Worksheet 10 Solution

Fred Azizi

2023-11-28

1. Michael is running for president. The proportion of voters who favor Michael is 0.8. A random sample of 100 voters is taken.

a. What is the distribution of the sample proportion $\hat{p}$?

Answer:

$$p = 0.8 \quad n = 100 \rightarrow n \times p = 80, \quad n(1-p) = 20 \geqslant 5$$

Therefore $\hat{p}$ is Normal. $\hat{\mu}_{\hat{p}} = p = 0.8$

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.8 \times 0.2}{100}} = 0.04 \quad \therefore \hat{p} \sim N\left(0.8, 0.04^2\right)$$

b. What is the probability that the number of voters in the sample who will vote for Michael will be between 80 and 90 ?

Answer:

Let X be the number of voters in the sample voting for Michael

$$\begin{aligned} P[80 < X < 90] &= P\left[\frac{80}{100} < \frac{X}{100} < \frac{90}{100}\right] \\ &= P[0.8 < \hat{p} < 0.9] \\ &= P\left[\frac{0.8 - 0.8}{0.04} < \frac{\hat{p} - 0.8}{0.04} < \frac{0.9.0.8}{0.04}\right] \\ &= P[0 < z < 2.5] \\ &= P[z < 2.5] - P[z < 0] \\ &= 0.9938 - 0.5 \\ &= 0.4938 \end{aligned}$$

c. What is the probability that the number of voters in the sample who will not favor Michael will be more than 16?

Answer:

If we define X as what we had in part b, then we are looking for the probability of 100-X>16.

$$\begin{aligned} P[100 - X > 16] &= p[100 - 16 > X] \\ &= P[X < 84] \\ &= P\left[\frac{X}{100} < \frac{84}{100}\right] \\ &= P[\hat{p} < 0.84] \\ &= P\left[\frac{\hat{p} - 0.8}{0.04} < \frac{0.84 - 0.8}{0.04}\right] \\ &= P[Z < 1] \\ &= 0.8413 \end{aligned}$$

2. The chairman of the Biology department in a certain college believes that 70% of the department's graduate internships are given to international students. A random sample of 50 graduate interns is taken.

a. What is the distribution of the sample proportion?

Answer:

$$\begin{aligned} p = 0.7 \quad n &= 50 \\ \rightarrow np &= 35 \geq 5 \\ n(1-p) &= 15 \geq 5 \end{aligned}$$

Therefore, we can use Central Limit Theorem.

$$\mu_{\hat{p}} = p = 0.7 \quad \sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.7 \times 0.3}{50}} = 0.065$$
$$\therefore \hat{p} \sim N\left(0.7, 0.065^2\right)$$

b. What is the probability that the sample proportion $\hat{p}$ is between 0.65 and 0.73 ?

Answer:

$$\begin{aligned} P[0.65 < \hat{p} < 0.73] &= P\left[\frac{0.65 - 0.7}{0.065} < Z < \frac{0.73 - 0.7}{0.065}\right] \\ &= P[-0.77 < Z < 0.46] \\ &= P[Z < 0.46] - P[Z < -0.77] \\ &= 0.6772 - 0.2206 \\ &= 0.4566 \end{aligned}$$

c. What is the probability that the sample proportion $\hat{p}$ is within $\pm .05$ of the population proportion p?

Answer:

$$\begin{aligned} P[\hat{p} \text{ is between } p \pm 0.05] &= P[p - 0.05 < \hat{p} < p + 0.05] \\ &= P[0.7 - 0.05 < \hat{p} < 0.7 + 0.05] \\ &= P[0.65 < \hat{p} < 0.75] \\ &= P\left[\frac{0.65 - 0.7}{0.065} < Z < \frac{0.75 - 0.7}{0.065}\right] \\ &= P[-0.77 < z < 0.77] \\ &= P[z < 0.77] - P[z < -0.77] \\ &= 0.7794 - 0.2206 \\ &= 0.5588 \end{aligned}$$

3. A professor of statistics noticed that the marks in his course are normally distributed. He has also noticed that his morning classes average 73%, with a standard deviation of 12% on their final exams. His afternoon classes average 77%, with a standard deviation of 10%. What is the probability that the mean mark of four randomly selected students from a morning class is greater than the average mark of four randomly selected students from an afternoon class?

Answer:

Since both populations are Normally distributed, we don't care that sample sizes are less than 30.

We want $P(\bar{X}_{\text{Morning}} > \bar{X}_{\text{Afternoon}})$ or equivalently $P(\bar{X}_{\text{Morning}} - \bar{X}_{\text{Afternoon}} > 0)$. We know that $\bar{X}_{\text{Morning}} - \bar{X}_{\text{Afternoon}} \sim N(.73 - .77 = -0.04, \sqrt{\frac{.12^2}{4} + \frac{.1^2}{4}} = .078)$.

$$\begin{aligned} P(\bar{X}_{\text{Morning}} - \bar{X}_{\text{Afternoon}} > 0) &= P(Z > \frac{0 - (-0.04)}{.078}) \\ &= P(Z > 0.51) \\ &= 0.305 \end{aligned}$$

4. In 200 tosses of a fair coin:

a. What is the expected value and standard deviation of number of heads?

Answer:

Let X be the number of heads in 200 tosses of a fair coin. By properties of Binomial distribution, $E(X) = np = 200 \times 0.5 = 100$ and $sd(X) = \sqrt{np(1-p)} = \sqrt{200 \times 0.5 \times 0.5} = 7.07$.

b. Use the normal distribution approximation to find the probability of exactly 110 heads.

