1. Error consists of what two aspects?

   (a) Sampling Error
   (b) Bias
   (c) Estimate
   (d) A and B
   (e) None of the Above

   Solution: D

2. In a random sample of 400 alumni at UCR, 43 say they use their phone in class. What is the estimated standard error for this point estimate? Round you answer to 4 decimal places.

   Solution: $\hat{p} = 43/400 = 0.1075$

$$\begin{aligned} SE_{\hat{p}} &= \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \\ &= \sqrt{\frac{0.1075(1-0.1075)}{400}} \\ &= \sqrt{2.3985 \times 10^{-4}} \\ &= 0.0155 \end{aligned}$$

3. A random sample of $n = 81$ observations has a mean of 30. Assume the standard deviation is known to be 3.9. Find the standard error of the sample mean and the 90% confidence interval for μ.

   Solution: $\bar{x} = 30$ and $z_{0.1/2} = 1.645$

$$\begin{aligned} SE &= \frac{\sigma}{\sqrt{n}} \\ &= \frac{3.9}{\sqrt{81}} \\ &= 0.4333 \end{aligned}$$

   So the confidence interval is

$$\begin{aligned} &\bar{x} \pm z_{\alpha/2} \times SE \\ \rightarrow\ &30 \pm 1.645 \times 0.4333 \\ \rightarrow\ &30 \pm 0.7128 \end{aligned}$$

   or (29.2872, 30.7128). We can be 90% confident that the true value of μ is between these two values.

4. Alexis and Travis run for their high school's junior varsity boy's cross-country team. There is one space left on the varsity team that the coach is trying to fill, and he has decided to measure both of their performance in the next race they run. Alexis ran a five-kilometer run in 18 minutes and 12 seconds. The average time for a varsity runner who ran that same course was 17 minutes and 6 seconds with a standard deviation of 57 seconds. Travis ran a three-mile run in 18 minutes and 49 seconds. The average time for a varsity runner who ran the same course Travis did was 16 minutes and 57 seconds with a standard deviation of 56.8 seconds.

   (a) What are μ and σ for both cases?
   (b) Calculate the z-score for both cases. (Assume a normal distribution.)
   (c) The coach wants to know who should get the spot on the varsity team. Who should make varsity and why?

   Solution:

(a) For the 5k (Alexis), μ =17 minutes and 6 seconds ($\mu = 1026$ seconds) and $\sigma = 57$ seconds. For the 3 mile (Travis), $\mu = 16$ minutes and 57 seconds ($\mu = 1017$ seconds) and $\sigma = 56.8$ seconds.

(b)

$$z_{Alexis} = \frac{18:12:00 - 17:06:00}{00:57:00} = \frac{1092 - 1026}{57} = 1.16$$

and

$$z_{Travis} = \frac{18:49:00 - 16:57:00}{00:56:08} = \frac{1129 - 1017}{56.8} = 1.97$$

(c) Both Z-scores are positive, so both athletes ran slower than the average varsity runner. Since Alexis's time is closer to the average, she is less slow (faster) than Travis. Therefore Alexis should get the varsity spot.

5. Suppose there was a midterm exam with 50 questions with an average of 35 correct questions and a standard deviation of 6.3 from a group of 75 students.

(a) When you change the mean (increase or decrease), what happens to the distribution model? (ex: shifts)

(b) When you change the standard deviation (increase and decrease), what happens tothe distribution model? (ex: shifts)

(c) How you would write out the notation for a normal distribution given the information above?

(d) Kim got a score of 43/50 on the exam. Calculate her Z-Score. Compare her score to James where his score was 27/50. Where do Kim and James lie in the class of 75 students?

Solution:

(a) When we increase the mean, the distribution shifts to the right. When we decrease the mean, it shifts to the left.

(b) Increasing the standard deviation makes the distribution wider (flatter). Decreasing the sd makes the distribution narrower (peakier).

(c) $N(\mu = 35, \sigma = 6.3)$.

(d)

$$z_{Kim} = \frac{43 - 35}{6.3} = 1.2698$$

and

$$z_{James} = \frac{27 - 35}{6.3} = -1.2698$$

Using the z-scores, Kim is 1.27 standard deviations above the mean. James is 1.27 standard deviations below the mean.

6. For extra credit, the professor illustrates a question for the class to answer. Out of 75 students, around 70% of the students will pass the class. The professor wants to know some probabilities about how many students will pass his course.

(a) What are the conditions under which a binomial distribution is allowed?

(b) Find the probability that exactly 30 students will pass his class.

