STATISTICS 2023 NAME IN PRINT Key

FINAL EXAM SIGNATURE IN INK ____________

FALL 2014 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 the box may be used more than once, or not 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

1. A point estimate is a single number calculated from observed data that is used to estimate a population parameter.

2. The p-value is the observed risk of error when the null hypothesis is rejected and the conclusion is that the data supports the alternative hypothesis.

3. The standard error associated with a statistical point estimate is the typical mistake made by the point estimate when it is used to estimate a population parameter.

4. The bound of error is half the width of the confidence interval.

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 rejection region of a hypothesis test identifies the unlikely values of the test statistic for which the null hypothesis will be rejected at the alpha significance level.

7. The statement stated as true that is tested by a hypothesis test is the null hypothesis.

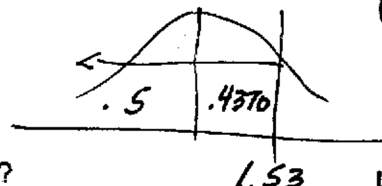
8. The process of Statistical inference involves forming conclusions about population parameters based on observed data.

9. If the sample size is small, then the standard error for the point estimator increases in magnitude.

10. In a hypothesis test, the null hypothesis is rejected if the value of the test statistic is an unlikely value in the distribution that it has if the null hypothesis is true.

11. The alternative hypothesis is the only hypothesis that can be supported by the data in a hypothesis test.

Z and t table questions **(2 points each)**

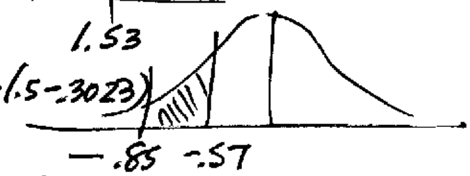


1.53 12. What is z_o, such that $P(Z< z_o)=0.9370$?

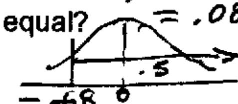
0.0866 13. What does $P(-0.85 < Z < -0.57)$ equal?

$P(Z<-.57)-P(Z<-.85)=(.5-.2157)-(.5-.3023)$

$= .0866$

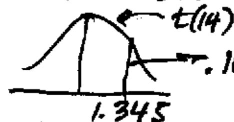


0.7517 14. What does $P(Z > -0.68)$ equal?



$=.5+.2517=.7517$

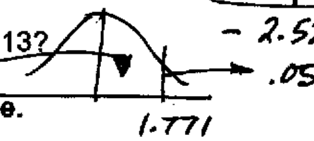
0.10 15. What is the $P(t > 1.345)$ in the t with df=14?



-2.528 16. What is the value of t_o if $P(t < t_o) = .01$, when df=20?

0.45 17. What is the $P(0 < t < 1.771)$, if df=13?

$.5-.05=.45$



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

0.7114 18. What is the p-value of a two-tail hypothesis test based on a large sample if the test statistic value is -0.37?

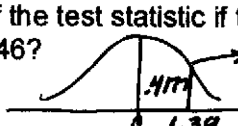
$P/2=.5-.1443=.3557 \Rightarrow$

$P=2(.3557)=.7114$

5.5 19. If a 92% confidence interval to estimate a population mean is (3, 8) what is the value of the point estimate for the population mean?

$\bar{X}=\hat{\mu}$ is the center of the interval, $\bar{X}=\frac{3+8}{2}=5.5$

1.39 20. What is the positive value of the test statistic if the p-value in a two-tail hypothesis test based on a Z test statistic is equal to 0.1646?



