Some questions have been adapted from *Statistics: The Art and Science of Learning from Data*, 3rd ed, by Agresti & Franklin.

1. **Don’t even open this until you are told to do so.**

2. Remember to put your phone on airplane mode. You may listen to your tunes as long as you do not disturb anyone.

3. Please turn your hats around backwards or take them off.

4. Please put your backpack and other things along the walls or at the front of the room.

5. You need a gray, $8\frac{1}{2} \times 11^\prime$ scantron, pencil, calculator and you may have 5 sheets of notes.

6. There are 20 multiple-choice questions on this exam, each worth 5 points. There is partial credit. Please mark your answers **clearly**. Multiple marks will be counted wrong.

7. You will have 60 minutes to finish this exam.

8. If you have questions, please write out what you are thinking so that we can discuss it after I return it to you.

9. If you are caught cheating or helping someone to cheat on this exam, you both will receive a grade of **zero** on the exam. You must work alone. Do not discuss anything about the exam with the other section until grades are posted.

10. When you are finished please make sure you have filled in your name and marked your FORM (A or B) and 20 answers, and leave everything on your desk. Try to sit on the same area next time so that I can return your materials to you.

11. Good luck!
1. At two different hospitals, boys and girls are equally likely to be born, 50% either. One is an urban hospital, visited each year by about 10 times as many mothers giving birth than the other rural hospital. In any given year, which of the following is most likely to be true? Think of a year as a sample, so the urban hospital has a larger sample size.

A. The rural hospital is more likely to have at least 70% boys born that year than the urban hospital since the rural hospital has more variability.
B. The urban hospital is more likely to have at least 70% boys born that year than the rural hospital since the urban hospital has more variability.
C. The hospitals are equally likely to have at least 70% boys born that year.
D. The rural hospital is more likely to have at least 70% boys born that year than the urban hospital since the rural hospital has less variability.
E. The urban hospital is more likely to have at least 70% boys born that year than the rural hospital since the rural hospital has less variability.

2. What affects the sampling distribution of the sample mean, \( \bar{X} \)?

A. whether the sample is random or not
B. the size of the sample, \( n \)
C. the parent population distribution (the population being sampled)
D. All of the above affect the sampling distribution of \( \bar{X} \).
E. Only two of the above affect the sampling distribution of \( \bar{X} \).

3. There exists at least 100 years of weather data, but suppose we only took a sample of years and calculated a 95% confidence interval for the true yearly average total rainfall of (16.5, 22) inches. Which of the following is best interpretation of this interval?

A. There is a 95% probability that the true yearly average total rainfall is between 16.5 and 22 inches.
B. If we sampled many years, 95% of the yearly totals would be between 16.5 and 22 inches.
C. As long as we sample enough years (at least 30), the probability that the true yearly average total rainfall is between 16.5 and 22 inches will be 0.95.
D. If we took many different samples of years, about 95% of the confidence intervals created from these samples would contain the true yearly average total rainfall.
E. If we took many different samples of years, about 95% of the yearly average totals would equal the true yearly average total rainfall.

4. Referring to the confidence interval in the last problem, (16.5, 22), which of the following statements are plausible for this data (with 95% confidence)?

A. The true yearly average total rainfall could be 20 inches.
B. The true yearly average total rainfall could be 16 inches.
C. The true yearly average total rainfall would NOT be 22 inches.
D. All of the above are plausible.
E. Only two of the above are plausible.

5. Suppose a 90% confidence interval for the true average on this exam is (62.3, 77.7). Which of the following is true?

A. No one flunked (made less than 60), but no one made above a B.
B. The true average of the class is 70.
C. Everyone was within 7.7 of the mean.
D. All of the above are true.
E. None of the above are true.

6. The results of a survey on the belief that stem cell research has merit are shown in the table below.

<table>
<thead>
<tr>
<th>X</th>
<th>N</th>
<th>Sample p</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1522</td>
<td>2104</td>
<td>0.7234</td>
<td>(0.7043, 0.7425)</td>
</tr>
</tbody>
</table>

Here, \( X \) denotes the number who believed that stem cell research has merit. The “Sample p” is the proportion of all respondents in the sample who believe stem cell research has merit. The “95% CI” is the 95% confidence interval for the population proportion. Interpret the margin of error in context of the problem.

