Note # 7: More Sampling Distributions and Statistical Inference

Problem 1. Sara wants to determine the average height of all students at Wittenberg University. In this context, what type of distribution is described in each of the scenarios below?

a. Sara surveys 40 students at Wittenberg University and asks them their height.

b. Sara and her 20 classmates each survey 40 students at Wittenberg University and ask them their height. Each person calculates the average of their sample.

c. Sara surveys all the students at Wittenberg University and asks them their height.

Solutions:

a. Data distribution.

b. Sampling distribution.

c. Population distribution.

Copyright © 2022 Aburweis

Problem 2. In a certain year, according to the National Census Bureau, the number of people living in a household had a mean of 4.2 and a standard deviation of 1.9. This is based on census information for the population. Suppose the Census Bureau instead had estimated this mean using a sample of 100 homes. Suppose the sample had a mean of 4.8 and a standard deviation of 1.5.

- **a.** What is the mean of the population distribution?
- **b.** What is the standard deviation of the population distribution?
- **c.** What is the mean of the data distribution?
- **d.** What is the standard deviation of the data distribution?
- **e.** What is the mean of the sampling distribution?
- **f.** What is the standard deviation of the sampling distribution?

Solution:

- **a.** $\mu = \mathbf{4.2}$
- **b.** $\sigma = \mathbf{1.9}$
- **c.** $\bar{x} = \mathbf{4.8}$
- **d.** $s = \mathbf{1.5}$
- **e.** $Mean = \mu = \mathbf{4.2}$
- **f.** Standard Error= SE $= \frac{\sigma}{\sqrt{n}} = \frac{1.9}{\sqrt{100}} = \frac{1.9}{10} = \mathbf{0.19}$

Problem 3. The National Center for Health Statistics reports that the systolic blood pressure for all males 35-44 years old has a mean of 122 and a standard deviation of 12. The medical director of a very large company looks at the medical records of 100 randomly selected male executives in this age group and finds that the mean systolic blood pressure in this sample is $\bar{x} = 128.4$, with a standard deviation of 8.

a. What is the mean of the sampling distribution of $\bar{x}$?

b. What is the standard deviation of the sampling distribution of $\bar{x}$?

c. What is the shape of the sampling distribution of $\bar{x}$?

Solution:

a. $Mean = \mu = \mathbf{122}$

b. Standard Error= SE $= \frac{\sigma}{\sqrt{n}} = \frac{12}{\sqrt{100}} = \frac{12}{\mathbf{10}} = \mathbf{1.2}$

c. $n = 100 \rightarrow large\ sample\ \ \rightarrow sampling\ distribution\ \ \rightarrow approximately\ normal.$

Copyright © 2022 Aburweis

Problem 4. A person's blood pressure is monitored by taking 5 readings daily. The probability distribution of his readings has a mean of 130 and a standard deviation of 6. Suppose the population distribution of his blood pressure readings is normal.

a. What is the shape of the sampling distribution of $\bar{x}$ for a sample size of 5?

b. What is the probability that the sample mean exceeds 135 for a sample size of 5?

Solution:

$X = blood\ pressure\ readings.$

$X \sim N(\mu = 130\,, \sigma = 6)$

a. $\boldsymbol{\bar{x} \sim N(mean = \mu = 130}$, Stdev $= \frac{\sigma}{\sqrt{n}} = \frac{6}{\sqrt{5}} = \mathbf{2.6833})$

b. $P(\bar{x} > 135) = P\left(Z > \frac{135-130}{2.6833}\right) = P(Z > 1.86) = 1 - P(Z < 1.86) = 1 - 0.9686 = \mathbf{0.0314}$

Copyright © 2022 Aburweis

Problem 5. Preschool aged children are on average 40 inches tall with a standard deviation of 3.1 inches.

- **a.** What is the probability that a randomly chosen preschool aged child is less than 37 inches tall?
- **b.** Describe the sampling distribution of the mean height of 36 randomly chosen preschool aged children.
- **c.** What is the probability that the mean height of 36 preschool aged children is less than 37 inches?