Solution:

(a) A binomial distribution is used when

- n is a fixed number of trials.
- Trials are independent.
- 2 possible outcomes (pass or fail)
- p is the probability of success

(b) Let X be the number of students who pass (out of the 75 total students).

$$
\begin{aligned}
P(X=k) &= \binom{n}{k} p^k (1-p)^{n-k} \\
P(X=30) &= \binom{75}{30} 0.7^{30} \times 0.3^{45} \\
&= 5.2064 \times 10^{-8}
\end{aligned}
$$

7. During a test, an average of 6.6 students leave the test room to go to the bathroom every 10 minutes.

   (a) Calculate the probability that only 3 students leave the class to go to the bathroom.

   (b) Calculate the probability that 0, 2, and 5 students leave the class to go to the bathroom.

   Solution:

   (a) Let X be the number of students who leave over the course of 10 minutes. $\lambda = 6.6$

$$
\begin{aligned}
P(X=k) &= \frac{\lambda^k e^{-\lambda}}{k!} \\
P(X=3) &= \frac{6.6^3 e^{-6.6}}{3!} \\
&= 0.0652
\end{aligned}
$$

   (b) For 0, 2, or 5 students

$$
\begin{aligned}
P(X=0,\ 2,\ \text{or } 5) &= P(X=0) + P(X=2) + P(X=5) \\
&= \frac{6.6^0 e^{-6.6}}{0!} + \frac{6.6^2 e^{-6.6}}{2!} + \frac{6.6^5 e^{-6.6}}{5!} \\
&= 0.0014 + 0.0296 + 0.1420 \\
&= 0.1730
\end{aligned}
$$

8. After a random sampling of 3000 students at UCR, around 34% of students agree that the school needs better roads for easier transportation.

   (a) Does this question follow the Central Limit Theorem? Explain your reasoning.

   (b) Find a 95% confidence interval for p. Note that $z_{0.05/2} = 1.96$

   Solution:

   (a) The samples are assumed independent. Success-failure condition:

$$np = 3000 \times 0.34 = 1020 \geq 10$$

   and

$$n(1-p) = 3000 \times 0.64 = 1980 \geq 10$$

   so the Central Limit Theorem applies.

   (b) A 95% CI is

$$
\begin{aligned}
&\hat{p} \pm z_{\alpha/2} \times SE \\
&0.34 \pm 1.96 \times \sqrt{\frac{0.34 \times 0.66}{3000}} \\
&0.34 \pm 1.96 \times 0.00865 \\
&0.34 \pm 0.016954
\end{aligned}
$$

   or the interval (0.3230, 0.3570).

9. A population within the UCR Soccer Club has around 73% of players randomly interviewed saying they need a new coach. After some statistical analysis, an analyst found the standard error to be around 0.036968. The analyst used a 95% confidence interval to estimate p. Find n, lower and upper bound, and whether or not this problem follows the Central Limit Theorem.

   Solution:

$$SE_{\hat{p}} = 0.036968 = \sqrt{\frac{0.73(1-0.73)}{n}}$$

   so

$$n = \frac{0.73(1-0.73)}{0.036968^2} = 144.2231$$

   with rounding error, we expect that about 144 or 145 players were interviewed.

   The bounds are

$$\hat{p} \pm z_{\alpha/2} SE$$
$$0.73 \pm 1.96 \times 0.036968$$

   or 0.6575 (lower) and 0.8024 (upper).

   To confirm the CLT is appropriate, the players were interviewed at random and

$$np \approx 144 \times 0.73 = 105.12 \geq 10$$
$$n(1-p) \approx 144 \times 0.26 = 51.84 \geq 10$$

   so the CLT applies.

10. Average daily high temperature in August in Riverside is 95 F with a standard deviation of 4 F. Suppose that the temperatures in August closely follow a normal distribution.

   (a) What is the probability of observing a 101 F temperature or higher in Riverside during a randomly chosen day in August?

   (b) How cool are the coldest 20% of the days (days with lowest average high temperature) during August in Riverside?

   Solution: Let X = temperature in August. $X \sim N(\mu = 95, \sigma = 4)$

   (a) To find $P(X \geq 101)$,

$$z_{101} = \frac{101 - 94}{4} = 1.5$$

   then

$$P(Z \geq 1.5) = 1 - P(Z < 1.5) = 1 - 0.9332 = 0.0668$$

   (b) To find the coldest 20% of days,

$$P(X < x) = P(Z < z) = 0.2$$

   Using `R` or a table, $z = -0.8416$ so

$$z = \frac{x - \mu}{\sigma} \to x = z\sigma + \mu = -0.8416 \times 4 + 95 = 91.6336$$

   So the coldest 20% of days will have temperatures no higher than 91.6 degrees.

