1. Two international students were talking in the wake of the Singapore Air disaster. The first student said she had been intending to fly Singapore Air for the semester break, but wasn't going to now that there had been an accident! The second said, "But it should be extra safe now. The chances of it crashing again so soon are very small." If the likelihood of an airplane crashing follows the binomial distribution, what are these students forgetting about a binomial random variable?

   A. the binomial random variable is discrete.
   B. the number of trials must be at least 30.
   C. Outcome of each trial is one of two mutually exclusive categories.
   D. the results of one trial are not dependent on the results of the other trials.

2. If the correlation coefficient (r) = 1.00, then

   A. the Y-intercept (b_0) must equal zero.
   B. the explained variation equals the unexplained variation.
   C. there is no unexplained variation.
   D. there is no explained variation.

3. The residuals in a linear regression represent

   A. the difference between the actual Y values and the mean of Y.
   B. the difference between the actual Y values and the predicted Y values.
   C. the square root of the slope.
   D. the predicted value of Y for the average X value.

4. We have created a 95% confidence interval for μ with the result [10, 15]. What conclusion will we make if we test H_0: $\mu = 16$ versus H_1: $\mu \neq 16$ at $\alpha = 0.05$?

   A. Reject the null and conclude the alternative.
   B. Fail to reject the null.
   C. Accept the null.
   D. We cannot tell what our decision will be from the information given.

5. If a test of hypothesis has a Type I error probability of .01, we mean

   A. if the null hypothesis is true, we don't reject it 1% of the time.
   B. if the null hypothesis is true, we reject it 1% of the time.
   C. if the null hypothesis is false, we don't reject it 1% of the time.
   D. if the null hypothesis is false, we reject it 1% of the time.

6. Kristin discovers a dependent relationship between the number of employees and the size of company inventory in the chemical industry: when the number of employees falls, the size of the inventory rises. Because of her discovery, Kristin calculates a correlation coefficient for Employees and Inventory. What is the sign on this coefficient?

   A. positive
   B. negative
   C. depends on which variable is dependent.
   D. it has no units, hence no sign.

7. The least squares method minimizes which of the following?

   A. Regression Sum of Squares
   B. Sum of Squared Error
   C. Total Sum of Squares
   D. All of the above.

The next ***TWO*** questions pertain to this information: You made an investment of $5000 on January 1, 2000. Your investment will be worth $1,000, $2,500 or $5,000 at the end of the year, with respective probabilities of 0.15, 0.60 and 0.25.

8. To the nearest dollar, what is the expected value of your investment?

   A. $4150 C. $4500
   B. $2900 D. $6500

9. Assume that the expected value of your investment is $1500. To the nearest dollar, what is the standard deviation of the value of your investment?

   A. 1064 C. 1400
   B. 1319 D. 1924

10. Which of the following is **TRUE** about the Student *t* distribution?

    A. It has more area in the center and less in the tails than does the normal distribution.
    B. It is used to construct confidence intervals for the population mean when the population standard deviation is known.
    C. It is bell shaped and symmetrical.
    D. As the number of degrees of freedom increases, the t distribution approaches the binomial.

11. For train travelers, one of the biggest complaints is of the "off-track" waiting time when the passenger train is bumped off the track to allow a freight train to pass. This waiting time is known to have a normal distribution with a mean of 25 minutes and a standard deviation of 5 minutes. Suppose 100 train trips have been randomly sampled. Describe the sampling distribution of the mean "off-track" waiting time for these 100 trips.

    A. Distribution is normal with mean = 25 minutes, standard error = 0.5 minutes.
    B. Distribution is normal with mean = 25 minutes, standard error = 5 minutes.
    C. Distribution is approximately normal with mean = 25 minutes, standard error = 0.5 minutes.
    D. Distribution is approximately normal with mean = 25 minutes, standard error = 5 minutes.

12. Refer to **Wine Equation**. How much will the quality of French Bordeaux change when harvest rain decreases by 15 millimeters?

