1. A p-value may be interpreted as the probability of
   (A) The null hypothesis
   (B) The alternative hypothesis
   (C) A type 1 error
   (D) A type 2 error

2. A study of PSU students failed (p-value = .32) to find a difference in pulse rates between men and women.
   (A) The null hypothesis is rejected
   (B) It is possible that they committed a Type 2 error
   (C) It is possible that they committed a Type 1 error
   (D) The research hypothesis is supported

3. A statistically significant result means that it is an important result.
   (A) True
   (B) False

4. In the case of a jury trial, the null hypothesis is innocent and the alternative hypothesis is guilty. In general our society considers:
   (A) A Type 1 error is more serious than a Type 2 error
   (B) A Type 2 error is more serious than a Type 1 error
   (C) Type 1 and 2 errors are equally serious
   (D) Neither error is particularly serious.

5. In a study to determine if putting newborn babies in an incubator contributed to claustrophobia in adult life, the researchers found a p-value of .023. This study supports:
   (A) Skeptic
   (B) Research Advocate
   (C) Both
   (D) Neither the Skeptic nor the Research Advocate

6. A study was conducted to see if PSU students sleep fewer than 8 hours. The study was based on a sample of 100 students. The sample mean number of hours of sleep was 7 and the SD was 5 hours. The p-value is:
   (A) .05
   (B) .025
   (C) .95
   (D) .975

7. A research experiment found that ABC bubblegum retains its flavor longer than XYZ bubblegum. We then know that:
   (A) the p-value was large
   (B) the p-value was small
   (C) they must have committed a Type 1 error
   (D) they must have committed a Type 2 error
8. Suppose after viewing the results of a study you decide, on the basis of the reported p-value, to support the Skeptic. Then
   (A) It is impossible for you to have committed either a Type 1 or Type 2 error
   (B) You must have committed a Type 1 error
   (C) It is impossible for you to have committed a Type 1 error
   (D) It is impossible for you to have committed a Type 2 error

9. In a randomized experiment, if the p-value is small when comparing the treatment and control groups, we can infer that the treatment caused the difference:
   (A) True
   (B) False

10. In a study to see if a new variety of popcorn pops faster than the old variety we collected the following data on time to complete popping in minutes:

<table>
<thead>
<tr>
<th></th>
<th>Old variety</th>
<th>New variety</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>15</td>
<td>10</td>
</tr>
<tr>
<td>SEM</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

   (A) We would support the Skeptic
   (B) We would support the Research Advocate
   (C) Not enough information to decide

11. If the sample size is large enough, almost any null hypothesis can be rejected.
   (A) True
   (B) False

12. Consider the research hypothesis that there is a difference in the proportions of men and women at PSU who own cell phones. Data from the class survey question: Do you own a cell phone?

<table>
<thead>
<tr>
<th></th>
<th>No</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>26</td>
<td>51</td>
</tr>
<tr>
<td>Male</td>
<td>19</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>45</td>
<td>67</td>
</tr>
</tbody>
</table>

If the chi-square statistic is 4.215, the p-value is:
   (A) equal to .05
   (B) greater than .05
   (C) less than .05
   (D) can’t tell

13. An experiment was conducted to see if adding Quaker State motor oil to Coppertone suntan lotion enhances tanning. A Type 1 error is:
   (A) claim QS enhances tanning when it does not
   (B) claim QS does not enhance tanning when it does
   (C) claim QS enhances tanning
   (D) claim QS does not enhance tanning
14. The following table shows some data summaries for amount of change carried by people in a sample taken from class.

<table>
<thead>
<tr>
<th></th>
<th>Men</th>
<th>Women</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>59</td>
<td>82</td>
</tr>
<tr>
<td>SEM</td>
<td>10</td>
<td>12</td>
</tr>
<tr>
<td>SD of difference</td>
<td>16</td>
<td></td>
</tr>
</tbody>
</table>

To test the research hypothesis that women carry more change than men, we must compute the test statistic:

(A) 2.30  
(B) 1.44  
(C) 23  
(D) can’t tell

15. Consider the research hypothesis: Working at least 5 hours per day at a computer contributes to deterioration of eyesight. The null hypothesis is:

(A) working at least 5 hours per day does not affect your eyesight  
(B) working at least 5 hours per day improves your eyesight  
(C) working at least 5 hours per day contributes to the deterioration of eyesight  
(D) cannot determine the null hypothesis

16. In the previous problem, the alternative hypothesis is:

(A) Working at least five hours a day at a computer does not contribute to the deterioration of eyesight.  
(B) Working at least five hours a day at a computer improves eyesight.  
(C) Working at least five hours a day at a computer contributes to deterioration of eyesight.  
(D) Insufficient information is given to allow us to determine the alternative hypothesis.