A. We are 95% sure that the standard deviation from one sample to another of the sample proportion of respondents who believe stem cell research has merit is 0.0191.
B. We are 95% sure that the sample proportion 0.7234 is no farther from the proportion of everyone who believes stem cell research has merit than 0.0191.
C. The amount of error made in calculations of the confidence interval is 0.0191; therefore we expect the sample proportion 0.7234 to be no more than 0.0191 from the proportion of everyone who believes stem cell research has merit.
D. Some people are biased about stem cell research and won’t answer the question correctly; that’s why we have some error in our sample proportion. Therefore we expect the sample proportion 0.7234 to be no more than 0.0191 from the proportion of everyone who believes stem cell research has merit.
E. None of the above are correct.
7. Assuming that the distribution of grades on this exam is approximately normal \( X \sim N(70, 5.8^2) \), what grade would you need to make to be in the top 25% of the class?

A. 70.25
B. 70 + 0.25 \times 5.8
C. 70 + 0.675 \times 5.8
D. 75
E. 0.25 \times (70 + 5.8)

8. The table below is from Ch 3 homework. One answer claimed there was no apparent association between happiness and income, but does that mean they are independent? Are Below and Very happy independent? Round to the nearest integer percent (no decimals).

<table>
<thead>
<tr>
<th></th>
<th>Not too</th>
<th>Pretty</th>
<th>Very</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Above</td>
<td>59</td>
<td>220</td>
<td>111</td>
<td>390</td>
</tr>
<tr>
<td>Average</td>
<td>7</td>
<td>270</td>
<td>115</td>
<td>459</td>
</tr>
<tr>
<td>Below</td>
<td>71</td>
<td>278</td>
<td>129</td>
<td>478</td>
</tr>
<tr>
<td>Total</td>
<td>204</td>
<td>768</td>
<td>355</td>
<td>1327</td>
</tr>
</tbody>
</table>

A. Yes, because \( 129/1327 = (478/1327) \times (355/1327) \)
B. Yes, because \( 478/1327 = 129/355 \)
C. Yes, because \( 355/1327 = 129/478 \)
D. All of the above are valid proof of independence.
E. None of the above are valid proof of independence.

9. Suppose we have a population of people’s ages with mean \( \mu = 35 \) years and standard deviation \( \sigma = 2.5 \) years. The proportion of people in that population that are 25 years or older is 0.3. If we take a simple random sample of size 40 from the population, what is the distribution of the sample mean age?

A. There is not enough information to decide.
B. Binomial (categorical data distribution) with \( n = 40 \) and \( p = 0.30 \) since the data is highly skewed
C. \( N(0.3, 0.07^2) \)
D. approximately \( N(35, 0.4^2) \) even though the sample could be highly skewed
E. approximately \( N(35, 2.5^2) \)

10. What are the \( z \) critical values, the \( z_{\alpha/2} \), for a 91% confidence interval?

A. \( \pm 0.48 \)
B. \( \pm 1.695 \)
C. \( \pm 0.82 \)
D. \( \pm 1.34 \)
E. \( \pm 0.67 \)

11. If the weights of students at A&M are normally distributed with a mean of 160 and a standard deviation of 25, what is the probability that a randomly selected student will weigh more than 150?

A. 0.6554
B. 0.9536
C. 0.4
D. 0.3446
E. 0.25

12. In one exit poll of \( n = 140 \) voters, 66 said they voted for the Democratic candidate and 74 said they voted for the Republican candidate. A 95% confidence interval for the true proportion in favor of the Democratic candidate is \((0.389, 0.554)\). Does allow you to predict the winner? Why or why not?

A. Yes, because the interval includes a majority of people voting for the Democratic candidate.
B. Yes, because all the values in the interval are positive (greater than 0) and less than 1.
C. No, because the interval includes a majority of people voting for the Democratic candidate and a majority of people voting for the Republican candidate.
D. No, because the values in the interval would be reversed if we define the proportion to be voting for the Republican candidate.
E. No, because this is just one sample and we would need more information to make an accurate prediction.