11. There are 600 students enrolled in a high school, but only 300 of the students plan to graduate.

   (a) What is the point estimate for the proportion of students planning to graduate?

   (b) Calculate the standard error (SE), Margin of Error(MoE), and upper and lower bounds of the interval, if we have a 95% confidence.

Solution

(a) $\hat{p} = 0.5$

(b)

$$SE_{\hat{p}} = \sqrt{\frac{0.5 * 0.5}{600}} = 0.0204$$

then

$$MoE = z_{\alpha/2} \times SE = 1.96 * .0204 = .0400$$

so the upper bound is $50 + .0400 = .540$ and the lower bound is $.50 - .0400 = .460$.

12. A gym junkie is told that there's a new protein powder that helps you build bigger muscles, repair tissues faster, and make a lot of enzymes and hormones. He is hesitant to try the new powder because he did not believe that the protein could be so effective. However, he caves in and after a couple of weeks on the supplement he thinks it may be working because he feels a major change in his body.

    (a) Write null and alternative hypotheses.

    (b) Identify the type I and type II errors.

    Solution:

    (a) H_0: the new protein powder has no effect on this person's body
    H_A: the new protein powder does have an effect on this person's body

    (b) Type I: the new protein powder is found to have an effect when in fact is does not have one.
    Type II: the new protein powder is not found to have any effects when it fact is does have effects.

13. An artist bought a new palette with seven spots for all the colors in the rainbow: red, orange, yellow, green, blue, indigo and violet. How many ways can she order these colors on her paint palette?

    Solution: $7! = 5040$

14. In FRIENDS, Joey Tribbiani says his famous line, "how you doing" about 18 times in 18 different episodes out of 236 episodes. The proportion of Joey saying his lines is 7.6%, what is the standard error? Round to three decimals.

    Solution: $\sqrt{\frac{.076(1-.076)}{236}} = .017$

15. A survey result says that 45% of Americans have a work commute longer than 1 hour. What's the probability that exactly 3 of 10 randomly sampled Americans have a commute longer that 1 hour?

    Solution: A random sample suggests independence. We will use a binomial distribution. $n = 10$ and $p = 0.45$. Let X be the number of people with commutes longer than 1 hours.

$$\begin{aligned} P(X = 3) &= \binom{10}{3} 0.45^3 (1 - 0.45)^{10-3} \\ &= 0.1664 \end{aligned}$$

16. A fair die has 6 sides. Suppose you roll the die 10 times.

    (a) What is the probability of rolling exactly three 6s?

    (b) What is the probability of rolling no more than two 6s?

    (c) What is the probability of rolling at least four 6s?

    Solution: Let X be the number of sixes rolled. (6 is a success, everything else is a failure.)

    (a)

$$P(X = 3) = \binom{10}{3} \times (0.167)^3 (1 - 0.167)^7 = 0.156$$

(b)

$$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 3) = 0.775$$

(c)

$$P(X \geq 4) = 1 - P(X \leq 3) = 1 - 0.156 - 0.775 = 0.0697$$

17. When is it useful to apply the Poisson distribution?

    Solution: When you are estimating the number of events in a large population over a unit of time.

18. Of all the girl scouts at the OC Girl Scouts, 15% received a gold star award. The girl scout leaders randomly sample 50 girl scouts and check to see if the girl scouts made the list. They repeated this 1000 times and built a distribution of sample proportions.

    (a) What is this distribution called?
    (b) What is the population under consideration in this data set?
    (c) What parameter is being estimated?
    (d) Calculate the variability of this distribution.
    (e) What is the formal name of the value you computed in (d)?

    Solution:

    (a) Sampling distribution
    (b) The girl scouts at OC Girls Scouts
    (c) The proportion of girl scouts who received the gold star award
    (d) SE = sqrt((p*(1-p))/n) = 0.0505
    (e) Standard Error

19. According to a study, 40% of U.S. college students drink coffee. Assume this value only represents a sample, but not necessarily the population of interest. Suppose the study reported a standard error of about 1.5%, and a normal model is used to represent the data. Find the margin of error and create a 95% confidence interval. Interpret the interval.

    Solution: Margin of error:

    $$MoE = z_{\alpha/2} \times SE = 1.96 \times .015 = 0.0294$$

    95% confidence interval:

    $$0.40 \pm 1.96 \times .015$$

    or (0.3706, 0.4294).

    We can be 95% confident that the proportion of U.S. college students who drink coffee falls within 0.3706 and 0.4294 confidence interval.