    A. Decrease by .00386 quality units.
    B. Decrease by 12.141 quality units
    C. Increase by .0579 quality units
    D. Increase by .00386 quality units.

13. If an economist interested in poverty wishes to determine whether there is evidence that average family income in a community falls short of $25,000,

    A. either a one-tailed or two-tailed test could be used with equivalent results.
    B. a one-tailed test should be used.
    C. a two-tailed test should be used.
    D. none of the above.

14. Refer to the **Filter Cake Data Set**. What is the point estimate of the proportion of solids in a filter cake?

    A. 11.09 C. 1.37
    B. 11.50 D. 0.188

15. Based on the information in the **Filter Cake Data Set** and assuming the population is normally distributed, the appropriate p-value to use to test the null hypothesis associated with the research question described in the printout is:

    A. =NORM.S.DIST(((11.09-11.5)/0.186),1) = 0.013752
    B. =NORM.S.DIST(((11.5-11.09)/1.37),1) = 0.617633
    C. =T.DIST(-0.41/0.186,53,1) = 0.015934
    D. =T.DIST(0.41/1.37,53,1) = 0.765904

16. Refer to the **Filter Cake Data Set**. Which of the following is the appropriate set of hypotheses for a test such as that described in the printout?

A. $H_0: \bar{x} = 11.5$ $H_1: \bar{x} \neq 11.5$
B. $H_0: \mu \geq 11.5$ $H_1: \mu < 11.5$
C. $H_0: \mu < 11.09$ $H_1: \mu \geq 11.09$
D. $H_0: \bar{x} \leq 11.5$ $H_1: \bar{x} > 11.5$

17. If the p-value associated with the hypothesis test from the **Filter Cake Data Set** is 0.014, this would mean

A. the probability of getting a sample mean less than 11.50 percent solids should be rejected, as it is quite small.
B. the probability of getting a sample mean less than 11.09 percent solids should be rejected, as it is quite large.
C. the probability of getting a sample mean of less than 11.09 percent solids, assuming the null is true, is quite small.
D. the probability of getting a sample mean of less than 11.50 percent solids, assuming the null is true, is quite large.

18. If the p-value associated with the hypothesis test from the **Filter Cake Data Set** is 0.014, the null hypothesis of the test would be

A. rejected at α=0.01 but not at α=0.001.
B. rejected at α=0.001 but not at α=0.01.
C. rejected at both α=0.01 and α=0.001.
D. not rejected at α=0.01 or α=0.001.

19. Refer to the **Candy Bar Sales Data Set.** If Smilin' Sam's management believes that the easiest change that can immediately be made to candy bar location is to shift candy bars from the bottom of the shelf to the top of the shelf. This action will be ordered only if management sees a significant reason to do so. What would be your recommendation?

A. Don't bother to move candy bars from the bottom to the top of the shelf because Shelf is significantly important to Sales at the 5% level.
B. Don't bother to move candy bars from the bottom to the top of the shelf because Shelf is irrelevant to Sales at the 5% level.
C. Move candy bars from the bottom to the top of the shelf because Shelf is significantly important to Sales at the 5% level.
D. Move candy bars from the bottom to the top of the shelf because Shelf is irrelevant to Sales at the 5% level.

20. According to the regression output in the **Candy Bar Sales Data Set,** about how much do candy bar sales change when candy bars are moved from the bottom of the shelf to the top?

    A. Sales increase by 2100 candy bars per month.
    B. Sales decrease by 2100 candy bars per month.
    C. Sales increase by 2850 candy bars per month.
    D. Sales decrease by 2850 candy bars per month.

21. According to the output in the **Candy Bar Sales Data Set**, what is the predicted monthly candy bar sales for a candy bar display located in the rear of the store on the top shelf?

    A. 2100 candy bars
    B. 4950 candy bars
    C. 2850 candy bars
    D. 8290 candy bars

22. Since the purpose of the regression from the **Candy Bar Sales Data Set** is to enable Smilin' Sam's stores to predict maximum candy bar sales, which of the variables in the equation might they be justified in eliminating?