17. A statistical study considers the question of whether the presence of plants in an office might lead to fewer sick days. In this study, the null hypothesis is:

(A) The presence of plants in an office leads to fewer sick days.  
(B) The presence of plants in an office does not lead to fewer sick days.  
(C) The presence of sick people in an office leads to fewer plants.  
(D) Insufficient information is given to allow us to determine the null hypothesis.

18. In the previous problem, the alternative hypothesis is:

(A) The presence of plants in an office leads to fewer sick days.  
(B) The presence of plants in an office does not lead to fewer sick days.  
(C) The presence of sick people in an office leads to fewer plants.  
(D) Insufficient information is given to allow us to determine the alternative hypothesis.

19. A statistical study considers the question of whether highly educated people are less likely to develop Alzheimer’s disease than others. In this study, the null hypothesis is:

(A) There is no relationship between level of education and the development of Alzheimer’s disease.  
(B) There is a relationship between level of education and the development of Alzheimer’s disease.  
(C) Highly educated people are certain of developing Alzheimer’s disease.  
(D) Insufficient information is given to allow us to determine the null hypothesis.
20. In the previous problem, the alternative hypothesis is:
   (A) There is no relationship between level of education and the development of Alzheimer’s disease.
   (B) There is a relationship between level of education and the development of Alzheimer’s disease.
   (C) Highly educated people are less likely than others to develop Alzheimer’s disease.
   (D) Insufficient information is given to allow us to determine the alternative hypothesis.

21. In a statistical test of hypotheses, a Type I error is committed if:
   (A) We reject the null hypothesis when, in fact, the null hypothesis is invalid.
   (B) We fail to reject the null hypothesis when, in fact, the null hypothesis is valid.
   (C) We reject the null hypothesis when, in fact, the null hypothesis is valid.
   (D) We fail to reject the null hypothesis when, in fact, the null hypothesis is invalid.

22. In a statistical test of hypotheses, a Type II error is committed if:
   (A) We reject the null hypothesis when, in fact, the null hypothesis is invalid.
   (B) We fail to reject the null hypothesis when, in fact, the null hypothesis is valid.
   (C) We reject the null hypothesis when, in fact, the null hypothesis is valid.
   (D) We fail to reject the null hypothesis when, in fact, the null hypothesis is invalid.

The following material pertains to the next five questions: Lee Salk exposed one group of newly born infants, the treatment group, to the sound of a human heartbeat. Next, Salk compared their weight gains to that of a group of newly born infants not exposed, the control group.

23. In Lee Salk’s experiment involving infants, the null hypothesis is:
   (A) Infants exposed to the sound of a human heartbeat will hear the heartbeat.
   (B) Infants not exposed to the sound of a human heartbeat will hear the heartbeat.
   (C) Infants exposed to the sound of a human heartbeat will gain a higher mean weight than infants who are not exposed to the sound of a heartbeat.
   (D) Infants exposed to the sound of a human heartbeat will gain the same mean weight as infants who are not exposed to the sound of a heartbeat.

24. In Salk’s experiment, the alternative hypothesis is:
   (A) Infants exposed to the sound of a human heartbeat will gain a higher mean weight than infants not exposed to the sound of a heartbeat.
   (B) Infants exposed to the sound of a human heartbeat will hear the heartbeat.
   (C) Infants not exposed to the sound of a human heartbeat will hear the heartbeat.
   (D) Infants exposed to the sound of a human heartbeat will gain the same mean weight as infants who are not exposed to the sound of a heartbeat.

