The following questions are similar to the types of questions you will see on the midterm exam. The actual midterm will contain 33 to 40 multiple-choice questions, a few of which draw on specific material from lectures.

1. In a random sample of 1,600 PSU students, 64% reported a preference for root beer (over birch beer). In estimating the population proportion of all PSU students who prefer root beer, we measure the precision of the sample proportion by:
   (A) The sample proportion: 64%.
   (B) The sample size as a proportion of all PSU students: \( \frac{1,600}{82,000} \), or 1.95%.
   (C) The margin of error: \( \frac{1}{\sqrt{1,600}} \), or 2.5%.
   (D) The square of the sample size: \( 1,600^2 \), or 960,000.

2. In the previous question, to obtain a margin of error of 1%, we need a sample of size:
   (A) 100
   (B) 100,000
   (C) 1,000
   (D) 10,000

3. The standard deviation of the histogram of a large number of sample means is, approximately:
   (A) The true mean of the population.
   (B) The mean of a proportion of the population which is never sampled.
   (C) \( \frac{\text{population standard deviation}}{\sqrt{\text{sample size}}} \).
   (D) The area below the normal curve and between -2 and +2.

4. The histogram of sample means from a large number of equally-sized random samples will be shaped approximately like a:
   (A) Semi-circle.
   (B) Triangle.
   (C) Skewed histogram, with a long right tail.
   (D) Normal (bell-shaped) curve.
5. The mean of a large number of sample means from equally-sized random samples will be approximately:

(A) The true mean of the population.

(B) The mean of a proportion of the population which is never sampled.

(C) \( \text{population standard deviation} / \sqrt{\text{sample size}} \).

(D) The area below the normal curve and between -1.96 and +1.96.

6. A confidence interval for a population mean is given as “Sample mean \( \pm 2.33 \times \text{SEM} \)”.
   The corresponding level of confidence is:

(A) 64% 

(B) 95% 

(C) 98% 

(D) 99% 

7. The standard deviation of the histogram of a large number of sample proportions is, approximately:

(A) The true proportion of the population.

(B) A proportion of the population which is never sampled.

(C) The square-root of: \( \text{true proportion} \times (1 - \text{true proportion}) / (\text{sample size}) \).

(D) The area below the normal curve and between -1.96 and +1.96.

(E) The square-root of: \( \text{true proportion} \times (\text{sample size}) / (1 - \text{true proportion}) \).

8. It is known that in the presidential elections of 1992, 56% of all adults actually voted. If we collect a large number of simple random samples each of size 1,600 adults then, about 95% of the time, the sample proportion of adults who voted will fall between:

(A) .56 \( \pm 1 \times \sqrt{.56(1-.56)}/1600 \), or 0.548 and 0.572.

(B) .56 \( \pm 2 \times \sqrt{.56(1-.56)}/1600 \), or 0.536 and 0.584.

(C) .56 \( \pm 3 \times \sqrt{.56(1-.56)}/1600 \), or 0.524 and 0.596.

(D) .56 \( \pm 4 \times \sqrt{.56(1-.56)}/1600 \), or 0.512 and 0.608.

9. In the previous problem, a researcher obtained a particular random sample of 1,600 adults of whom 61% claimed to have voted. This claim is:

(A) Plausible, because .61 falls outside the interval 0.548 to 0.572.

(B) Implausible, because .61 falls outside the interval 0.536 to 0.584.

(C) Implausible, because people probably can’t recall whether or not they voted.

(D) Plausible, because .61 falls outside the interval 0.524 to 0.596.
10. The mean of a large number of sample proportions from equally-sized random samples will be approximately:
   (A) The true proportion of the population.
   (B) A proportion of the population which is never sampled.
   (C) The square-root of: \((\text{true proportion}) \times (1 - \text{true proportion})/(\text{sample size})\).
   (D) The area below the normal curve and between -1.96 and +1.96.
   (E) The square-root of: \((\text{true proportion}) \times (\text{sample size})/(1 - \text{true proportion})\).

