MULTIPLE CHOICE. Choose the one alternative that best completes the statement or answers the question. 2.5% Each.

1) The chancellor of a major university was concerned about alcohol abuse on her campus and wanted to find out the proportion of students at her university who visited campus bars on the weekend before the final exam week. Her advisor took a random sample of 250 students. The portion of students in the sample who visited campus bars on the weekend before the final exam week is an example of
   1) a statistic.
   2) a discrete random variable.
   3) a parameter.
   4) a categorical random variable.

2) In a right-skewed distribution,
   1) the median is larger than the arithmetic mean.
   2) the median is less than the arithmetic mean.
   3) the median equals the arithmetic mean.
   4) none of the above

3) When extreme values are present in a set of data, which of the following descriptive summary measures are most appropriate?
   1) CV and range
   2) variance and interquartile range
   3) arithmetic mean and standard deviation
   4) interquartile range and median

4) When using the general multiplication rule, $P(A \text{ and } B)$ is equal to
   1) $P(A)P(B)$
   2) $P(A \mid B)P(B)$
   3) $P(A)P(B)$
   4) $P(B)/P(A)$

5) According to a survey of American households, the probability that the residents own 2 cars if annual household income is over $25,000 is 80%. Of the households surveyed, 60% had incomes over $25,000 and 70% had 2 cars. The probability that the residents of a household own 2 cars and have an income over $25,000 a year is:
   1) 0.48
   2) 0.12
   3) 0.22
   4) 0.18

6) Following question (5), the probability that the residents own 2 cars if annual household income is over $25,000 is 80%. Of the households surveyed, 60% had incomes over $25,000 and 70% had 2 cars. The probability that the residents of a household do not own 2 cars and have an income over $25,000 a year is:
   1) 0.22
   2) 0.18
   3) 0.48
   4) 0.12

7) An airport has been criticized for the waiting times associated with departing flights. While the critics acknowledge that many flights have little or no waiting times, their complaints deal more specifically with the longer waits attributed to some flights. The critics are interested in showing, mathematically, exactly what the problems are. Which type of distribution would best model the waiting times of the departing flights at the Airport?
   1) uniform distribution
   2) normal distribution
   3) binomial distribution
   4) exponential distribution

8) For a given sample size n, if the level of significance (α) is decreased, the power of the test
   1) will increase.
   2) will decrease.
   3) will remain the same.
   4) cannot be determined.

9) The standard error of the mean
   1) measures the variability of the mean from sample to sample.
   2) is never larger than the standard deviation of the population.
   3) decreases as the sample size increases.
   4) all of the above

10) If the expectation of a sampling distribution is located at the parameter it is estimating, then we call that sampling distribution
    1) random.
    2) unbiased.
    3) biased.
    4) a minimum variance.
11) The Central Limit Theorem is important in statistics because
   1) for any population, it says the sampling distribution of the sample mean is approximately normal, regardless of the sample size.
   2) for a large sample, it says the population is approximately normal.
   3) for any size sample, it says the sampling distribution of the sample mean is approximately normal.
   4) for a large sample, it says the sampling distribution of the sample mean is approximately normal, regardless of the shape of the population.

12) Which of the following statements about the sampling distribution of the sample mean is incorrect?
   1) The sampling distribution of the sample mean is approximately normal whenever the sample size is sufficiently large (n ≥ 30).
   2) The mean of the sampling distribution of the sample mean is equal to μ.
   3) The standard deviation of the sampling distribution of the sample mean is equal to σ.
   4) The sampling distribution of the sample mean is generated by repeatedly taking samples of size n and computing the sample means.

13) The Dean of Students mailed a survey to a total of 400 students. The sample included 100 students randomly selected from each of the freshman, sophomore, junior, and senior classes on campus last term. What sampling method was used?
   1) systematic sample
   2) simple random sample
   3) cluster sample
   4) stratified sample

14) It is desired to estimate the average total compensation of CEOs in the service industry. Data were randomly collected from 18 CEOs and the 97% confidence interval was calculated to be ($2,181,260, $5,836,180). Which of the following interpretations is correct?
   1) In the population of service-industry CEOs, 97% of them will have total compensations that fall in the interval $2,181,260 to $5,836,180.
   2) 97% of the sampled total compensation values fell between $2,181,260 and $5,836,180.
   3) We are 97% confident that the mean of the sampled CEOs falls in the interval $2,181,260 to $5,836,180.
   4) We are 97% confident that the average total compensation of all CEOs in the service industry falls in the interval $2,181,260 to $5,836,180.