$\frac{P}{2}=\frac{.1646}{2}=.0823$, so $.5-.0823=$

$=.4177$

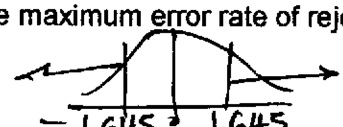
28.96 21. Consider a 98% confidence interval to estimate a population mean based on a sample of 9 observations with a sample mean of 4,325 and a sample variance of 225. How wide is this interval? Round to two digits past the decimal.

$$\text{width}=2B=2\left(t_{\frac{.02}{2}(8)}\right)S_{\bar{x}}=2\,t_{.01(8)}\frac{S}{\sqrt{n}}=2(2.896)\frac{\sqrt{225}}{\sqrt{9}}=28.96$$

38.8 22. If a sample of 8 observations has the values of 8, -2, 5, 3, 9, -4, 7, -8, what is the value of the point estimate for the population variance? Round your answer to one digit past the decimal.

$$S^2=\frac{\Sigma x^2-\frac{(\Sigma x)^2}{n}}{n-1}=\frac{312-\frac{18^2}{8}}{8-1}=\frac{271.5}{7}=38.7857$$

0.10 23. If the rejection region in a two-tail hypothesis test based on a large sample is above 1.645 and below -1.645, what is the maximum error rate of rejecting a true null hypothesis that this researcher will tolerate?

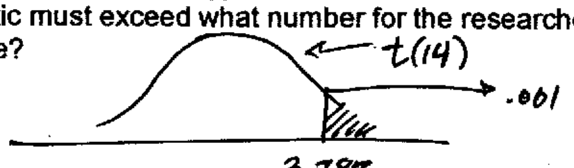


$.05=\frac{\alpha}{2} \Rightarrow \alpha=2(.05)=.10$

$H_A: \mu_1-\mu_2>24$ 24. If the alternative hypothesis to compare two population means is H_a: $\mu_2-\mu_1 < -24$, then what would be the equivalent alternative hypothesis if the statement was written for the parameter $\mu_1-\mu_2$?

$H_A: \mu_2-\mu_1<-24 \Rightarrow H_A: \mu_1-\mu_2>24$

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



State the answer on the line. **(3 points each)**

Eye-level grocery store shelves, referred to as middle shelves, are believed to result in higher numbers of product sales than lower-level shelves. A local retail grocery store has collected data to test this idea. The following data, which are total numbers of weekly sales of a certain type of product, were recorded for middle and lower shelves for eight weeks. Use this data to answer the questions on this page.

MIDDLE:	68	63	59	82	74	60	94	83	$\Sigma x_m = 583$, $\Sigma x_m^2 = 43{,}599$
LOWER:	53	60	39	59	48	46	62	53	$\Sigma x_L = 420$, $\Sigma x_L^2 = 22{,}484$

$\bar{X}_m = 72.875$, $\bar{X}_L = 52.5$

52.5 26. State the point estimate for the mean of the population of weekly sales from the lower shelves. State your answer with one digit past the decimal.

$\bar{X}_L = \hat{\mu}_L = 52.5$

12.6 27. State the point estimate for the standard deviation for the population of sales from the middle shelves. Round your answer to one digit past the decimal.

$$S_m = \hat{\sigma}_m = \sqrt{\frac{\Sigma x^2 - \frac{(\Sigma x)^2}{n}}{n-1}} = \sqrt{\frac{43599 - \frac{583^2}{8}}{8-1}} = 12.6088$$

62 28. State the point estimate for the variance of the population of sales from the lower shelves. State your answer as a whole number with no digits to the right of the decimal.

$$S_L^2 = \hat{\sigma}_L^2 = \frac{\Sigma x^2 - \frac{(\Sigma x)^2}{n}}{n-1} = \frac{22{,}484 - \frac{420^2}{8}}{8-1} = 62$$

20.375 29. State the point estimate for the difference between the mean number of sales from the middle shelves and the mean number of sales from the lower shelves.

