STATISTICS 2023 **NAME IN PRINT** 2E4S97V1. WRD

FINAL EXAM **SIGNATURE**____________________

SPRING 1997 **SS OR OSU ID**____________________

TRUE OR FALSE. Answer with a capital T or F. **(2 points each)**

T 1. The center value in a confidence interval is the point estimate for the parameter being estimated.

T 2. If a null hypothesis is a true statement then the set of values for the test statistic is known; if the observed test statistic is an unlikely value in that set then the null hypothesis is rejected.

F 3. In a hypothesis test on a population mean based on a large sample if the test statistic value is 1.0, then the null hypothesis could be rejected with a small error rate.

F 4. Parameters are known constant values calculated from the sample data.

F 5. When a researcher does not reject the null hypothesis the error rate that must be tolerated is the p-value of the hypothesis test.

F 6. The width of a confidence interval to estimate a population parameter depends on the value of the point estimate for the parameter and the standard error of the point estimate.

F 7. Part of the time when a point estimate is used to estimate a population parameter it is correct and has a standard error of zero.

T 8. If a small sample is used to build a confidence interval to estimate a population mean or to test a hypothesis about the mean then we must assume that the population from which the sample was drawn is normally distributed.

T 9. The Z-distribution and the t-distribution are similar in the sense that they are both symmetric with a mean of zero, but for low degree of freedom the t-distribution has greater variance.

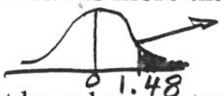
F 10. Hypotheses are statements about values of point estimates.

F 11. Variance estimators with low degree of freedom are preferable to variance estimators with high degree of freedom.

F 12. The mean of a sample must always be a larger number than the standard deviation of a sample.

STATISTICS 2023 FINAL EXAM SPRING 1997 PAGE 2
FILL IN THE BLANK. (2 points each)

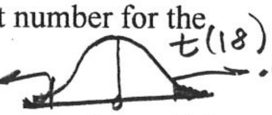
0.0694 13. If a hypothesis test based on a large sample has a test statistic value of 1.48 and the researcher is trying to prove that the population mean is more than some stated value what is the p-value of the hypothesis test?



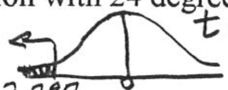
$P = .5 - .4306 = .0694$

1.72 14. If the p-value in a right-tail hypothesis test based on a z test statistic is equal to 0.0427 what is the numerical value of the test statistic?

3.922 15. In a two-tail hypothesis test based on a small sample of only 19 observations the absolute value of the test statistic must exceed what number for the researcher to reject the null hypothesis with only a 0.001 error rate?



0.005 16. What maximum error rate is the researcher willing to tolerate if the rejection rule in a left-tail hypothesis test based on a t-distribution with 24 degree of freedom is, "Reject Ho if the test statistic is less than -2.797?"



0.02 17. If a 95% confidence interval to estimate a population proportion is (0.3608, 0.4392) what is the standard error of the point estimate for the population proportion? width $= 2B = .0784 = 2(1.96\, S_{\hat{p}}) \Rightarrow S_{\hat{p}} = \frac{.0784}{2(1.96)} = .02$

189 18. How many observations are required if a biologist wants to estimate the average number of days required for a specific virus to develop correct to within 3 days with a 95% confidence when it is known that the standard deviation for the virus development time is 21 days? $n \geq \frac{z^2\sigma^2}{B^2} = \frac{1.96^2(21)^2}{3^2} = 188.24 \Rightarrow 189$

64 19. If a 90% confidence interval to estimate a population mean is (58, 70), what is the numerical value of the point estimate for the population mean?

$\bar{X}$ is the center of the interval.

14 20. If a sample of 200 observations produces a sum of 2800 units what is the numerical value of the mean of the sample? $\bar{X} = \frac{\Sigma X}{n} = \frac{2800}{200} = 14$

259 21. If the sum of the data from a sample with 100 observations is 1,420 and the uncorrected sum of squares is 45,805 then what is the value of the variance estimate from this sample? $S^2 = \frac{\Sigma X^2 - \frac{(\Sigma X)^2}{n}}{n-1} = \frac{45{,}805 - \frac{1420^2}{100}}{99} = 259$

(228.34, 261.66) 22. If a sample of 400 observations produces a mean of 245 units and a sample standard deviation of 170 units what is the 95% confidence interval from this sample to estimate the population mean? Round your answer to 2 digits past the decimal. $\bar{X} \pm z \cdot S_{\bar{X}} \Rightarrow 245 \pm 1.96\left(\frac{170}{\sqrt{400}}\right) \Rightarrow 245 \pm 16.66$

4.2 23. A sample of 6 observations is the following: 4, 9, 11, 17, 12, 11, what is the numerical value of the estimate of the standard deviation for the population from which the sample was drawn? Round to 1 digit past the decimal.

$S = 4.22689996$

COMPARING TWO POPULATIONS **(3 points each)**

A financial consultant wants to study the difference between the average returns on $1,000 investments for two different types of stock. One hundred investments of each type of stock were made. The data below indicates the mean and variance of the return on each type of stock for a $1,000 investment held for only 3 months.