Solution:

$X =$ Height of a preschool aged child.

$X \sim N(\mu = 40, \sigma = 3.1)$

a. $P(X < 37) = P\left(Z < \frac{37-40}{3.1}\right) = P\left(Z < \frac{-3}{3.1}\right) = P(Z < -0.97) = \mathbf{0.1660}$

b. $\bar{x} = average$ height of 36 preschool aged children.

$\boldsymbol{\bar{x} \sim N(mean = \mu = 40}$, Stdev $= \frac{\sigma}{\sqrt{n}} = \frac{3.1}{\sqrt{36}} = \mathbf{0.5167})$

c. $\boldsymbol{\bar{x} \sim N(mean = \mu = 40}$, Stdev $= \mathbf{0.5167})$

$P(\bar{x} < 37) = P\left(Z < \frac{37-40}{0.5167}\right) = P(Z < -5.81) = \mathbf{0.0002} \approx \mathbf{0}$

Problem 6. Jenna wants to know what the true average price of a gallon of gasoline in her city is, so she takes a random sample and creates a 95% confidence interval. The interval from her study is ($2.47, $2.85).

- **a.** Based on her interval, is it plausible or possible that the true average price of a gallon of gasoline in her city is $2.99?
- **b.** Based on her interval, is it plausible or possible that the true average price of a gallon of gasoline in her city is $2.59?
- **c.** Based on her interval, is it plausible or possible that the true average price of a gallon of gasoline in her city is -$2.09?

Solution:

- **a.** It is possible, but not plausible.
- **b.** It is both possible and plausible.
- **c.** It is neither possible nor plausible.

Copyright © 2022 Aburweis

Problem 7. Your local school board wants to determine the proportion of people who plan on voting for the school levy in the upcoming election. They conduct a random phone poll, where they contact 150 individuals and ask them whether or not they plan on voting for the levy. Of these 150 respondents, 78 people say they plan on voting for the levy. The school board wants to determine whether or not the data supports the idea that more than 50% of people plan on voting for the levy. **Create a 98% confidence interval for the true proportion of all people who plan on voting for the levy.**

a. What is the 98% confidence interval?
b. What is the correct interpretation of the confidence interval?
c. Are the assumptions met? Explain.
d. Based on the confidence interval, what can you say about the school board's question?

Solution:

a. $CI: \hat{p} \pm Z^* \sqrt{\frac{\hat{p}\,(1-\hat{p})}{n}} = 0.52 \pm (2.326)\sqrt{\frac{0.52\,(0.48)}{150}} = 0.52 \pm (2.326)(0.0408)$
$= 0.52 \pm 0.0949 \rightarrow \mathbf{98\%}\ \boldsymbol{CI}: (\mathbf{0.4251, 0.6149})$

b. We are 98% confident that the true proportion of all people who plan on voting for the levy is between 0.4251 and 0.6149.

c.

i. **Independence:**
- Random
- $n < 10\%\ of\ the\ population\ ?\ unkown$

ii. **Sample size:**

$n\hat{p} \geq 10 \quad \rightarrow \quad (150)(0.52) = \mathbf{78 \geq 10}$
$n(1-\hat{p}) \geq 10 \quad \rightarrow \quad (150)(0.48) = \mathbf{72 \geq 10}$

d. 98% CI: (0.4251,0.6149). This interval includes: values < 0.5, value = 0.5, and values > 0.5. we don't know if more than 50% will vote for the levy.

Problem 8. Your local school board wants to determine the proportion of people who plan on voting for the school levy in the upcoming election. They conduct a random phone poll, where they contact 150 individuals and ask them whether or not they plan on voting for the levy. Of these 150 respondents, 78 people say they plan on voting for the levy. The school board wants to determine whether or not the data supports the idea that more than 50% of people plan on voting for the levy. **Conduct a hypothesis test at the 0.05 significance level to test this claim.**

a. What are the hypotheses?
b. What is the significance level?
c. What is the value of the test statistic?
d. What is the p-value?
e. What is the correct decision?
f. What is the appropriate conclusion/interpretation?
g. Are the assumptions met? Explain.