15) It is desired to estimate the average total compensation of CEOs in the Service industry. Data were randomly collected from 18 CEOs and the 97% confidence interval was calculated to be ($2,181,260, $5,836,180). Based on the interval above, do you believe the average total compensation of CEOs in the service industry is more than $3,000,000?
   1) Yes, and I am 97% confident of it.
   2) I am 97% confident that the average compensation is $3,000,000.
   3) Yes, and I am 78% confident of it.
   4) I cannot conclude that the average exceeds $3,000,000 at the 97% confidence level.

16) Suppose a 95% confidence interval for μ turns out to be (1,000. 2,100). To make more useful inferences from the data, it is desired to reduce the width of the confidence interval. Which of the following will result in a reduced interval width?
   1) Increase the sample size.
   2) Decrease the confidence level.
   3) Both increase the sample size and decrease the confidence level.
   4) Both increase the confidence level and decrease the sample size.

17) The owner of a local nightclub has recently surveyed a random sample of n = 250 customers of the club. She would now like to determine whether or not the mean age of her customers is over 30. If so, she plans to alter the entertainment to appeal to an older crowd. If not, no entertainment changes will be made. The appropriate hypotheses to test are:
   1) Hp: μ ≤ 30 versus H1: μ > 30.
   2) Hp: X ≥ 30 versus H1: X < 30.
18) Following question (17), if she wants to be 99% confident in her decision, what rejection region should she use?
   1) Reject $H_0$ if $t = Z < -2.58$.
   2) Reject $H_0$ if $t = Z < -2.34$.
   3) Reject $H_0$ if $t = Z > 2.58$.
   4) Reject $H_0$ if $t = Z > 2.34$.

19) Following questions (17)-(18), suppose she found that the sample mean was 30.45 years and the sample standard deviation was 5 years. If she wants to be 99% confident in her decision, what decision should she make?
   1) Accept $H_0$.
   2) Reject $H_0$.
   3) Fail to reject $H_0$.
   4) We cannot tell what her decision should be from the information given.

20) Following questions (17)-(19), suppose she found that the sample mean was 30.45 years and the sample standard deviation was 5 years. If she wants to be 99% confident in her decision, what conclusion can she make?
   1) There is sufficient evidence that the mean age of her customers is not over 30.
   2) There is sufficient evidence that the mean age of her customers is over 30.
   3) There is not sufficient evidence that the mean age of her customers is over 30.
   4) There is not sufficient evidence that the mean age of her customers is not over 30.

A student claims that he can correctly identify whether a person is a business major or an agriculture major by the way the person dresses. Suppose in actuality that if someone is a business major, he can correctly identify that person as a business major 87% of the time. When a person is an agriculture major, the student will incorrectly identify that person as a business major 16% of the time. Presented with one person and asked to identify the major of this person (who is either a business or agriculture major), he 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 agriculture major.

21) Referring to the statement above, what would be a Type I error?
   1) Saying that the person is an agriculture major when in fact the person is an agriculture major.
   2) Saying that the person is a business major when in fact the person is an agriculture major.
   3) Saying that the person is a business major when in fact the person is a business major.
   4) Saying that the person is an agriculture major when in fact the person is a business major.

22) Following question (21), what would be a Type II error?
   1) Saying that the person is an agriculture major when in fact the person is a business major.
   2) Saying that the person is a business major when in fact the person is an agriculture major.
   3) Saying that the person is an agriculture major when in fact the person is a business major.
   4) Saying that the person is a business major when in fact the person is a business major.

23) If the Durbin–Watson statistic has a value close to 4, which assumption is violated?
   1) normality of the errors
   2) homoscedasticity
   3) independence of errors
   4) none of the above
TABLE 1
A researcher randomly sampled 30 graduates of an MBA program and recorded data concerning their starting salaries. Of primary interest to the researcher was the effect of gender on starting salaries. Analysis of the mean salaries of the females and males in the sample is given below.