	Stock Type One	Stock Type Two
Mean	$1032.50	$1022.90
Variance	$ 45.80	$ 98.20

9.6 24. What is the numerical value of the point estimate of the difference between the average returns on the $1,000 investment for the two different types of stock?

$$\bar{X}_1 - \bar{X}_2 = 1032.50 - 1022.90 = 9.60$$

1.2 25. What is the estimate of the standard error for the point estimate for the difference between the average returns on the $1,000 investment for the two different types of stock based on these two independent samples?

$$S_{\bar{X}_1-\bar{X}_2} = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} = \sqrt{\frac{45.80}{100} + \frac{98.20}{100}} = \sqrt{1.44} = 1.2$$

$M_1 - M_2 \neq 0$, $M_1 \neq M_2$ 26. State the alternative hypothesis if the financial consultant wishes to answer the question, "Do these two samples provide evidence that the average returns for these two types of stock are not equal for a $1,000 investment held only for 3 months?"

$$H_A: M_1 \neq M_2 \quad \text{OR} \quad H_A: M_1 - M_2 \neq 0$$

4 27. If the estimated standard error for the difference between the sample means is 2.4 what is the value of the test statistic to check the hypothesis of equal average returns for these two types of stock? If $S_{\bar{X}_1-\bar{X}_2} = 2.4$ then,

$$t = \frac{\bar{X}_1 - \bar{X}_2}{S_{\bar{X}_1-\bar{X}_2}} = \frac{9.6}{2.4} = 4.$$

72 28. A researcher would probably not use a pooled variance estimator in this case since there are 100 observations for each sample, but if the pooled variance estimator was used what would the numerical value be based on these two samples?

$$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2} = 72$$

YES 29. If the P-value of this hypothesis test is 0.007 and the significance level is 2% should the hypothesis of equal average returns be rejected? Answer YES or NO.

$$P = .007 < .02 = \alpha \Rightarrow \text{Reject } H_0.$$

NO 30. If the bound of error on a 95% confidence interval to estimate the difference in average returns for the two types of stock is 4.8 would the confidence interval to estimate the average difference in returns for the two types of stock contain zero or not? Answer YES to indicate the interval would contain zero; NO to indicate the interval would not contain zero.

LINEAR REGRESSION QUESTION **(2 points each blank)**

31. A set of six bivariate data points were recorded as follows:

x	4	2	8	6	7	3
y	10	8	16	14	16	10

Use the six bivariate data points above to calculate the sufficient sums:

$\Sigma x =$ 30, $\Sigma y =$ 74, $\Sigma x^2 =$ 178, $\Sigma y^2 =$ 972, $\Sigma xy =$ 410.

Use the six bivariate data points above to calculate the corrected sums of squares:

$SS_x =$ 28, $SS_y =$ $59.\overline{3}$ ✓, $SS_{xy} =$ 40.

The least squares estimate of the slope for this regression situation is: 1.428571429 ✓

✓ round to 1.

ANOTHER LINEAR REGRESSION QUESTION **(3 points each)**

The following questions are not related to the above.

Twenty bivariate data points were used to estimate a linear regression equation to estimate TOTAL COST (y) based on LABOR COST (x). The resulting corrected sums of squares for the data set are $\mathbf{SS_x = 170, SS_y = 1{,}320, SS_{xy} = 425}$. Use these corrected sums of squares to answer the remaining questions on this page.

2.5 32. What is the estimate of the slope of the regression line that fits these data the best in terms of squared vertical distance between data and line?

$$\hat{\beta}_1 = \frac{SS_{xy}}{SS_x} = \frac{425}{170} = 2.5$$

27.8 33. If $\Sigma x = 680$, $\Sigma y = 2{,}460$ and the estimate of the slope is 2.8 what is the estimate of the y-intercept for the regression line that best fits the bivariate data?

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = \frac{\Sigma y}{n} - \hat{\beta}_1 \frac{\Sigma x}{n} = \frac{2{,}460}{20} - 2.8\left(\frac{680}{20}\right) = 27.8$$

7.22 34. If the estimate of the slope is 2.8 what is the estimate of the variance of the residuals for this regression model? Round your answer to two digits past the decimal.

$$S_\varepsilon^2 = \frac{SS_y - \hat{\beta}_1 SS_{xy}}{n-2} = \frac{1{,}320 - 2.8(425)}{20-2} = 7.\overline{2}$$

1.4 35. If the estimate of the slope is 2.8 and the estimate of the standard error for the slope estimate is 2, what is the t test statistic value to test whether the slope is zero in this regression model?

$$t = \frac{\hat{\beta}_1}{S_{\hat{\beta}_1}} = \frac{2.8}{2.0} = 1.4$$

.897 36. What is the numerical value of the estimated linear correlation coefficient that results from this regression situation? Round your answer to three digits past the decimal.

$$r = \frac{SS_{xy}}{\sqrt{SS_x \cdot SS_y}} = \frac{425}{\sqrt{170 \cdot 1{,}320}} = .897175703$$