25. In Salk’s experiment, a Type I error occurs if:
   (A) The study rejects the hypothesis that exposed infants have the same mean weight gain as unexposed infants when, in fact, this hypothesis is valid.
   (B) Infants exposed to the sound of a human heartbeat actually hear the heartbeat.
   (C) Infants not exposed to the sound of a human heartbeat do not hear the heartbeat.
   (D) The study fails to reject the hypothesis that exposed infants have the same mean weight gain as unexposed infants when, in fact, this hypothesis is not valid.
26. In Salk’s experiment, a Type II error occurs if:

(A) The study rejects the hypothesis that exposed infants have the same mean weight gain as unexposed infants when, in fact, this hypothesis is valid.

(B) Infants exposed to the sound of a human heartbeat actually hear the heartbeat.

(C) Infants not exposed to the sound of a human heartbeat do not hear the heartbeat.

(D) The study fails to reject the hypothesis that exposed infants have the same mean weight gain as unexposed infants when, in fact, this hypothesis is not valid.

27. Salk concluded from his data that the treatment group had higher mean weight gain than the control group. The error which he possibly commits here is:

(A) A Type I error.

(B) A Type II error.

(C) No error at all; we know that infants exposed to a heartbeat are healthier.

(D) In thinking that the sound of a heartbeat could have an effect on an infant’s weight.

28. During lunch today, I found a shiny new dime! To study the problem of whether or not the coin is fair, I choose as my test statistic the number of heads obtained in 20 tosses. When I tossed the coin 20 times, I obtained 13 heads. Bearing in mind that the probability of 13 or more heads in 20 tosses of a fair coin is 13.16%, my decision is to:

(A) Reject the null hypothesis: I have strong evidence against the hypothesis that the coin is fair.

(B) Fail to reject the null hypothesis: I do not have strong evidence against the hypothesis that the coin is fair.

(C) Reject the null hypothesis: I have strong evidence against the hypothesis that the coin is unfair.

(D) Fail to reject the null hypothesis: I do not have strong evidence against the hypothesis that the coin is unfair.

29. Fiona receives a beautiful four-sided die for her eighteenth birthday. After playing with it for two hours, she starts to suspect that her die is more favorable to rolling “1” than to any other number. Fiona’s alternative hypothesis is

(A) One-sided

(B) Two-sided

30. Fiona’s null hypothesis is that

(A) The die rolls “1” with probability less than 0.25.

(B) The die rolls “1” with probability greater than 0.25.

(C) The die rolls “1” with probability equal to 0.25.

(D) The die rolls “1” with probability not equal to 0.25.

Fiona chooses for her test statistic the standardized score corresponding to the sample proportion of 1’s obtained in 300 rolls of her die. The standardized score is

\[
\frac{\text{Sample proportion} - \text{Population proportion}}{\text{S.D. of the sample proportion}}
\]

Fiona rolled her die 300 times and obtained a “1” on 93, or 31%, of her rolls.

31. The value of Fiona’s test statistic is:

(A) \( \frac{0.31-0.25}{\sqrt{\frac{(0.31)(1-0.31)}{300}}} = 2.25 \)

(B) \( \frac{0.25-0.31}{\sqrt{\frac{(0.25)(1-0.25)}{300}}} = -2.4 \)

(C) \( \frac{300-93}{\sqrt{\frac{(0.25)(1-0.25)}{300}}} = 8.280 \)

(D) \( \frac{0.31-0.25}{\sqrt{\frac{(0.25)(1-0.25)}{300}}} = 2.4 \)
32. Using the table of standard normal scores, we find that the $P$-value for Fiona’s die-rolling problem is:

(A) .99, i.e., the area below the normal curve and to the left of 2.05
(B) .98, i.e., the area below the normal curve and to the right of -2.4
(C) 0, i.e., the area below the normal curve and to the right of 8.280
(D) .01, i.e., the area below the normal curve and to the right of 2.4

The following material pertains to the next four questions: Joe Palermo interviewed 507 randomly chosen PSU students and found that 59% of the students in his sample like to play chess. Consider the research question of whether or not a majority of PSU students like to play chess.

33. The test for this research question is a:

(A) One-sided test.
(B) Two-sided test.
(C) Both a one-sided and two-sided test.
(D) Neither a one-sided nor two-sided test.

34. The null hypothesis for this research question is the statement:

(A) The proportion of all PSU students who like to play chess is less than one-half.
(B) The proportion of all PSU students who like to play chess is one-half.
(C) The proportion of all PSU students who like to play chess is greater than one-half.
(D) The proportion of all PSU students who like to play chess is not equal to one-half.