Hypothesized Difference = 0; Significance level = 0.05
Population 1 female sample: n=18; sample mean = 48266.7; sample standard deviation = 13577.63
Population 2 male sample: n=12; sample mean = 55000; sample standard deviation = 11741.29
t-Test statistic=-1.40193; p-value=0.085962; Lower critical value=-1.70113

24) Referring to Table 1, the researcher was attempting to show statistically that the female MBA graduates have a significantly lower mean starting salary than the male MBA graduates. According to the test run, which of the following is an appropriate alternative hypothesis?
1) $H_1: \mu_{females} < \mu_{males}$
2) $H_1: \mu_{females} = \mu_{males}$
3) $H_1: \mu_{females} > \mu_{males}$
4) $H_1: \mu_{females} \neq \mu_{males}$

25) Referring to Table 1, the researcher attempted to show that the female MBA graduates have a significantly lower mean starting salary than the male MBA graduates. From the analysis, the correct test statistic is:
1) -1.7011
2) 0.0860
3) -6.733.33
4) -1.4019

26) Referring to Table 1, the researcher attempted to show statistically that the female MBA graduates have a significantly lower mean starting salary than the male MBA graduates. The proper conclusion for this test, at the $\alpha = 0.10$ level, is:
1) there is sufficient evidence to indicate a difference in the mean starting salaries of male and female MBA.
2) there is insufficient evidence to indicate that females have a lower mean starting salary than male MBA.
3) there is insufficient evidence to indicate any difference in the mean starting salaries of male and female MBA.
4) there is sufficient evidence to indicate that females have a higher mean starting salary than male MBA.

27) Referring to Table 1, the researcher attempted to show that the female MBA graduates have a significantly lower mean starting salary than the male MBA graduates. What assumptions were necessary to conduct this test?
1) The population variances are approximately equal.
2) The samples were randomly and independently selected.
3) Both populations of salaries (male and female) must have approximate normal distributions.
4) All of the above assumptions were necessary.

28) The standard error of the estimate is a measure of the
1) variation of the X variable.
2) variation around the sample regression line.
3) total variation of the Y variable.
4) explained variation.

A few years ago, Pepsi invited consumers to take the "Pepsi Challenge." Consumers were asked to decide which of the two sodas, Coke or Pepsi, they preferred in a blind taste test. Pepsi was interested in determining what factors played a role in people's taste preferences. One of the factors studied was the gender of the consumer. Below are the results of analyses comparing the taste preferences of men and women with the proportions depicting preference for Pepsi.

Males: $n = 109$, $p_M = 0.422018$
Females: $n = 52$, $p_F = 0.25$
$P_M - p_F = 0.172018$
$Z = 2.11825$

29) Referring to the statement above, to determine if a difference exists in the taste preferences of men and women, give the correct alternative hypothesis that Pepsi would test.
1) $H_1: \mu_M - \mu_F > 0$
2) $H_1: \pi_M - \pi_F \neq 0$
3) $H_1: \mu_M - \mu_F = 0$
4) $H_1: \pi_M - \pi_F = 0$
30) Referring to the statement above, suppose that the two-tailed p value was really 0.0734. State the conclusion.

1) At α = 0.05, there is sufficient evidence to indicate that the proportion of males preferring Pepsi equals the proportion of females preferring Pepsi.
2) At α = 0.08, there is insufficient evidence to indicate that the proportion of males preferring Pepsi differs from the proportion of females preferring Pepsi.
3) At α = 0.10, there is sufficient evidence to indicate that the proportion of males preferring Pepsi differs from the proportion of females preferring Pepsi.
4) At α = 0.05, there is sufficient evidence to indicate that the proportion of males preferring Pepsi differs from the proportion of females preferring Pepsi.

TABLE 2
A campus researcher wanted to investigate the factors that affect visitor travel time in a complex, multilevel building on campus. Specifically, he wanted to determine whether different building signs (building maps versus wall signage) affect the total amount of time visitors require to reach their destination and whether that time depends on whether the starting location is inside or outside the building. Three subjects were assigned to each of the combinations of signs and starting locations, and travel time in seconds from beginning to destination was recorded. An output of the appropriate analysis is given below:

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>P-value</th>
<th>F crit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signs</td>
<td>14008.33</td>
<td>14008.33</td>
<td>0.11267</td>
<td>5.317645</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Starting Location</td>
<td>12288</td>
<td>48</td>
<td>2784395</td>
<td>0.13374</td>
<td>5.317645</td>
<td></td>
</tr>
<tr>
<td>Interaction</td>
<td>48</td>
<td>48</td>
<td>0.919506</td>
<td>5.317645</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Within</td>
<td>35305.33</td>
<td>4413.167</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