$\bar{X}_m - \bar{X}_L = \hat{\mu}_m - \hat{\mu}_L = 72.875 - 52.5 = 20.375$

110.4 30. What is the pooled variance estimate that would result from these two samples? Round your answer to one digit past the decimal.

$$S_p^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2} \underset{\text{if } n_1=n_2}{=} \frac{S_1^2 + S_2^2}{2} = \frac{158.76 + 62}{2} = 110.38$$

14 31. How many degree of freedom are associated with the pooled variance estimate from these two samples?

df for S_p^2 is $n_1+n_2-2 = 8+8-2 = 14$

4.075 32. If the estimated standard error of the difference between the sample means is 5, then what is the value of the test statistic to test whether the mean of the population of sales from the middle shelves is equal to the mean of the population of sales from the lower shelves? State your answer with three digits past the decimal. $S_{\bar{x}_1-\bar{x}_2} = 5$ then $t \Rightarrow$

$$t = \frac{\bar{X}_1 - \bar{X}_2 - D_0}{S_{\bar{x}_1-\bar{x}_2}} = \frac{72.875 - 52.5 - 0}{5} = 4.075$$

10.725 33. What is the value of the bound of error that would be used to construct a 95% confidence interval to estimate the difference between the mean of the population of sales from the middle shelves and the mean of the population of sales from the lower shelves if the estimated standard error for the difference between the sample means is 5? Assume equal population variances. State your answer with three digits past the decimal. If $S_{\bar{x}_1-\bar{x}_2} = 5$ then Bound is

$$B = t_{\frac{.05}{2}(14)} \cdot S_{\bar{x}_1-\bar{x}_2} = 2.145(5) = 10.725$$

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

Last year oil production in the United States was growing faster than anywhere else in the world. The number of active drilling rigs in the US is related to the price of a barrel of oil. The following bivariate data are X=price per barrel of oil and Y=number of active drilling rigs in the US. The questions on this page are associated with fitting a simple linear regression equation to the bivariate data to estimate the number of active drilling rigs in the US based on the price of a barrel of oil.

X=price per barrel of oil	55	58	42	85	80	75	98	95	$\Sigma x = 588$
Y=number of active drilling rigs	990	1080	912	1670	1550	1470	1920	1880	$\Sigma y = 11472$

$\Sigma x^2 = 55^2 + \dots + 95^2 = 46{,}032$, $\Sigma y^2 = 17{,}551{,}344$, $\Sigma xy = 898{,}354$

$SS_{xy} = \Sigma xy - \frac{\Sigma x(\Sigma y)}{n} = 898{,}354 - \frac{588(11472)}{8} = 55162$

$SS_y = \Sigma y^2 - \frac{(\Sigma y)^2}{n} = 17{,}551{,}344 - \frac{11472^2}{8} = 1{,}100{,}496$

898, 354 34. What is the sum of the cross product for the price per barrel of oil and the number of active drilling rigs data provided above?

$\Sigma xy = 55(990) + \dots + 95(1880) = 898{,}354$

2,814 35. What is the corrected sum of squares for the x-variable based on the price per barrel of oil and the number of active drilling rigs data provided above?

$SS_x = \Sigma x^2 - \frac{(\Sigma x)^2}{n} = 46032 - \frac{588^2}{8} = 2{,}814$

19.6 36. What is the least squares estimate of the slope in the linear regression equation to estimate the number of active drilling rigs based on the price of a barrel of oil? Round your answer to one digit past the decimal.

$\hat{\beta}_1 = \frac{SS_{xy}}{SS_x} = \frac{55162}{2814} = 19.6$

− 36 37. If the least squares estimate of the slope is 20, what is the least squares estimate of the y-intercept in the linear regression equation to estimate the number of active drilling rigs based on the price of a barrel of oil? This answer is a negative number. If $\hat{\beta}_1 = 20$, then,

$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} = \frac{\Sigma y}{n} - \hat{\beta}_1\left(\frac{\Sigma x}{n}\right) = \frac{11472}{8} - 20\left(\frac{588}{8}\right) = -36$

1,720 38. If the estimated regression equation is y-hat = -40 + 20 x, what is the least squares estimate of the number of active drilling rigs based on a price of a barrel of oil of $88?

$\hat{y}_{x=88} = -40 + 20(88) = 1{,}720$

0.99 39. What is the numerical value of the estimated linear correlation between the price of a barrel of oil and the number of active drilling rigs? Round your answer to two digits past the decimal.

$r = \frac{SS_{xy}}{\sqrt{SS_x \cdot SS_y}} = \frac{55162}{\sqrt{2814(1100496)}} = .99125$