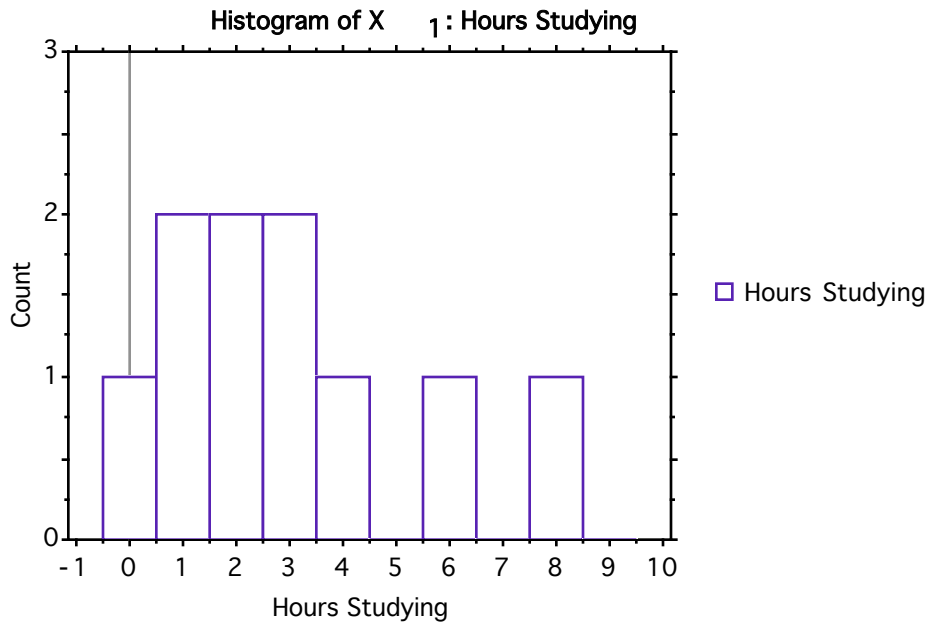


1. Suppose that you are interested in the number of hours that a person spends studying in a course (statistics, for instance). You question the students in a particular class about the number of hours spent studying in a typical week and obtain the results shown below.

a. Use your typical estimation procedures to make inferences about the parameters of the population from which this sample is likely to have been drawn. [10 pts]



$$\bar{X} = \frac{\sum X}{n} = \frac{30}{10} = 3 = \hat{\mu}$$

$$SS = \sum X^2 - \frac{(\sum X)^2}{n} = 144 - \frac{30^2}{10} = 54$$

$$s^2 = \frac{SS}{n-1} = \frac{54}{9} = 6 = \hat{\sigma}^2$$

$$s = \sqrt{6} = 2.45 = \hat{\sigma}$$

1b. Test the hypothesis that the above sample mean was drawn from a normally distributed population with  $\mu = 5$  and  $\sigma = 3$ . [5 pts]

$$H_0: \mu = 5$$

$$H_1: \mu \neq 5$$

**Decision Rule: If  $|z| \geq 1.96$ , reject  $H_0$ .**

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{3}{\sqrt{10}} = .95$$

$$z = \frac{\bar{X} - \mu}{\sigma_{\bar{x}}} = \frac{3 - 5}{.95} = -2.11$$

**Because  $|z| \geq 1.96$ , I would reject  $H_0$  and conclude that it's more likely that this sample was drawn from a population with  $\mu < 5$ .**

1c. Test the hypothesis that the above sample mean was drawn from a normally distributed population with  $\mu = 5$  and  $\sigma = 6$ . [5 pts]

$$H_0: \mu = 5$$

$$H_1: \mu \neq 5$$

**Decision Rule: If  $|z| \geq 1.96$ , reject  $H_0$ .**

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{6}{\sqrt{10}} = 1.9$$

$$z = \frac{\bar{X} - \mu}{\sigma_{\bar{x}}} = \frac{3 - 5}{1.9} = -1.05$$

**Because  $|z| < 1.96$ , I would retain  $H_0$  and conclude that it's possible that the sample was drawn from a population with  $\mu = 5$ .**

1d. Explain why you might have reached a different conclusion in 1b and 1c. [2 pts]