31) Referring to Table 2, the degrees of freedom for the different building signs (factor A) is

1) 1
2) 2
3) 3
4) 8

32) Referring to Table 2, the mean squares for starting location (factor B) is

1) 14,008.3
2) 4,413.17
3) 48
4) 12,288

33) Referring to Table 2, the F test statistic for testing the main effect of types of signs is

1) 3.1742
2) 2.7844
3) 5.3176
4) 0.0109

34) Referring to Table 2, at 1% level of significance,

1) there is no sufficient evidence to conclude that the difference between the average traveling time for the different starting locations depends on the types of signs.
2) there is no sufficient evidence to conclude that the relationship between traveling time and the types of signs depends on the starting locations.
3) there is no sufficient evidence to conclude that the difference between the average traveling time for the different types of signs depends on the starting locations.
4) All of the above.

TABLE 3
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:

<table>
<thead>
<tr>
<th>City</th>
<th>Price ($)</th>
<th>Sales</th>
</tr>
</thead>
<tbody>
<tr>
<td>River Falls</td>
<td>1.30</td>
<td>100</td>
</tr>
<tr>
<td>Hudson</td>
<td>1.60</td>
<td>90</td>
</tr>
<tr>
<td>Ellsworth</td>
<td>1.80</td>
<td>90</td>
</tr>
<tr>
<td>Prescott</td>
<td>2.00</td>
<td>40</td>
</tr>
<tr>
<td>Rock Elm</td>
<td>2.40</td>
<td>38</td>
</tr>
<tr>
<td>Stillwater</td>
<td>2.90</td>
<td>32</td>
</tr>
</tbody>
</table>

35) Referring to Table 3, what is the estimated slope parameter for the candy bar price and sales data?

1) -48.193
2) 0.784
3) 161.386
4) -3.810
36) Referring to Table 3, what is the estimated average change in the sales of the candy bar if price goes up by $1.00?
   1) -3.810  2) 161.386  3) -48.193  4) 0.784

37) Based on the residual plot below, you will conclude that there might be a violation of which of the following assumptions?

![Footage Residual Plot]

1) Normality of errors  2) Linearity of the relationship  3) Independence of errors  4) Homoscedasticity

**TABLE 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. An analysis output is provided below.

**SUMMARY OUTPUT**

<table>
<thead>
<tr>
<th>Regression Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple R</td>
</tr>
<tr>
<td>R Square</td>
</tr>
<tr>
<td>Adjusted R Square</td>
</tr>
<tr>
<td>Standard Error</td>
</tr>
<tr>
<td>Observations</td>
</tr>
</tbody>
</table>

**ANOVA**

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>Significance F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>4</td>
<td>3605.7736</td>
<td>901.4434</td>
<td>0.0001</td>
<td></td>
</tr>
<tr>
<td>Residual</td>
<td></td>
<td>1214.2264</td>
<td>26.9828</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>49</td>
<td>4820.0000</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Coefficients**

<table>
<thead>
<tr>
<th></th>
<th>Standard Error</th>
<th>t Stat</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>5.8078</td>
<td>-0.281</td>
<td>0.7798</td>
</tr>
<tr>
<td>Income</td>
<td>0.4485</td>
<td>3.9545</td>
<td>0.0003</td>
</tr>
<tr>
<td>Size</td>
<td>0.8062</td>
<td>5.286</td>
<td>0.0001</td>
</tr>
<tr>
<td>School</td>
<td>0.4319</td>
<td>-1.509</td>
<td>0.1383</td>
</tr>
</tbody>
</table>

38) Referring to Table 4, what fraction of the variability in house size is explained by income, size of family, and education?
   1) 74.8%  2) 27.0%  3) 86.5%  4) 33.4%

39) Referring to Table 4, which of the independent variables in the model are significant at the 2% level?
   1) Income, Size, School  2) Income, Size  3) Income, School  4) Size, School

40) Referring to Table 4, what is the predicted house size (in hundreds of square feet) for an individual earning an annual income of $40,000, having a family size of 4, and going to school a total of 13 years?
   1) 11.43  2) 15.15  3) 53.87  4) 24.88