    A. none
    B. Shelf
    C. Aisle
    D. Shelf, Aisle

23. 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 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.

24. Refer to the **Home Heating Oil Data Set**. If a consumer were interested in explaining the monthly amount of heating oil consumed, what variables, if any, might the consumer be justified in dropping from the model?

    A. Insulation, InsuSq
    B. Temp(F), InsuSq
    C. InsuSq
    D. None of the above.

25. Refer to the **Home Heating Oil Data Set**. What is the margin of error for the 95% confidence interval associate with Temp(F)?

A. 0.7
B. -5.39
C. 1.4
D. -1.4

Use this information to answer the next two questions. The Wall Street Journal reported that the mean value of a share of Quaker Oats stock was $95 with a standard deviation of $8 and the mean value of a share of PepsiCo was $45 with a standard deviation of $3. You purchase a portfolio of these stocks that includes one share of Quaker Oats and two shares of PepsiCo.

26. Assuming that the two stocks are independent, what is the standard deviation of your portfolio?

A. $10
B. $14
C. $20
D. $100

27. What is the expected value of your portfolio?

A. $61.5
B. $140
C. $185
D. $275

28. In a simple linear regression, r and b_1

A. may have opposite signs.
B. must have the same sign.
C. must have opposite signs.
D. are equal.

29. What is the r^2 in the **Farmingdale, NY Data Set?**

A. 0.64
B. 0.36
C. 0.56
D. 0.60

30. Suppose that the r^2 in the **Farmingdale, NY Data Set** is .47. This means

A. the proportion of home characteristics to value is 0.47.
B. the proportion of value to home characteristics is 0.47.
C. the proportion of the variation in home characteristics due to variation in value is 0.47.
D. the proportion of the variation in value due to variation in home characteristics is 0.47.

31. With reference to **Farmingdale, NY Data Set**, the number that the least squares regression line minimizes is

A. 743.31
B. 27.26
C. 40138.85
D. 62713.23

32. Refer to the **Farmingdale, NY Data Set** and perform the following hypothesis test:
H_0: $\beta_1 = 0$
H_1: $\beta_1 \neq 0$
At the 5% level of significance you would

A. fail to reject the null and conclude that lot size is significantly important to explaining the assessed value of a home.
B. fail to reject the null and conclude that lot size is completely irrelevant to explaining the assessed value of a home.
C. reject the null and conclude that lot size is significantly important to explaining the assessed value of a home.
D. reject the null and conclude that lot size is completely irrelevant to explaining the assessed value of a home.

33. Refer to **Farmingdale, NY Data Set**. What would be the p-value of the hypothesis test H_0: $\beta_2 \geq 0$ versus H_1: $\beta_2 < 0$?

A. 0.005
B. 0.020
C. 0.010
D. 0.990

34. Which of the following statements about multiple regression is **TRUE**?

A. The total sum of squares in a regression model will never exceed the regression sum of squares.
B. The coefficient of multiple determination is calculated by taking the ratio of the regression sum of squares over the total sum of squares and subtracting that value from 1.
C. A multiple regression is called "multiple" because it has several data points.
D. If we have taken into account all relevant explanatory factors, the residuals from a multiple regression should be random.

35. A sample of 39 statistics students contained 14 females. What size sample would be necessary if we wanted to estimate the true proportion of female statistics students to within ±0.08 with 95% confidence? The z-statistic for a 95% confidence interval is ±1.96.

A. 2606
B. 564
C. 1300
D. 139

36. A student claims that she can correctly identify whether a person is a business major or an economics major by the way the person dresses. She is presented with one person and considers this to be a hypothesis test with the null hypothesis being that the person is a business major and the alternative that the person is an economics major. With reference to this test, what would be a Type II error?

A. Identifying the person as an economics major when the person is an economics major.
B. Identifying the person as an economics major when the person is a business major.
C. Identifying the person as a business major when the person is an economics major.
D. Identifying the person as a business major when the person is a business major.