13. A 95% confidence interval with \( n = 1400 \) voters and counts 660 and 740 would give different results than those above. Explain why.

A. We have a larger margin of error when we have a larger sample size, giving us more precision to estimate the parameter.
B. The larger sample size helps to reduce people’s bias for one candidate or the other.
C. The proportions of people who voted for the Democratic and Republican candidates would be different from those above.
D. The \( z \)-scores in the confidence intervals would be different for this confidence interval from those above.
E. The larger sample size provides more information, so the same confidence level, I have more precision to estimate the parameter.

14. Suppose the IQ of children with Fetal Alcohol Syndrome is normally distributed with true mean \( \mu = 70 \) and true standard deviation \( \sigma = 12 \). What does \( P(X < 50) \), where \( X \) is the IQ’s of these children, actually mean?

A. how likely an average child with FAS would have an IQ of 50 or less
B. how likely any child with FAS would have an IQ of 50 or less
C. how likely a sample of children with FAS would have an average of 50 or less
D. how likely one child out of a sample of children with FAS would have an IQ of 50 or less
E. how likely the true mean IQ of children with FAS is 50 or less
15. The Pew Research Center recently took an online poll of \( n = 1048 \) U.S. drivers and found that 38% of the respondents said that they “shouted, cursed or made gestures to other drivers” in the last year. A 95% confidence interval for the proportion of all U.S. drivers who “shouted, cursed or made gestures to other drivers” in the last year is thus (0.3506, 0.4094). Does the fact that the respondent is self reporting these actions affect the way you interpret this confidence interval? Why or why not?

A. Yes, because this is an example of undercoverage; the Pew Research Center only polled 1048 drivers instead of a larger sample size. With such a tiny percentage of all drivers, we can’t make an inference to all U.S. drivers.

B. Yes, because this is an example of response bias; the sample proportion, 0.38, may be incorrect because people may feel it is socially unacceptable to curse and not admit that they do it.

C. Yes, because this is an example of voluntary response bias; the respondents must volunteer the information that they have “shouted, cursed, or made gestures to other drivers,” so they may lie about their actions.

D. No, because the sample proportion is unbiased for the population proportion, thus there is no bias in this study.

E. No, the Pew Research Center is very reliable.

16. Suppose we have a fairly normal looking distribution of probabilities of getting a date to the next Aggie football game where the average is 40% (on average you get a date 40% of the time) with a standard deviation of 4.9%. So \( P \sim N(0.4, 0.049^2) \). What is the 95th percentile of this distribution?

A. 82.89%

B. 1.96

C. ±2 sd

D. 48%

E. 0.495

17. For the population of people who suffer occasionally from migraine headaches, suppose \( p = 0.30 \) is the true proportion who get some relief from taking a certain medicine. For a particular subject, let \( x = 1 \) if they get relief and \( x = 0 \) if they do not. For a random sample of 60 people who suffer from migraines which of the following is true?

A. The population distribution will have have 18 1’s and 42 0’s.

B. The data (sample) distribution will have have 18 1’s and 42 0’s.

C. The sampling distribution of the sample proportion, \( \hat{p} \) will have 18 1’s and 42 0’s.

D. The sampling distribution of the sample proportion, \( \hat{p} \) will centered at 0.30.

E. None of the above are true.

18. Which of the following best describes the standard deviation of the sampling distribution of the sample proportion in the previous problem?

A. It is equal to the standard deviation of the population distribution.

B. It is equal to the standard deviation of the data distribution.

C. It is equal to the standard deviation of the population distribution divided by the square root of 60.

D. It is the approximate average difference between an observation, one person, and the average of all.

E. It is the approximate average difference between a sample proportion and the true proportion.

19. The weight of a package of mints is believed to be normally distributed with mean 21.37 grams and standard deviation 0.4, \( X \sim N(21.37, 0.4^2) \). What is the chance that a sample of 4 packages of mints has an average weight, \( \bar{X}_4 \), between 21 and 22 grams?

A. 0.0314

B. 0.9032

C. 0.9426

D. 0.9670

E. 0.9938

20. What’s the chance a standard normal random variable is between -2.1 and 1.37?

A. 0.73

B. 0.7673

C. 0.8968

D. 0.9326

E. 0.2327