11. \textit{Note: This problem originally contained an error; it was changed on Wednesday morning.}
In a random sample of 400 PSU graduates, 64\% stated that they prefer root beer (over birch beer). Therefore, a 90\% confidence interval for the proportion of all PSU graduates who prefer root beer is:
   (A) \(0.64 \pm 1.64 \times \sqrt{0.64 \times 0.36/400}\)
   (B) \(0.64 \pm 2 \times \sqrt{0.64 \times 0.36/400}\)
   (C) \(0.64 \pm 1.64 \times \sqrt{0.64/400}\)
   (D) \(0.64 \pm 2 \times \sqrt{0.64/400}\)

12. The histogram of sample proportions from a large number of equally-sized random samples will be shaped approximately like a:
   (A) Semi-circle.
   (B) Triangle.
   (C) Skewed curve, with a long right tail.
   (D) Normal (bell-shaped) curve.

13. To calculate a 99\% confidence interval for a population proportion, we use:
   (A) Sample proportion \(\pm 1.64\times\text{S.D.}\)
   (B) Sample proportion \(\pm 2\times\text{S.D.}\)
   (C) Sample proportion \(\pm 2.33\times\text{S.D.}\)
   (D) Sample proportion \(\pm 2.576\times\text{S.D.}\)

14. All other things remaining constant, if the sample size increases by a factor of nine then the confidence interval for the population mean will:
   (A) Become nine times as wide.
   (B) Become one-ninth as wide.
   (C) Triple in width.
   (D) Become one-third as wide.
15. A random sample of 25 farmers was examined in a study of caffeine levels. From the data collected, a 95% confidence interval for the population mean caffeine level was calculated to be 21.5 to 23.0. We can conclude that:

(A) For any random sample of 25 farmers, the resulting sample mean will always fall between 21.5 and 23.0.

(B) 95% of all farmers have caffeine levels between 21.5 and 23.0.

(C) If we repeatedly collect random samples of size 25 and calculate the corresponding confidence intervals then, over the long run, 95% of these intervals will capture the population mean and 5% will fail to capture the population mean. The interval 21.5 to 23.0 is one of these many possible intervals and we cannot be certain that it has captured the population mean.

(D) If we repeatedly sample the entire population then, about 95% of the time, the population mean will fall between 21.5 and 23.0.

(E) None of the above.

16. In a previous question, it was stated that the interval 21.5 to 23.0 is a 95% confidence interval for the population mean. We may conclude that:

(A) 22.3 is a plausible value for the population mean.

(B) 21.8 is not a plausible value for the population mean.

(C) 23.3 is a plausible value for the population mean.

(D) 19.7 is a plausible value for the population mean.

17. All other things remaining constant, increasing the confidence coefficient causes the width of a confidence interval to:

(A) Decrease.

(B) Increase and then decrease.

(C) Remain unchanged.

(D) Increase.

18. All other things remaining constant, if the population size quadruples from 10 million to 40 million then the width of a confidence interval will:

(A) Decrease by a factor of two.

(B) Increase and then decrease.

(C) Remain unchanged.

(D) Increase by a factor of two.
19. A student claims that, for any data set of size two or more, the standard error of the mean (SEM) is smaller than the sample standard deviation. This claim is:
   (A) Always true.
   (B) Always false.
   (C) Not always true; it depends on the actual numbers in the sample.
   (D) Not always false; it depends on the sample size.

20. A confidence interval for a population proportion is a range of numbers which:
   (A) Is certain to contain the population proportion.
   (B) Has a 95% probability of containing the population proportion.
   (C) Increases in width as the sample size increases.
   (D) Is a plausible range of values for the population proportion.

21. All other things remaining constant, if the sample size decreases then the standard error of the sample mean:
   (A) Increases, levels off, and then increases again.
   (B) Decreases and then increases.
   (C) Decreases.
   (D) Increases.
   (E) Will remain unchanged.

22. In a study of farmers’ caffeine levels, a random sample of 25 farmers yielded a sample mean of 22 and a sample standard deviation (S.D.) of 4. Therefore, the standard error of the mean (SEM) is:
   (A) \( \frac{22}{\sqrt{25}} \)
   (B) \( \frac{4}{\sqrt{25}} \)
   (C) \( 22 \times \sqrt{25} \)
   (D) \( 4 \times \sqrt{25} \)

23. All other things remaining constant, if the sample size increases by a factor of 25 then the standard error of the mean:
   (A) Decreases by a factor of 5.
   (B) Increases by a factor of 5.
   (C) Decreases by a factor of 25.
   (D) Increases by a factor of 25.
   (E) Will remain unchanged.
24. All other things remaining constant, which of the following sample proportions will result in the widest confidence interval:

(A) .1
(B) .2
(C) .3
(D) .4
(E) .5

25. An advertiser of No-Pain aspirin claims it is the pain-killer most preferred by consumers. This claim was based on a consumer survey in which the choices were: Lavid, Acinna, No-Pain, and Yellnot. This is an example of:

(A) A closed question.
(B) An open question.
(C) A difficult question.
(D) An easy question.