Solution:

a. $H_0: p = 0.50$ VS. $H_A: p > 0.50$

b. $\alpha = 0.05$

c. $TS = \frac{\hat{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.52-0.5}{\sqrt{\frac{(0.50)(0.50)}{150}}} = \frac{0.02}{0.0408} = \mathbf{0.49}$

d. $p-value = P(get\ our\ results\ or\ more\ extereme \mid H_0\ is\ true)$

$p-value = P(\hat{p} > 0.52 \mid p = 0.50) = P(Z > 0.49) = 1 - P(Z < 0.49) = 1 - 0.6879 = \mathbf{0.3121}$

a. $\because 0.3121 > \alpha = 0.05$, $\therefore$ Fail to reject H_0.

b. The data does not provide statistically significant evidence that the true proportion of all Americans who are planning to vote for the levy is greater than 0.50.

c.

i. **Independence:**
- Random
- $n < 10\%\ of\ the\ population\ ?\ unknown$

ii. **Sample size:**

$n\hat{p} \geq 10 \rightarrow (150)(0.52) = \mathbf{78 \geq 10}$

$n(1-\hat{p}) \geq 10 \rightarrow (150)(0.48) = \mathbf{72 \geq 10}$

Copyright © 2022 Aburweis

Problem 9. Your local school board wants to determine if the proportion of people who plan on voting for the school levy in the upcoming election is different for families who have elementary aged students and families that do not. They conduct a random phone poll, where they contact 200 families. Of the 100 families with elementary-aged students, 75 say they plan on voting for the levy. Of the 100 families without elementary-aged students, 68 say they plan on voting for the levy. Let Group A = Families with Elementary Aged Students and Group B = Families without Elementary Aged Students. **Create a 98% confidence interval for the difference between the two proportions.**

- **a.** What is the 98% confidence interval?
- **b.** What is the correct interpretation of the confidence interval?
- **c.** Are the assumptions met? Explain.
- **d.** Based on the confidence interval, what can you say about the school board's question?

Solution:

a. $CI: (\hat{p}_A - \hat{p}_B) \pm Z^* \sqrt{\frac{\hat{p}_A\,(1-\hat{p}_A)}{n_A} + \frac{\hat{p}_B\,(1-\hat{p}_B)}{n_B}}$

$\hat{p}_A = \frac{75}{100} = 0.75\ , n_A = n_E = 100\ (Group\ A)$

$\hat{p}_B = \frac{68}{100} = 0.68\ , n_{AB} = n_{NE} = 100\ (Group\ B)$

$Z^* = 2.326$

$$(0.75 - 0.68) \pm (2.326)\sqrt{\frac{0.75\,(0.25)}{100} + \frac{0.68\,(0.32)}{100}} = 0.07 \pm (2.326)(0.06365)$$

$$= 0.07 \pm 0.148 \rightarrow \mathbf{98\%\ CI: (-0.078, 0.218)}$$

b. We are 98% confident that the true difference between the proportion of all families with elementary kids and the proportion of all families without elementary kids who plan on voting for the levy is between -0.078 and 0.218.

c.

i. **Independence:**

- Independence within groups? Random.
- Independence between groups? Separate samples.

ii. **Sample size:**

$Elementry - Yes = 75 \geq 10 \quad None\ elementry - Yes = 68 \geq 10$

$Elementry - No = 25 \geq 10 \quad None\ elementry - No = 32 \geq 10$

d. Zero is included in the interval, there may not be a difference.