37. With respect to hypothesis testing, which of the following is **FALSE**?

A. The larger the p-value the more likely one is to reject the null hypothesis.
B. The statement of the null hypothesis always contains an equality.
C. The test statistic measures how close the computed sample statistic has come to the hypothesized population parameter.
D. In testing a hypothesis, statements for the null and alternative hypothesis as well as the selection of the level of significance should precede the collection and examination of the data.

38. Refer to **University Total Costs Data Set.** What is the value of the test statistic for the hypothesis test of Type of School?

A. z =-11.29 C. t =11.29
B. z =-1.13 D. t =1.13

39. Using Excel, at a significance level of 5%, the critical value against which the above test statistic would be compared would be calculated by

A. =T.INV(0.975,df) C. =NORM.S.INV(ABS(.05))
B. =T.INV(0.95,df) D. =NORM.S.INV(0.95)

40. Refer to the **University Total Costs Data Set**. The meaning of the coefficient on Type is

A. ceteris paribus, for each additional dollar of educational cost, the school's type changes by 8.13 units.
B. ceteris paribus, the difference between the educational cost of a public school and a private school is $8130.
C. ceteris paribus, a public school costs $8130 more on average than a private school.
D. ceteris paribus, for each additional dollar of educational cost, the school moves 8.13 units closer to being private.

A certain type of business succeeds 60% of the time. Three such businesses open where they do not compete with one another. For the next two questions, indicate the Excel command that would generate the requested probabilities.

41. What is the probability that at least two of the businesses succeed?

A. =1- BINOM.DIST(1,3,0.60,1)
B. =BINOM.DIST(2,3,0.60,0)
C. =NORM.DIST(2,1.2,0.85,1)
D. =1 - NORM.DIST(2,1.2,0.85,1)
E. B and D

42. What is the probability that at least two of the businesses **FAIL**?

A. =BINOM.DIST(2,3,0.40,0)
B. =1- BINOM.DIST(1,3,0.40,1)
C. =NORM.DIST(2,1.2,0.85,1)
D. =1 - NORMDIST(2,1.2,0.85,1)
E. B and D.

43. Assuming a linear relationship between X and Y, if the coefficient of correlation (r) = -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.

44. We have created a 95% confidence interval for : with the result [10, 15]. We are not happy with the precision of our confidence interval. How can we make it more precise?

A. Increase σ
B. Increase α
C. Increase *n*
D. Increase μ.

45. A national trend predicts that women will account for half of all business travelers in the next three years. To attract them, hotels are providing more amenities that women like. A survey found that 70% offer hairdryers in the room. Based on a random sample of 20 hotels, how many would you expect to offer hairdryers?

A. 20 B. 17 C. 7 D. 14

46. Refer to the **University Total Costs Data Set**. What is the sample size?

A. 79 B. 77 C. 89 D. 80

Candy Bar Sales Data Set

Smilin' Sam's chain of convenience stores is attempting to decide where to position their primary displays of candy bars, in order to maximize candy bar sales. A selection of 8 stores from the chain was randomly drawn. Data were gathered about monthly sales of candy bars (Sales, in hundreds), the aisle where candy bars were located (Aisle, front = 1, rear = 0) and the location on the shelf where candy bars were placed (Shelf, top = 1, bottom = 0). The following regression output was generated using Excel.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.95
R Square	0.89
Adjusted R Square	
Standard Error	
Observations	8

ANOVA

	df	SS
Regression	2	3060
Residual		
Total	7	

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	28.50	5.21	5.47	0.00	15.11	41.89
Aisle	33.00	6.02	5.48	0.00	17.53	48.47
Shelf	21.00	6.02	3.49	0.02	5.53	36.47

University Total Costs Data Set

Indiana University Trustees are interested in investigating average total student costs. Below is the output of a regression of annual total costs (ATC, in thousands of dollars) on Type of School (TYPE, private = 1, public = 0) and Total Mean SAT (SAT).