Answer:

To use Normal approximation, we need $Y \sim N(100, 7.07)$. Since $P(Y = 110)$ is zero, we need to use continuity correction factor. Hence:

$$\begin{aligned} P(X = 110) &\approx P(110 - 0.5 \leq Y \leq 110 + 0.5) \\ &= P(109.5 \leq Y \leq 110.5) \\ &= P(Y \leq 110.5) - P(Y \leq 109.5) \\ &= P\left(Z \leq \frac{110.5 - 100}{7.07}\right) - P\left(Z \leq \frac{109.5 - 100}{7.07}\right) \\ &= P(Z \leq 1.49) - P(Z \leq 1.34) \\ &= 0.9318 - 0.9099 = 0.0219 \end{aligned} \tag{1}$$

c. Use formula of binomial probability to compare the results.

Answer:

Exact probability can be calculated as:

$$P(X = 110) = \binom{200}{110}(0.5)^{110}(1-0.5)^{200-110} = 0.021$$

d. What is the probability that we have less than or equal to 95 heads? (Use continuity correction factor)

Answer:

$$\begin{aligned} P(X \leq 95) &\approx P(Y < 95 + 0.5) && \text{(Using continuity correction factor)} \\ &= P(Z < \frac{95.5 - 100}{7.07}) \\ &= P(Z < -0.64) = 0.2611 \end{aligned}$$

Fun fact! using a modern programming languages, $P(X \leq 95) = 0.2623112$. Therefore, our approximation is within 2 decimal places of the correct answer.

5. Suppose that the amount of time teenagers spend weekly working at part-time jobs is normally distributed with a mean of 300 minutes and standard deviation of 40 minutes. Suppose that we sampled this population with a sample size of n and the average of the sample is $\bar{X}_n = 360$. Construct confidence intervals for the population mean with the following confidence levels and sample sizes:

   a. Does the CI become larger or smaller as the confidence level increases?

   Answer:

   We first need to figure out the $z_{\alpha/2}$. For 90% CI, $z_{\alpha/2} = 1.645$. For 95% CI, $z_{\alpha/2} = 1.96$. For 99% CI, $z_{\alpha/2} = 2.575$.

   The confidence interval for μ can be constructed as $\bar{x}_n \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$. Therefore, for $n = 9$ and 90% confidence level:

$$\bar{x}_n \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}} = 360 \pm 1.645 \times \frac{36}{\sqrt{9}} = 360 \pm 19.74 = (340.26,\ 379.74)$$

   Using similar calculation, we get:

Confidence Level	$n = 9$	$n = 25$
90%	(340.26, 379.74)	(348.156, 371.844)
95%	(336.48, 383.52)	(345.888, 374.112)
99%	(329.1, 390.9)	(341.46, 378.54)

   b. Does the CI become larger or smaller as the confidence level increases?

   Answer:

   It becomes wider (larger). It makes sense since in the formula, $z_{\alpha/2}$ is in the numerator, therefore as $z_{\alpha/2}$ gets bigger $z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$ becomes bigger

   c. Does the CI become larger or smaller as the sample size increases?

   Answer:

   It shrinks (becomes smaller). It makes sense since in the formula, n is in the denominator, therefore as n gets bigger $\frac{\sigma}{\sqrt{n}}$ becomes smaller.

   d. With fixed confidence level and sample size, would the CI become larger/smaller/not change if the sample mean were smaller than 360?

   Answer:

It won't change since the width of the confidence interval depends only on confidence level and sample size. Therefore, the confidence interval will only **shift** when we change the sample mean.

6. A sample of 121 cans of coffee showed an average weight of 16 ounces and a standard deviation of 1 ounces. Find an 80% and a 98% confidence interval for the population mean.

   Answer:

   Since σ is unknown, We must use t-distribution. $df = n - 1 = 121 - 1 = 120$. We note that for a 80% CI, we need to find $t_{0.2/2} = t_{0.1,df=120} = 1.289$ and for 98% CI, we need to find $t_{0.02/2} = t_{0.01,df=120} = 2.358$.

   The 80% CI: $\bar{x} \pm t_{\alpha/2}\frac{s}{\sqrt{n}} = 16 \pm 1.289 \times \frac{1}{\sqrt{121}} = (15.88, 16.12)$.

   The 98% CI: $\bar{x} \pm t_{\alpha/2}\frac{s}{\sqrt{n}} = 16 \pm 2.358 \times \frac{1}{\sqrt{121}} = (15.79, 16.21)$.

7. Among 81 individuals sampled from the population, 24 smokers were observed.

   a. Develop the 90%Cl for the population proportion.

      Answer:

      $\hat{p} = \frac{24}{81} = 0.2963$. Therefore, $E = z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 1.645 \times 0.0507 = 0.0835$

      Now we can find the 90 % CI as $\hat{p} \pm E = 0.2963 \pm 0.0835 = (0.2128, 0.3798)$

   b. If now you have a new sample of 150 individuals, determine an interval for the number of smokers based on your answer from question 3.

      Answer:

      If we know the true proportion $\frac{x}{N}$ of smokers is between $(0.2128, 0.3798)$ then number of smokers among 150 people is 150× (0.2128, 0.3798)= (31.92, 56.97). But since we want the **number** of smokers (i.e. an integer), this should be an integer. We need to round it to (31, 57).