) 1 (b) 1.5 (c) 2 (d) 2.5 (e) 3

14. An agricultural researcher plants twenty-five plots with a new variety of corn. The average yield for these plots is 150 bushels per acre. Assume that the yield per acre for the new variety of corn follows a normal distribution with unknown mean μ and standard deviation of 10 bushels per acre. The following confidence interval for the mean yield μ was obtained: 150±4.66. The confidence level of the confidence interval is

(a) 90% (b) 95% (c) 97% **(d) 98%** (e) 99%

15. A machine produces metal rods used in an automobile suspension system. A random sample of twenty two rods is selected and the diameter is measured for each of them. The sample mean is 13.71 and the sample standard deviation is 3.55. The 95% confidence interval for the mean diameter is

(a) 13.71±1.48 **(b) 13.71±1.57** (c) 13.71±1.66 (d) 13.71±1.78 (e) 13.71±1.84

16. Refer to the previous question. We test H_o: μ = 15 against H_1: μ < 15, where μ stands for the mean diameter of rods. The p-value of the test is between

(a) 0.005 and 0.01 (b) 0.01 and 0.025 (c) 0.025 and 0.05 **(d) 0.05 and 0.10** (d) 0.10 and 0.25

17. Random samples of size n=9 are taken repeatedly from a normal population with the mean of 1 and standard deviation of 0.6 and sample means are calculated for each sample. The sampling distribution of the sample mean is

(a) normal with the the mean of 1 and the standard deviation of 0.6
(b) normal with the the mean of 1 and the standard deviation of 1/ 0.6
(c) normal with the the mean of 1 and the standard deviation of 0.2
(d) non-normal with the the mean of 0 and the standard deviation of 1
(e) none of these

18. An insurance company is attempting to see if two different chains of auto-body repair shops give significantly different estimates of repair costs. They take five randomly selected cars to one chain, and then the cars to another chain, and obtain estimates. These estimates, in hundreds of dollars, are:

CAR	1	2	3	4	5
CHAIN 1	1	3	5	6	7
CHAIN 2	2	4	6	7	6

What is the absolute value of the appropriate test statistic?

(a) 1 **(b) 1.5** (c) 2 (d) 2 (e) 2.5

19. Refer to the value obtained in Problem 18. What are the values bracketing the P-value of the test?

(a) 0.05 and 0.1 (b) 0.10 and 0.20 (c) 0.10 and 0.30 (d) 0.10 and 0.15 **(e) larger than 0.20**

20. Sarah's parents are concerned that she seems short for her age. The following linear regression model $\mu_{HEIGHT} = \alpha + \beta \cdot AGE$, was used to study the relationship between AGE (in months) and HEIGHT (in cm) for Sarah. Here is the related computer output for the variable AGE between ages 4 (48 months) and 5 (60 months).

Predictor	Coef	SE Coef	T	P-Value
Constant	69.200	1.806	38.33	0.000
AGE	0.4333	0.03333	13.00	0.001

s = 0.3162 R-Sq = 98.3%

What is the approximate value of the correlation between AGE and HEIGHT for Sarah?

(a) 0.912 (b) 0.924 (c) 0.946 (d) 0.983 **(d) 0.991**

21. Refer to Question 20. What would be the predicted Sarah's height at 60 months?

(a) 89.333 **(b) 95.198** (c) 97.327 (d) 98.211 (e) 99.536

22. Refer to Question 20. Normally growing girls gain 6 cm on the average between ages 4 and 5. In order to see whether Sarah's growth is less rapid than normal, the null and alternative hypotheses should be defined as follows:

(a) H_0: ß= 6, vs. H_A: ß> 6,
(b) H_0: ß= 6, vs. H_A: ß< 6,
(c) H_0: ß= 0.4333, vs. H_A: ß< 0.433,
(d) H_0: ß= 0.4333, vs. H_A: ß >0.433,
(e) H_0: ß= 0.5, vs. H_A: ß< 0.5.

23. Refer to Question 29. Use the output in Question 20 to calculate the value of the test statistic for the test in Question 22. The value of the test statistic is

(a) -2.00 (b) -1.00 (c) 1.27 (d) 2.79 (e) 13.00

24. In order to compare the effectiveness of three new drugs A, B, and C to reduce blood pressure of patients with hypertension, a random sample of 30 people suffering from the condition is obtained and each subject in the sample is assigned to one of the three treatment groups (A, or B, or C), each group consisting of 10 subjects. The blood pressure reductions for each subject are in three treatment groups are then obtained. What is the distribution of the test statistic to test the null hypothesis of no difference in the mean systolic blood pressure reductions among the three groups?