20. A manufacturer is performing a structural-quality test on toothpicks. A device applies force to a toothpick until it cracks. If a toothpick cracks only after exceeding a predetermined threshold of force, it is deemed to be of acceptable quality. From the manufacturer's total stock of toothpicks, 5000 were randomly selected for testing, 79.4% of which were deemed acceptable. Construct a 95% confidence interval for the population proportion of acceptable toothpicks and interpret the result. (Use —z— = 1.96).

    Solution: The 95% confidence interval is

    $$0.794 \pm 1.96 \times \sqrt{\frac{0.794 \times (1 - 0.794)}{5000}}$$

    or 0.794 ± 0.0012. As an interval, $(0.7828, 0.8052)$.

    One can be 95% confident that the true proportion of acceptable toothpicks is between 0.78 and 0.81.

21. A research survey on Facebook adult users within the U.S. found that 51% use the platform to receive at least some news. The standard error for this estimate was 3.4%. Determine whether the following statements are true or false and explain your answer.

   (a) We can say that 98% of all U.S. adult Facebook users were included in the study, since the standard error is 3.4%.
   (b) If we collect less data, the standard error of estimate will decrease.

   Solution:

   (a) False! Standard error does not have anything to do with the proportion of the sample to the population.
   (b) False! As the standard error inversely proportional to the square root of n, we should collect more data for it to decrease.

22. Joselyn took the Statistics 100A test on Friday morning. She scored 89 on the multiple choice section and 74 on the reasoning section. The average score for the multiple choice section was 72 with a standard deviation of 3.5, and the average score for the reasoning section was 67 with a standard deviation of 4.21 What is Joselyn's Z-score on the multiple choice section? On the reasoning section? On which section did she score better?

   Solution: Multiple choice
$$z_{mult} = \frac{89 - 72}{3.5} = 4.86$$
   Reasoning
$$z_{reas} = \frac{74 - 67}{4.21} = 1.66$$
   So she did better on the multiple choice section.

23. A high school teacher wants to see the proportion of high school students who do not like school. Past research has demonstrated that about 73% do not like school. The high school teacher plans to use a 99% confidence level and wants a margin of error to be no more than 4%. How many high school students should he ask?

   Solution:
$$n \geq \left(\frac{z_{\alpha/2} \times sd}{MoE}\right)^2 = \left(\frac{2.576}{0.04}\right)^2 \times 0.73 \times (1 - 0.73) = 817.444656$$
   so he should ask 818 students.

24. Mary's goal is to be at least at the 70th percentile for the SATs. Mary scored a 1230 on her SAT and wanted to know if she had reached her goal. Given that the distribution of SAT scores is normal with a mean of SAT scores is 1100 and standard deviation of 200. What percentile did Mary get, and did she achieve her goal?

   Solution:
$$z = \frac{1230 - 1100}{200} = 0.65$$
   and
$$P(Z < 0.65) = 0.74$$
   So Mary's score is at the 74th percentile and she achieved her goal.

25. Suppose the probability of winning a game of UNO is 60% Find the probability of winning 7 matches out of 10.

   Solution: Assume UNO matches are independent. Then we will use a binomial distribution with $n = 10$ and $p = 0.6$.
$$\binom{10}{7} 0.6^7 0.4^{10-7} = 2.1496$$

26. Kelly has been dealing with acne since she was 13 years old. A friend of hers came over one day and brought her a new topical acne medication called ProAct. Kelly has tried every acne medication out there, so she's skeptical of the effects of this new acne treatment. Kelly's friend convinced her to try it for a month, and Kelly actually starts to see her face Start to clear up.

    (a) What are the null and alternative hypothesis for Kelly's situation?

    (b) What is a Type 1 error?

    (c) Type 2 error?

    Solution:

    (a) H_0: The ProAct acne medication has no effect on Kelly's acne condition.
    H_A: The ProAct Acne medication will either worsen or improve Kelly's Acne

    (b) Type I error (false rejection) is concluding that the medication has an effecg when in fact it does not.

    (c) Type II error is failing to conclude that the medication has an effect when in fact it does.

27. Systolic blood pressure is the amount of pressure that blood exerts on blood vessels while the heart is beating. The mean systolic blood pressure for all adults aged 18 and over in the U.S. is reported to be 122 millimeters of mercury (mmHg) with a standard deviation of 15 mmHg. In a random sample of 1000 American adults aged 18 and over, 67.8% of the adults in the sample population had a systolic blood pressure that was higher than the reported mean of 122 mmHg. Find a 95% confidence interval for the sample proportion.

