Math 240 Final Exam
Name_____

1. A confidence interval was used to estimate the proportion of statistics students that are female. A random sample of 72 statistics students generated the following 90% confidence interval: (0.438, 0.642). Using the information above, what size sample would be necessary if we wanted to estimate the true proportion to within ±0.08 using 95% confidence?
   a. 105
   b. 150
   c. 420
   d. 597

2. The head librarian at the Library of Congress has asked her assistant for an interval estimate of the mean number of books checked out each day. The assistant provides the following interval estimate: from 740 to 920 books per day. What is an efficient, unbiased point estimate of the number of books checked out each day at the Library of Congress?
   a. 740
   b. 830
   c. 920
   d. 1,660

3. Suppose we wish to test H_0: $\mu \leq 47$ versus H_1: $\mu > 47$. What will result if we conclude that the mean is greater than 47 when its true value is really 52?
   a. We have made a Type I error.
   b. We have made a Type II error.
   c. We have made a correct decision.
   d. None of the above are correct.

4. The marketing manager for an automobile manufacturer is interested in determining the proportion of new compact-car owners who would have purchased a passenger-side inflatable air bag if it had been available for an additional cost of $300. The manager believes from previous information that the proportion is 0.30. Suppose that a survey of 200 new compact-car owners is selected and 79 indicate that they would have purchased the inflatable air bags. If you were to conduct a test to determine whether there is evidence that the proportion is different from 0.30, which test would you use?
   a. Z-test of a population mean
   b. Z-test of a population proportion
   c. t-test of population mean
   d. χ^2-test of population proportion

TABLE 7-1

Microsoft Excel was used on a set of data involving the number of parasites found on 46 Monarch butterflies captured in Pismo Beach State Park. A biologist wants to know if the mean number of parasites per butterfly is over 20. She will make her decision using a test with a level of 0.10. The following information was extracted from the Microsoft Excel output for the sample of 46 Monarch butterflies:

n = 46; Arithmetic Mean = 28.00; Standard Deviation = 25.92;
Standard Error = 3.82; Null Hypothesis: H_0: $\mu \leq 20.000$;
α = 0.10; df = 45; T Test Statistic = 2.09;
One-Tailed Test Upper Critical Value = 1.3006; p-value = 0.021;
Decision = Reject.

5. Referring to Table 7-1, the parameter the biologist is interested in is:
   a. the mean number of butterflies in Pismo Beach State Park.
   b. the mean number of parasites on these 46 butterflies.
   c. the mean number of parasites on Monarch butterflies in Pismo Beach State Park.
   d. the proportion of butterflies with parasites.

6. In testing for differences between the means of 2 related populations where the variance of the differences is unknown, the degrees of freedom are
   a. $n - 1$.
   b. $n_1 + n_2 - 1$.
   c. $n_1 + n_2 - 2$.
   d. $n - 2$.

TABLE 8-5

To test the effects of a business school preparation course, eight (8) students took a general business test before and after the course. The results are given below.

Student	Exam Score Before Course (1)	Exam Score After Course (2)
1	530	670
2	690	770
3	910	1000
4	700	710
5	450	550
6	820	870
7	820	770
8	630	610

7. Referring to Table 8-5, the number of degrees of freedom is
   a. 14.
   b. 13.
   c. 8.
   d. 7.

8. Referring to Table 8-5, the value of the standard error of the difference scores is
   a. 65.027
   b. 60.828
   c. 22.991
   d. 14.696

9. Referring to Table 8-5, at the 0.05 level of significance, the conclusion for this hypothesis test would be:
   a. the business school preparation course does improve exam score.
   b. the business school preparation course does not improve exam score.
   c. the business school preparation course has no impact on exam score.
   d. It cannot be drawn from the information given.

TABLE 8-4

A real estate company is interested in testing whether, on average, families in Gotham have been living in their current homes for less time than families in Metropolis have. A random sample of 100 families from Gotham and a random sample of 150 families in Metropolis yield the following data on length of residence in current homes.

Gotham: $\bar{X}_G$ = 35 months, s_G^2 = 900 Metropolis: $\bar{X}_M$ = 50 months, s_M^2 = 1050

10. Referring to Table 8-4, suppose α = 0.10. Which of the following represents the result of the relevant hypothesis test?
    a. The alternative hypothesis is rejected.
    b. The null hypothesis is rejected.
    c. The null hypothesis is not rejected.
    d. Insufficient information exists on which to make a decision.