**These two problems highlight the implications of population variability. In the first problem, because the population variability is smaller, the standard error is also smaller. That results in a larger z-score, which leads to the rejection of  $H_0$ . Note that the observed difference in the size of the standard error is similar to the change you'd observe with changes in sample size. In either case, of course, when your standard error is smaller, you will find it easier to reject  $H_0$  (that is, you will have more power).**

2. What is the general formula for all standard scores? [2 pts]

$$\frac{\text{Score} - \text{Mean of the distribution from which the score was obtained}}{\text{Standard Deviation of the distribution from which the score was obtained}}$$

3. What does the Central Limit Theorem tell us about the sampling distribution of means. [2 pts.]

**The CLT tells us that with increasingly larger sample sizes, we can expect that the sampling distribution of the mean will become increasingly normal and will be centered around  $\mu$  with decreasing variability.**

4. Why is the sum of squares (SS) not a good measure of variability? [2 pts.]

**In general, the SS is not a good measure of variability because it continues to increase with increases in sample size (with the rare exception of adding scores at the mean). You'd like to have a measure of variability that compensated for sample size (as does the variance).**

5. When might the median be a better measure of central tendency than the mean? [2 pt.]

**When you're dealing with a skewed distribution, the median is usually a better index of central tendency.**

6. As we saw in class, human gestation periods are normally distributed with  $\mu = 268$  days and  $\sigma = 16$  days.

a. Suppose that a woman was interested in getting the whole thing over with fairly quickly. She would also like to avoid having a baby that's really premature. Suppose that she's interested in the probability that her gestation period might fall between 244 and 252 days. What would you tell her? [5 pts]

$$z = \frac{244 - 268}{16} = -1.5$$

$$z = \frac{252 - 268}{16} = -1.0$$

**.1587 - .0655 = .0932 OR, in other words, not too many women carry their babies for that period (only 9%)**

b. Suppose that she wonders what the gestation period might be for the briefest 35% of women (again, hoping that she might fall somewhere within this group). What gestation period cuts off the lowest 35% of gestation periods? [5 pts]

**A z-score of -.385 would cut off the lower 35% of the distribution.**

$$-.385 = \frac{X - 268}{16}$$

$$261.8 = X$$

**Thus, only 35% of women would have a gestation period of ~262 days or less.**

c. Suppose that this woman is in a natural-birth class of 16 pregnant women. What is the probability that the mean gestation period of this class would fall between 244 and 252 days? [5 pts]

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{16}{\sqrt{16}} = 4$$

$$z = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}} = \frac{244 - 268}{4} = -6$$

$$z = \frac{\bar{X} - \mu}{\sigma_{\bar{X}}} = \frac{252 - 268}{4} = -4$$

**Give it up! Virtually no group of 16 women would have a mean gestation period between 244 and 252 days. (The overall probability is less than .00003!)**

7. In class, we talked about the poor woman who was pregnant in San Diego. She argued that she had a *very* long gestation period (~308 days!!!). Essentially, she is claiming that her gestation period was sampled from the normal population ( $\mu = 268$  days and  $\sigma = 16$  days). When faced with a situation such as this, a statistician has to make a decision. Given that her gestation period would yield a z-score of 2.5, what decision would you make? Why? What kind of error might you be making with that decision? Why is it that you can never make a Type I Error *and* a Type II Error on the same decision? [ 5 pts]

**Given our statistical decision-making machinery, you'd have to treat a z-score of 2.5 as well into the rejection region (greater than 1.96). As such, you'd conclude that it's unlikely that she was drawn from a population with  $\mu = 268$ ...or that she was and her conception date was later than she is claiming! In essence, you would be denying her claim that she did not play around while hubby was out to sea. Of course, you could well be making a Type I error in making such a claim. (I doubt that acknowledging as much would be much comfort to the poor woman in San Diego!)**

**Thank the Random Number God, you cannot make both a Type I and a Type II error simultaneously. You can only make a Type I error when you reject  $H_0$ . You can only make a Type II error when you retain  $H_0$ . Because you make one decision or the other, you cannot make both errors on any given decision.**