    Solution:

    $$0.678 \pm 1.96 \times \sqrt{\frac{0.678 \times (1 - 0.678)}{1000}} = 0.678 \pm 1.96 \times 0.0148$$

    We can be 95% confident that 64.9% to 70.7% of American adults aged 18 and over has a blood pressure higher than the reported mean of 122 mmHg.

28. We wish to assess the average student height. Say that we take 500 randomly sampled UCR students and ask them of their height. Approximately 268 students reported that they were 5'4. Calculate a 95% confidence interval and explain it in the context of the problem.

    Solution: We will consider the proportion of UCR students who have a height of 5'4. The sample proportion is $\hat{p} = 268/500 = 0.536$.

    $$SE = \sqrt{\frac{0.536 * (1 - 0.536)}{500}} = 0.022$$

    so the CI is

    $$0.536 \pm 1.96 \times 0.022$$

    The 95% confidence interval is (.4923, .5797). We can be 95% confident that between .4923 and .5979 of students at UCR have a height of 5'4.

29. Monsanto is an agricultural company that specializes in producing GMO crop seeds. The company hires inspectors to perform routine quality control on germinating apples derived from their GMO seeds. An inspector performs an inspection on a random sample of 200 apples and found that 28 had genetic defects. Compute a 95% confidence interval and interpret it in the context of this problem.

    Solution: $\hat{p} = 28/200 = 0.14$

    $$0.14 \pm 1.94\sqrt{\frac{0.14 \times 0.86}{200}}$$

    or (0.0919, 0.1881). We can be 95% confident that between 9.19% and 18.81% of the apples have genetic defects.

30. Assume the standard deviation of the weights among newborn girls is 3.5 kg. How manynewborn girls need to be sampled if we want to be 90% sure that the population mean height is estimated correctly to within 0.5 kg?

    Solution:

    $$n \geq \left(z_{\alpha/2} \times \frac{\sigma}{MoE}\right)^2 = \left(1.645 \times \frac{3.5}{0.5}\right)^2 = 132.595225$$

    or 133 newborns.

31. The margin of error will *decrease* when (select all that apply)

    (a) n increases

    (b) $1 - \alpha$ decreases

    (c) $\alpha/2$ increases

    (d) $z_{\alpha/2}$ increases

    Solution: a, b, and c are all true.

32. Is the following statement about hypothesis testing true or false? If false, correct the statement and provide a brief explanation.

    If you cannot reject the null hypothesis, you must assume it to be true.

    Solution: False. Failing to reject the null hypothesis is different from concluding the null hypothesis. Explanation: It is possible that there is a difference that was undetectable with a single study.

33. A student playing Dungeons and Dragons rolls a 20-sided die 3 times. Each side has an equal chance of occurring. What is the probability he rolls 3 20s in a row?

    Solution: Let X be the number of successes in 3 rolls.

    $$P(X = 3) = \binom{3}{3} 0.05^3 (1 - 0.05)^0 = 0.000125$$

34. How does sample size correlate to Type 1 Error and Type 2 Error in hypothesis testing?

    Solution: Sample size correlates to Type 1 Error and Type 2 Error in hypothesis testing because increasing data will reduce Type 2 Error but increase Type 1 Error.

    Note: There is always a trade-off when it comes to Type 1 Error and Type 2 Error because they are inversely proportional; therefore, when one decreases, the other will increase.

35. I surveyed 100 students to see how many approved of having pets in dorms. 42 percent agreed that having pets in the dorms is a good thing.

    (a) What is the sample proportion?

    (b) What is the null hypothesis?

    (c) If I were to repeat survey multiple times and then find out the statistics behind the number of success, what would I use to do this mathematically?

    Solution:

    (a) sample proportion ($\hat{p}$) is the number of success; therefore, it is 0.42.

    (b) The null hypothesis is the default hypothesis; therefore, the null hypothesis is that less than or equal to half of the participants will approve of having pets in the dorms.

    (c) Given that it is repeated over and over again, and we are looking for the number of success only (and that the variable is independent), we will want to use a binomial distribution when solving for this mathematically.

36. For the following questions, state whether the parameter of interest is a mean or proportion and explain why.

   (a) In a survey, 3,000 UCR students were asked if they had purchased coffee or not in the past three days.
   (b) In the latest US census, all Americans were asked if they had paid their taxes on time.
   (c) In an online poll, redditors were asked how many times have they made posts within the past week.

   (a) Proportion because UCR students are responding with a yes or no (proportion of UCR students who purchased coffee).
   (b) Proportion because Americans would respond with a yes or no (proportion of Americans who paid taxes on time).
   (c) Mean because the redditors are reporting a numerical value, how many posts they've made within the past week (mean number of posts).