11. Assuming a linear relationship between *X* and *Y*, if the coefficient of correlation (*r*) equals −0.30,
    a. there is no correlation.
    b. the slope (b_1) is negative.
    c. variable *X* is larger than variable *Y*.
    d. the variance of *X* is negative.

TABLE 11-9

It is believed that, the average numbers of hours spent studying per day (HOURS) during undergraduate education should have a positive linear relationship with the starting salary (SALARY, measured in thousands of dollars per month) after graduation. Given below is the Excel output from regressing starting salary on number of hours spent studying per day for a sample of 51 students.
NOTE: Some of the numbers in the output are purposely erased.

Regression Statistics	
Multiple R	0.8857
R Square	0.7845
Adjusted R Square	0.7801
Standard Error	1.3704
Observations	51

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	335.0472	335.0473	178.3859	
Residual			1.8782		
Total	50	427.0798			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-1.8940	0.4018	-4.7134	2.051E-05	-2.7015	-1.0865
Hours	0.9795	0.0733	13.3561	5.944E-18	0.8321	1.1269

12. Referring to Table 11-9, to test the claim that average SALARY depends positively on HOURS against the null hypothesis that average SALARY does not depend linearly on HOURS, the *p*-value of the test statistic is
    a. (5.944E-18)/2
    b. 5.944E-18
    c. (2.051E-05)/2
    d. 2.051E-05

TABLE 11-2

A candy bar manufacturer is interested in trying to estimate how sales are influenced by the price of their product. To do this, the company randomly chooses 6 small cities and offers the candy bar at different prices. Using candy bar sales as the dependent variable, the company will conduct a simple linear regression on the data below:

City	Price ($)	Sales
River Falls	1.30	100
Hudson	1.60	90
Ellsworth	1.80	90
Prescott	2.00	40
Rock Elm	2.40	38
Stillwater	2.90	32

13. Referring to Table 11-2, to test that the regression coefficient, β_1, is not equal to 0, what would be the limits of the rejection region for b_1? Use $\alpha = 0.05$.
    a. -48.193 ± 45.245
    b. -48.193 ± 35.117
    c. -48.193 ± 2.776
    d. 0 ± 35.117

14. In a multiple regression problem involving two independent variables, if b_1 is computed to be +2.0, it means that
    a. the relationship between X_1 and Y is significant.
    b. the estimated average of Y increases by 2 units for each increase of 1 unit of X_1, holding X_2 constant.
    c. the estimated average of Y increases by 2 units for each increase of 1 unit of X_1, without regard to X_2.
    d. the estimated average of Y is 2 when X_1 equals zero.

15. Which of the following is not used to find a "best" model?
    a. Adjusted r^2
    b. Mallow's C_p
    c. Odds ratio
    d. All of the above

TABLE 12-6

One of the most common questions of prospective house buyers pertains to the average cost of heating in dollars (Y). To provide its customers with information on that matter, a large real estate firm used the following 4 variables to predict heating costs: the daily minimum outside temperature in degrees of Fahrenheit (X_1), the amount of insulation in inches (X_2), the number of windows in the house (X_3), and the age of the furnace in years (X_4). Given below are the EXCEL outputs of two regression models.

Model 1

Regression Statistics	
R Square	0.8080
Adjusted R Square	0.7568
Observations	20

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	4	169503.4241	42375.86	15.7874	2.96869E-05
Residual	15	40262.3259	2684.155		
Total	19	209765.75			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 90.0%*	*Upper 90.0%*
Intercept	421.4277	77.8614	5.4125	7.2E-05	284.9327	557.9227
X_1 (Temperature)	-4.5098	0.8129	-5.5476	5.58E-05	-5.9349	-3.0847
X_2 (Insulation)	-14.9029	5.0508	-2.9505	0.0099	-23.7573	-6.0485
X_3 (Windows)	0.2151	4.8675	0.0442	0.9653	-8.3181	8.7484
X_4 (Furnace Age)	6.3780	4.1026	1.5546	0.1408	-0.8140	13.5702

Model 2

Regression Statistics	
R Square	0.7768
Adjusted R Square	0.7506
Observations	20

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	162958.2277	81479.11	29.5923	2.9036E-06
Residual	17	46807.5222	2753.384		

