**STATISTICS 2023** **NAME IN PRINT** Key

**FINAL EXAM** **SIGNATURE IN INK** _____

**FALL 2009** **CWID IN INK** _____

**FILL IN THE BLANK. Write the word or phrase from the list on the right that belongs in the blank.**
**(Each blank 2 points each)**

THE WORDS OR PHRASES IN THIS LIST MAY BE USED MORE THAN ONCE OR NOT USED AT ALL.

| |
|---|
| Standard error |
| Test statistic |
| Null hypothesis |
| Alternative hypothesis |
| Parameter |
| Sample statistic |
| Confidence interval |
| Point estimate |
| Bound of error |
| Population |
| Sample |
| Statistical inference |
| Alpha |
| P-value |
| Rejection region |
| Critical value |

1. The conclusion of a hypothesis test indicates that the data do or do not provide support for the alternative hypothesis.

2 In the field of statistics a population is described by observing a subset of the population known as a sample.

3. A population parameter is an unknown constant that describes the population.

4. A point estimate is a single number calculated from a sample to estimate a population parameter.

5. A confidence interval is a way to construct an interval estimate so that there is a certain degree of accuracy associated with the estimator.

6. The measure of relative standing of the point estimate in a hypothesis test is the test statistic.

7. A test statistic is used to form a decision about the validity of the null hypothesis.

8. The area of the rejection region is equal to alpha.

9. The error rate that must be tolerated if the decision in a hypothesis test is to reject the null hypothesis is called the p-value of a test.

10. Statistical inference uses information from samples to form conclusions about population parameters.

11. The rejection region divides the distribution of the test statistic under a true null hypothesis into likely and unlikely values.

**TRUE OR FALSE. Answer with a capital T or F.** **(3 points each)**

T 12. If the slope of the regression line is not shown to be different from zero then the x-variable in the regression equation does not significantly influence the y-variable.

T 13. If the value of the estimated linear correlation, r, is close to positive one then a researcher could conclude that the bivariate data lies close to an upward sloping line.

F 14. If a hypothesis test is performed on the mean of two populations whose variances are assumed equal then the estimated standard error of the point estimate for the difference between the means would not be based on a pooled variance estimator, but would be based on separate variance estimators from the two independent samples.

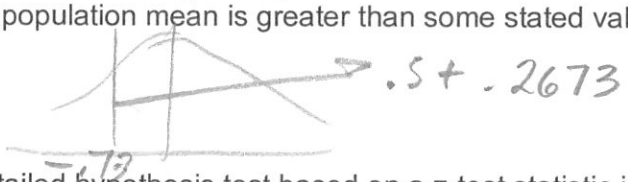
**STATE THE ANSWER. Write the answer on the line.** **(3 points each)**

55 15. What is the numerical value of the mean of a sample of six observations, 61, 84, 42, 51, 60, and 32?

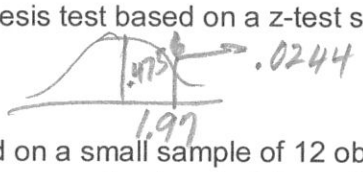
17.98 16. What is the numerical value of the standard deviation of a sample of six observations, 61, 84, 42, 51, 60, and 32? Round your answer to two digits past the decimal.

17.977764

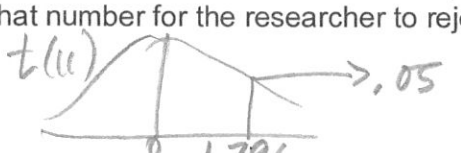
.7673 17. If a Z hypothesis test based on a large sample has a test statistic value of -.73 and the researcher is trying to prove that the population mean is greater than some stated value, what is the p-value of the hypothesis test?




1.97 18. If the p-value in a two tailed hypothesis test based on a z-test statistic is equal to 0.0488 what is the magnitude of the test statistic?




1.796 19. In a right-tail hypothesis test based on a small sample of 12 observations the value of the test statistic must exceed what number for the researcher to reject the null hypothesis with only a 0.05 error rate?




$M_2 - M_1 > 12$

OR $M_1 - M_2 < -12$ 20. If a researcher who was comparing the means of two populations was attempting to prove that the mean of population two is more than 12 units larger than the mean of population one what is the appropriate alternative hypothesis?

$M_2 - M_1 > 12$ OR $M_1 - M_2 < -12$

-42.5 21. If the mean of the sample from population one is 42.3 and the mean of the sample from population two is 84.8 what is the numerical value of the point estimate for the mean of population one minus the mean of population two?