Copyright © 2022 Aburweis

Problem 10. Your local school board wants to determine if the proportion of people who plan on voting for the school levy in the upcoming election is different for families who have elementary aged students and families that do not. They conduct a random phone poll, where they contact 200 families. Of the 100 families with elementary-aged students, 75 say they plan on voting for the levy. Of the 100 families without elementary-aged students, 68 say they plan on voting for the levy. Let Group A = Families with Elementary Aged Students and Group B = Families without Elementary Aged Students. **Conduct a hypothesis test at the 0.02 significance level to test this.**

- **a.** What are the hypotheses?
- **b.** What is the significance level?
- **c.** What is the value of the test statistic?
- **d.** What is the p-value?
- **e.** What is the correct decision?
- **f.** What is the appropriate conclusion/interpretation?
- **g.** Are the assumptions met? Explain.

Solution:

a. $H_0: p_E = p_{NE}$ vs. $H_A: p_E \neq p_{NE}$

b. $\alpha = 0.05$

c. First, we need to calculate $\hat{p}_{pooled} \rightarrow \quad \hat{p}_{pooled} \frac{S_E+S_{NE}}{N_E+N_{NE}} = \frac{75+68}{100+100} = \frac{143}{200} = 0.715$

$$TS = \frac{\hat{p}_A - \hat{p}_B}{\sqrt{\frac{\hat{p}_{pooled}\,(1-\hat{p}_{pooled})}{n_A} + \frac{\hat{p}_{pooled}\,(1-\hat{p}_{pooled})}{n_B}}} = \frac{0.75 - 0.68}{\sqrt{\frac{(0.715)(0.285)}{100} + \sqrt{\frac{(0.715)(0.285)}{100}}}}$$

$$= \frac{0.07}{0.06384} = \mathbf{0.10}$$

d. $p - value = P(|Z| > 1.10\,) = P(Z < -1.10 + Z > 1.10) = \ 2 \times P(Z < -1.10) = 2 \times 0.1357 =$ $\mathbf{0.2714}$

e. $p - \because 0.2714 > \ \alpha = 0.02, \ \therefore$ Fail to reject H_0.

f. The data does not provide statistically significant evidence that the true proportion of all elementary school families that will vote for the school levy and the true proportion of all non-elementary school families that will vote for the school levy are different.

g.

i. **Independence:**

- Independence within groups? Random.
- Independence between groups? Separate samples.

ii. **Sample size / success-failure:**

$$n_A \times \hat{p}_{pooled} = (100)(0.715) = 71.5 \geq 10$$

$$n_B \times \hat{p}_{pooled} = (100)(0.715) = 71.5 \geq 10$$

$$n_A \times (1 - \hat{p}_{pooled}) = (100)(0.285) = 28.5 \geq 10$$

$$n_B \times (1 - \hat{p}_{pooled}) = (100)(0.285) = 28.5 \geq 10$$

Problem 11. The p-value for a two-sided hypothesis test of the null hypothesis $H_0: \mu = 12$ is 0.07. Which of the following confidence interval would include the value 12? Select all that apply.

a. 90% Confidence interval.

b. 95% Confidence interval.

c. 99% Confidence interval.

Solution:

Moved to next week

Problem 12. The level of calcium in the blood of healthy young adults follows a normal distribution with $\mu = 10$ milligrams per deciliter and $\sigma = 4$. A clinic measures the blood calcium of 25 healthy pregnant young women at their first visit for prenatal care. The mean of these 25 measurements is $\bar{x} = 9.6$. We want to test the hypotheses $H_0: \mu = 10$; $H_A: \mu < 10$. What does it mean if the p-value is 0.0002?

a. If the true population mean is less than 10, the probability that we get a sample mean of 9.6 is 0.0002.

b. If the true population mean is less than 10, the probability that we get a sample mean of 9.6 or less is 0.0002.

c. If the true population mean is 10, the probability that we get a sample mean of 9.6 is 0.0002.

d. If the true population mean is 10, the probability that we get a sample mean of 9.6 or less is 0.0002.

e. None of the above.

Solution:

Moved to next week

Copyright © 2022 Aburweis