Total	19	209765.75

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	489.3227	43.9826	11.1253	3.17E-09	396.5273	582.1180
X_1 (Temperature)	-5.1103	0.6951	-7.3515	1.13E-06	-6.5769	-3.6437
X_2 (Insulation)	-14.7195	4.8864	-3.0123	0.0078	-25.0290	-4.4099

16. Referring to Table 12-6 and allowing for a 1% probability of committing a type I error, what is the decision and conclusion for the test H_0: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ vs. H_1: At least one $\beta_j \neq 0$, $j = 1,2,...,4$ using **Model 1**?
    a. Do not reject H_0 and conclude that the 4 independent variables have significant individual linear effects on average heating costs.
    b. Reject H_0 and conclude that the 4 independent variables taken as a group have significant linear effects on average heating costs.
    c. Do not reject H_0 and conclude that the 4 independent variables taken as a group do not have significant linear effects on average heating costs.
    d. Reject H_0 and conclude that the 4 independent variables taken as a group do not have significant linear effects on average heating costs.

TABLE 12-4

A real estate builder wishes to determine how house size (House) is influenced by family income (Income), family size (Size), and education of the head of household (School). House size is measured in hundreds of square feet, income is measured in thousands of dollars, and education is in years. The builder randomly selected 50 families and ran the multiple regression. Microsoft Excel output is provided below:

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.865
R Square	0.748
Adjusted R Square	0.726
Standard Error	5.195
Observations	50

ANOVA

	df	SS	MS	*F*	*Signif F*
Regression		3605.7736	901.4434		0.0001
Residual		1214.2264	26.9828		
Total	49	4820.0000			

	Coeff	StdError	*t Stat*	*P-value*
Intercept	-1.6335	5.8078	-0.281	0.7798
Income	0.4485	0.1137	3.9545	0.0003
Size	4.2615	0.8062	5.286	0.0001
School	-0.6517	0.4319	-1.509	0.1383

17. Referring to Table 12-4, which of the following values for the level of significance is the smallest for which the regression model as a whole is significant?
    a. 0.00005
    b. 0.001
    c. 0.01
    d. 0.05

18. Referring to Table 12-4, one individual in the sample had an annual income of $100,000, a family size of 10, and an education of 16 years. This individual owned a home with an area of 7,000 square feet (House = 70.00). What is the residual (in hundreds of square feet) for this data point?
    a. 7.40
    b. 2.52
    c. -2.52
    d. -5.40

TABLE 12-16

A business statistics student decides to predict gasoline prices in different cities and towns in the United States for the term project. The dependent variable is price of gasoline per gallon and the explanatory variables are per capita income, the number of firms that manufacture automobile parts in and around the city, the number of new business starts in the last year, population density of the city, percentage of local taxes on gasoline, and the number of people using public transportation. Data of 32 cities was collected and a regression sum of squares SSR = 122.8821 was obtained. The computed value of standard error of the estimate is 1.9549.

19. Referring to Table 12-16, if variables that measure the number of new business starts in the last year and population density of the city are removed from the multiple regression model, which of the following is true?
    a. The adjusted r^2 will definitely increase.
    b. The adjusted r^2 can never increase.
    c. The coefficient of multiple determination can never increase.
    d. The coefficient of multiple determination will definitely increase.

TABLE 12-11

To explain personal consumption (CONS) measured in dollars, data is collected for

INC:	personal income in dollars
CRDTLIM:	$1 plus the credit limit in dollars available to the individual
APR:	average annualized percentage interest rate for borrowing for the individual
ADVT:	per person advertising expenditure in dollars by manufacturers in the city where the individual lives
SEX:	gender of the individual; 1 if female, 0 if male

A regression analysis was performed with CONS as the dependent variable and ln(CRDTLIM), ln(APR), ln(ADVT) and SEX as the independent variables. The estimated model was

$$\hat{y} = 2.28 - 0.29 \ln(\text{CRDTLIM}) + 5.77 \ln(\text{APR}) + 2.35 \ln(\text{ADVT}) + 0.39 \text{ SEX}$$

20. Referring to Table 12-11, what is the correct interpretation for the estimated coefficient for SEX?
    a. Holding everything else fixed, personal consumption for females is estimated to be $0.39 higher than males on the average.
    b. Holding everything else fixed, personal consumption for males is estimated to be $0.39 higher than females on the average.
    c. Holding everything else fixed, personal consumption for females is estimated to be 0.39% higher than males on the average.
    d. Holding everything else fixed, personal consumption for males is estimated to be 0.39% higher than females on the average.