42.3 - 84.3

4.995 22. If a sample of 23 observations had a sample variance of 4.5 and a sample of 19 observations had a sample variance of 5.6 then what is the numerical value of the pooled variance estimate based on these two samples? State three digits past the decimal.

$$\frac{22(4.5) + 18(5.6)}{40} = 4.995$$

**STATE THE ANSWER. Write the answer on the line. (3 points each)**

A loan officer compares the interest rates for 48-month fixed-rate auto loans and 48-month variable-rate auto loans. Two independent, random samples of auto loan rates are selected. A sample of eight 48-month fixed-rate auto loans had the following loan rates. Do not assume equal population variances for the questions on this page.

12.29% 11.75% 9.50% 8.99% 9.25% 10.99% 9.40% 11.55%

while a sample of five 48-month variable-rate auto loans had loan rates as follows:

8.59% 9.75% 7.99% 9.50% 8.29%

**Hypothesis Test:** **Independent Groups**

**(t-test, pooled variance)**

Not assuming equal population variances

| Fixed | Variable | |
|---|---|---|
| | | Mean |
| | | Std. dev. |
| | | n |

$\bar{X}_1 = 10.465$, $S_1 = 1.31757$

$\bar{X}_2 = 8.824$, $S_2 = .766472$

| | |
|---|---|
| | df |
| | difference (Fixed - Variable) |
| | pooled variance |
| 0.57836 | standard error of difference |
| 0 | hypothesized difference |
| 2.8 | t |
| 0.0032 | p-value (two-tailed) |

10.465 23. State the mean of the sample of fixed rate loan percentages. State your answer with three digits past the decimal.

.766 24. State the estimate for the standard deviation for the population of variable rate loan percentages. State your answer with three digits past the decimal.

1.641 .766472 25. State the estimate of the difference between the mean of the population of fixed rate loan percentages and the mean of the population of variable rate loan percentages. State your answer with three digits past the decimal.

$\mu_1 - \mu_2 > 0$ $10.465 - 8.824 =$ 26. What is the alternative hypothesis if the question is whether the data supports the conclusion that the mean of the population of loan percentages for fixed rate loans is greater than the mean of the population of loan percentages for variable rate loans?

2.837 $\mu_F > \mu_V$, $\mu_1 > \mu_2$ or $\mu_1 - \mu_2 > 0$ 27. What is the value of the test statistic to test whether the mean of the population of fixed rate loan percentages is equal to the mean of the population of variable rate loan percentages? State your answer with three digits past the decimal.

1.108 2.83735 28. What is the value of the test statistic to test whether the mean of the population of fixed rate loan percentages exceeds the mean of the population of variable rate loan percentages by one percent? State your answer with three digits past the decimal.

1.27 1.10831 29. What is the value of the bound of error test statistic that would be used to construct a 95% confidence interval to estimate the difference between the mean of the population of fixed rate loan percentages and the mean of the population of variable rate loan percentages? Use df=11.

$t_{.025(11)} \cdot .57836$

$2.201 (.57836) = 1.27297$

**LINEAR REGRESSION QUESTIONS. Write the answer on the line. (3 point each)**

The average hourly temperature affects the weekly fuel consumption of natural gas in a small city. The following data are the average hourly temperature measured in Fahrenheit degrees (X) and the weekly fuel consumption of natural gas measured in millions of cubic feet (Y) for eight randomly chosen weeks for a small city in the mid-west United States. Use this data to answer the following questions.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| X | 28.0 | 28.0 | 32.5 | 39.0 | 45.9 | 57.8 | 58.1 | 62.5 |
| Y | 12.4 | 11.7 | 12.4 | 10.8 | 9.4 | 9.5 | 8.0 | 7.5 |

3413.11 30. What is the sum of the cross product for the daily average temperature and fuel consumption data provided above? $\Sigma XY = X_1Y_1 + \ldots + X_n Y_n = 3413.11$

25.54875 31. What is the numerical value of the corrected sum of squares for the y-variable based on the daily average temperature and fuel consumption data provided above? $SS_Y = \Sigma Y^2 - \frac{(\Sigma Y)^2}{n} =$

−.1 32. What is the least squares estimate of the slope in the linear regression equation to estimate the weekly fuel consumption based on the average daily temperature? Round your answer to one digit past the decimal.

15.8 33. What is the least squares estimate of the y-intercept in the linear regression equation to estimate the weekly fuel consumption based on the average daily temperature? Round your answer to one digit past the decimal.

$\hat{y} = 15.8 - .1X$ 34. Write the estimated regression equation to estimate the class grade from the number of classes missed. Use the estimates of the slope and y-intercept that you calculated in the two problems directly above.

11 35. What is the least squares estimate of the weekly fuel consumption based on an average daily temperature of 48F?

−.948 36. What is the numerical value of the estimated linear correlation between the two variables? Round your answer to three digits past the decimal.

89.95 37. What percent of the variance in weekly fuel consumption is determined by the variance in average daily temperature?

.89948 .8995