(a) Normal with mean 10 and standard deviation 30,
(b) t-distribution with 30 degrees of freedom,
(c) F distribution with 3 degrees of freedom for the numerator and 30 degrees of freedom for denominator,
(d) F distribution with 3 degrees of freedom for the numerator and 27 degrees of freedom for denominator,
(e) F distribution with 2 degrees of freedom for the numerator and 27 degrees of freedom for denominator.

25. Refer to Question 24. In order to reject the null hypothesis of no difference among the three means at the level 0.05, the value of the test statistic has to be

(a) smaller than 2.96 (b) larger than 2.96 (c) larger than 3.32 (d) larger than 3.34 **(e) larger than 3.35**

Use the appropriate table to answer the question.

PART 2 (LONG-ANSWER PROBLEMS)

1. The diameter (in cm) of steel rods manufactured on two different extrusion machines is being investigated. Two random samples of 25 and 21 observations are selected. The corresponding sample means for machines 1 and 2 are 9.62 and 10.10. The corresponding sample variances for machines 1 and 2 are 1.00 and 0.90, respectively.

(a) Manufacturing specifications indicate that the mean diameter of steel rods must be at least 10cm. Is there significant evidence at $\alpha = 0.05$ that machine 1 is not meeting specifications? Carry out an appropriate test to support your answer (6 points)

Denote by μ = the mean diameter of steel rods for machine 1. Define the null and alternative hypotheses as follows H_0: $\mu \geq 10$ vs. H_A: $\mu < 10$. The value of the test statistic is

$$t = \frac{\overline{x} - \mu_0}{s/\sqrt{n}} = \frac{9.62 - 10}{\sqrt{1}/\sqrt{25}} = -1.9$$

The t statistic follows a t-distribution with DF= n-1= 25-1=24. Given that the alternative hypothesis is H_A: $\mu<10$, the p-value is the area under the t(24) density curve to the left of -1.9. Based on the table of the t-distribution critical values with DF=24, 0.025<p-value<0.05. Thus there is an evidence that machine 1 is not meeting specifications.

(b) Assume that $\sigma_1 = \sigma_2$ and that the data are drawn from normal populations. Construct a 95% confidence interval for the difference in mean rod diameter between machine 1 and machine 2. Does the interval provide evidence that the two machines produce rods with different mean diameters? (4 points)

$$(\bar{x}_1 - \bar{x}_2) \pm t^* \bullet s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}, \quad \text{DF}= n_1 + n_2 - 2 = 25+21-2 = 44$$

As the critical value of t for DF=44 is not available in the tables, we will take the critical value for DF=40 as this will make the confidence interval more conservative (wider, because the larger t for DF=40 relative to DF=44).
The estimate of the common variance is

$$s^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2} = \frac{(25-1)\cdot 1 + (21-1)\cdot 0.9}{25+21-2} = 0.9545.$$

Thus the 95% confidence interval is

$$(\bar{x}_1 - \bar{x}_2) \pm t^* \bullet s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}} = (9.62-10.10) \pm 2.021 \bullet 0.977\sqrt{\frac{1}{25}+\frac{1}{21}}$$

2. A new blood pressure drug is advertised to reduce, after 1 week of medication, a patient's blood pressure an average of 10 units. Blood pressure reductions were recorded for 16 patients after treatment with the drug. The mean and standard deviation for this sample were 9.0 and 2.0, respectively.

(a) Do the data appear to contradict the advertising claim? Test at the 0.05 level of significance (6 marks). Find the values bracketing the P-value. (2 points)

Denote by μ = the mean reduction in blood pressure. Define the null and alternative hypotheses as follows H_0: $\mu = 10$ vs. H_A: $\mu<10$ (the drug is not that effective in reducing blood pressure as advertised). The value of the test statistic is

$$t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}} = \frac{9-10}{2/\sqrt{16}} = -2$$

The t statistic follows a t-distribution with DF= n-1= 16-1=15. Based on the table of the t-distribution critical values, 0.025<p-value<0.05. The null hypothesis is rejected. Thus there is strong evidence that the new drug does not reduce blood pressure as advertised.

(b) What assumption should be imposed on the distribution of blood pressure reductions in order to make the test valid? What would you do to make the test valid if the assumption is not satisfied? (2 points)

We can assume that the data come from a normal population to make the test valid. However, given the sample size of n=16>15 and the robustness of the t-procedures for non-normal data, the minimum assumption is that the data do not show any outliers or strong skewness.

(c) Estimate with 95% confidence the average blood pressure reduction for the drug (3 points)

$$\bar{x} \pm t^{*} \cdot \frac{s}{\sqrt{n}} = 9 \pm 2.131 \cdot \frac{2}{\sqrt{16}}$$

(d) Will a 99% confidence interval be wider or narrower than the 95% confidence interval found in the previous problem? Note that actually you don't have to calculate the intervals to answer the questions (2 points)

The 99% confidence interval will be wider.