35. The alternative hypothesis for this research question is the statement:

(A) The proportion of all PSU students who like to play chess is less than one-half.
(B) The proportion of all PSU students who like to play chess is one-half.
(C) The proportion of all PSU students who like to play chess is greater than one-half.
(D) The proportion of all PSU students who like to play chess is not equal to one-half.

36. The value of Joe’s test statistic (standardized score) is:

(A) $\frac{0.59-0.5}{\sqrt{0.59(1-0.5)}/507} = 4.12$
(B) $\frac{0.5-0.5}{\sqrt{0.5(1-0.5)/507}} = 0$
(C) $\frac{0.59-0.5}{\sqrt{507}} = 2.03$
(D) $\frac{0.59-0.5}{\sqrt{0.5(1-0.5)/507}} = 4.05$

The following material pertains to the next four questions: To study the effects of exercise on lean body (muscle) weight change, a random sample of 36 students was placed on a two-month long exercise program. At the end of the program, all 36 students’ changes in lean body weight were measured. The sample mean change in muscle weight was 1.05 pounds and the sample standard deviation was 3.6 pounds. The study organizers wish to know if the results of this sample provide good evidence that this exercise program causes a statistically significant change in the population mean lean body weight.

37. The null hypothesis for this research question is the statement:

(A) The population mean change in lean body weight due to this exercise program is greater than zero.
(B) The population mean change in lean body weight due to this exercise program is less than zero.
(C) The population mean change in lean body weight due to this exercise program is equal to zero.
(D) The population mean change in lean body weight due to this exercise program is not equal to zero.
38. The alternative hypothesis for this research question is the statement:

(A) The population mean change in lean body weight due to this exercise program is greater than zero.
(B) The population mean change in lean body weight due to this exercise program is less than zero.
(C) The population mean change in lean body weight due to this exercise program is equal to zero.
(D) The population mean change in lean body weight due to this exercise program is not equal to zero.

39. The test for this research question is a:

(A) One-sided test.
(B) Two-sided test.
(C) Both a one-sided and two-sided test.
(D) Neither a one-sided nor two-sided test.

40. The value of this test statistic (standardized score) is:

(A) \( \frac{1.05}{3.6} = 0.29 \)
(B) \( \frac{1.05}{\sqrt{36}} = 1.75 \)
(C) \( \frac{3.6}{1.05} = 3.43 \)
(D) \( \frac{3.6}{1.05/\sqrt{36}} = 20.57 \)

41. A p-value is a probability that is computed under what assumption?

(A) The null hypothesis is true
(B) The alternative hypothesis is true
(C) A type 1 error has been committed
(D) A type 2 error has been committed

42. Meta-analysis is

(A) a collection of statistical techniques for combining studies
(B) a way to make statistical significance equivalent to practical significance
(C) the use of stratified or cluster sampling
(D) the computation of a test statistic followed by a decision regarding a null hypothesis for data presented in the form of a table

43. The “file drawer” problem refers to

(A) the bias resulting from considering only published studies, which are more likely to contain statistically significant results than those unpublished studies sitting in file drawers.
(B) the challenge of building quality furniture.
(C) a gambling choice in which the player must choose between two drawers.
(D) the fact that studies that have been sitting around for many years are out-of-date.

44. Suppose ten studies were conducted to assess the relationship between watching violence on television and subsequent violent behavior in children. Suppose that none of the ten studies detected a statistically significant relationship. What would be the result of applying the vote-counting method to this example?

(A) Vote-counting will detect a relationship in this example.
(B) Vote-counting will not detect a relationship in this example.
(C) Vote-counting might detect a relationship in this example, but we would need more information.
45. Suppose ten studies were conducted to assess the relationship between watching violence on television and subsequent violent behavior in children. Suppose that none of the ten studies detected a statistically significant relationship. True or false: It is impossible for a meta-analysis to detect a statistically significant relationship in this example.

   (A) True
   (B) False

46. Which of the following is a problem that can occur in a meta-analysis when data sets are combined inappropriately?

   (A) the vote-counting paradox
   (B) the Fibonacci effect
   (C) Simpson’s paradox
   (D) the file drawer problem