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.89
R Square	0.79
Adjusted R Square	0.79
Standard Error	2.72
Observations	

ANOVA

	df	SS
Regression		2176
Residual	77	567.75
Total		2743.75

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	-3.47	3.08		0.26	-9.61	2.67
Type of School	8.13	0.72		0.00	6.7	9.56
Average Total SAT	0.015	0.00		0.00	0.01	0.02

Farmingdale, NY Data Set

A citizens group in Farmingdale, New York is interested in suing their community for inappropriate property taxes. The group believe that the community government is not accurately revealing how it assesses property values. Taxes are based on assessed property values. The Farmingdale citizens group gathered data for the variables that the community government reports using as a basis for assessing property values and regressed Assessed Value (in thousands of dollars) on these variables: Lotsize (in linear feet of road frontage), Bed (number of bedrooms), Bath (number of bathrooms), Rooms (number of rooms of any kind), and Age (in years.)

SUMMARY OUTPUT

Regression Statistics	
Multiple R	
R Square	
Adjusted R Square	
Standard Error	27.26
Observations	60
Observations	

ANOVA

	df	SS
Regression	5	
Residual	54	40138.85
Total		62713.23

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	113.28	26.11			60.92	
Lotsize	-0.35	0.98	-0.35	0.73		1.63
Bed	-14.93	5.35	-2.79	0.01	-25.64	-4.21
Bath	33.62	7.96	4.23	0	17.67	
Rooms	9.52	3.6	2.64	0.01	2.3	16.75
Age	0.28	0.28				

Home Heating Oil Data Set

The results of the following regression were published in Consumer Reports to enable consumers to predict the amount of heating oil they might need for the next month. Gallons of heating oil per month were regressed on mean temperature (Temp in degrees Fahrenheit), depth of attic insulation (Insulation in inches) and amount of insulation (InsuSq, in square feet.)

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.99
R Square	
Adjusted R Square	
Standard Error	24.29
Observations	

ANOVA

	df	SS
Regression	3	
Residual		6492.06
Total	14	236135.23

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%
Intercept	624.59	42.44	14.72	0	531.19	717.99
Temp(F)	-5.36	0.32	-16.91	0	-6.06	-4.66
Insulation	-44.59	14.95	-2.98	0.01	-77.5	-11.67
InsuSq	1.87	1.12	1.66	0.12	-0.61	4.34

PARTIAL RESIDUAL OUTPUT

Observation	Predicted Gallons	Residuals
1	293.12	-17.82
2	362.84	0.96
3	150.88	13.42
4	32.8	8

Filter Cake Data Set

A by-product of potatoes processed by The Mountain States Potato Company is a filter cake. It is sold to area feedlots for cattle feed. Recently a feedlot owner complained that cattle were not gaining weight and believed that it was due to a reduction in the quality of the filter cakes. The Company scrambled to assemble data. Historical records indicated that the percentage of solids in the filter cakes had been running at 11.5%. The following is part of an Excel generated set of descriptive statistics for a random sample of filter cakes produced over a 20 day period. The percent of solids in each filter cake selected was measured.

Solids	
Mean	11.09
Standard Deviation	1.37
Range	8.4
Minimum	9.4
Maximum	17.8
Sum	599
Count	54

Wine Equation

$$Q = -12.145 + .00117WR + .6164TMP - .00386HR$$

Q = index of quality of French Bordeaux
WR = winter rain (October through March) in millimeters
TMP = average temperature during growing season (April through September) in degrees Celsius.
HR = harvest rain (August to September) in millimeters

Sample Final A Answers																	
1	D	6	B	11	A	16	B	21	B	26	A	31	C	36	C	41	A
2	C	7	B	12	C	17	C	22	A	27	C	32	B	37	A	42	B
3	B	8	B	13	B	18	D	23	B	28	B	33	A	38	C	43	B
4	A	9	D	14	A	19	C	24	D	29	B	34	D	39	A	44	C
5	B	10	C	15	C	20	A	25	A	30	D	35	D	40	B	45	D
																46	D