26. In a study of car ownership in Pennsylvania, the variable “Brand of car owned” is:

(A) A measurement variable.
(B) An ambiguous variable.
(C) A categorical variable.
(D) A discrete variable.

27. The survey question, “Don’t you agree that our tax system, which is too complicated for anyone to understand, should be overhauled?,” is:

(A) Deliberately biased.
(B) Unnecessarily complicated.
(C) Unintentionally biased.
(D) Likely to cause the respondent not to tell the truth.

28. Lee Salk exposed one group of infants to the sound of a heartbeat and compared their weight gain to that of a group not exposed. The greater the amount of natural variability within each group:

(A) The more difficult it will be to detect a significant difference in weight gains between the two groups.
(B) The less difficult it will be to detect a significant difference in weight gains between the two groups.
(C) There will be no change in our ability to detect a significant difference in weight gains between the two groups.
(D) None of the above.
A study of caffeine levels of farmers and doctors reported the following data, which is the subject of the next three questions:

<table>
<thead>
<tr>
<th></th>
<th>Farmers</th>
<th>Doctors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample Size</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>Sample Mean</td>
<td>13</td>
<td>10</td>
</tr>
<tr>
<td>Sample S.D.</td>
<td>4</td>
<td>3</td>
</tr>
</tbody>
</table>

29. The standard error of the difference between the two sample means (also called the “measure of variability”) is:
   (A) The standard deviation of the data in the combined samples.
   (B) \( \sqrt{(\text{SEM}_1)^2 + (\text{SEM}_2)^2} = \sqrt{\left(\frac{4}{\sqrt{25}}\right)^2 + \left(\frac{3}{\sqrt{25}}\right)^2} = 1.0 \)
   (C) \( \sqrt{\frac{(\text{Sample Mean}_1)^2}{\text{SD}_1} + \frac{(\text{Sample Mean}_2)^2}{\text{SD}_2}} = \sqrt{\left(\frac{13}{\sqrt{25}}\right)^2 + \left(\frac{10}{\sqrt{25}}\right)^2} = 3.28 \)
   (D) \( \frac{(\text{Sample Mean}_1 - \text{Sample Mean}_2)}{\sqrt{\text{SD}_1 - \text{SD}_2}} = \frac{(13 - 10)}{\sqrt{4 - 3}} = 3.0 \)

30. A 95% confidence interval for the difference in population mean caffeine levels (farmers’ mean minus doctors’ mean) is:
   (A) \( (4 - 3) \pm 2 \times 3.28 = 1 \pm 6.56. \)
   (B) \( (4 - 3) \pm 2 \times 3.0 = 1 \pm 6. \)
   (C) \( (13 - 10) \pm 1.64 \times 1.0 = 3 \pm 1.64. \)
   (D) \( (13 - 10) \pm 2 \times 1.0 = 3 \pm 2. \)

31. Based on our interpretation of confidence intervals, we infer that the population means of farmers’ and doctors’ caffeine levels:
   (A) Are close to each other; there is no significant difference between them.
   (B) Are significantly different from each other. Indeed, we have strong evidence that the farmers have a higher population mean level than the doctors.
   (C) Are equal to each other; we are 95% confident that they are equal.
   (D) Have no relationship to each other; we have insufficient evidence to make an inference about their relative location.

32. A researcher repeatedly collects random samples of size 1,600 and computes a 95% confidence interval for the population mean using each sample. Over the long run, the proportion of confidence intervals which will fail to capture the population mean is:
   (A) 95%
   (B) 5%
   (C) \( 1/\sqrt{1,600} \), or 1/40, 2.5%
   (D) None of the above.
33. The area below the normal bell-shaped curve and within ±2 S.D. of the center is:
   (A) 1.96
   (B) .5
   (C) .95
   (D) .975

34. When measured with extreme accuracy, the variable “height of a building,” is:
   (A) A discrete variable.
   (B) A categorical variable.
   (C) A continuous variable.
   (D) All of the above.