HW4-Lin

Bin Lin

2016-10-10

Point estimates that are based on samples only approximate the population parameter, and they vary from one sample to another.

4.4 Heights of adults. (a) What is the point estimate for the average height of active individuals? What about the median? The point estimate for the average height is 171.1. The point estimate of median is 170.3.

b. What is the point estimate for the standard deviation of the heights of active individuals? What about the IQR? The point estimate for the standard deviation is 9.4. IQR = Q3-Q1 = 177.8 - 163.8 = 14

c. Is a person who is 1m 80cm (180 cm) tall considered unusually tall? And is a person who is 1m 55cm (155cm) considered unusually short? Explain your reasoning. 180cm has standard deviation of 0.947, which is less than 2 standard deviation, so he or she is not considered to be unusually tall. 155cm has standard deviation of -1.713, which is less than 2 standard deviation lower than the average height. So he or she is still not considerd to be unusually short.

d. The researchers take another random sample of physically active individuals. Would you expect the mean and the standard deviation of this new sample to be the ones given above? Explain your reasoning. No, because the sample is random again, it is going to be another point estimate of the population parameter. The distribution will be similar but not exactly the same.

e. The sample means obtained are point estimates for the mean height of all active individuals, if the sample of individuals is equivalent to a simple random sample. What measure do we use to quantify the variability of such an estimate (Hint: recall that SD¯x = ! pn)? Compute this quantity using the data from the original sample under the condition that the data are a simple random sample. SE = 9.4/(sqrt(507)) = 0.417

4.14 Thanksgiving spending, Part I. (a) We are 95% confident that the average spending of these 436 American adults is between $80.31 and $89.11. False, inference is made on the population parameter, not the point estimate. The point estimate is always in the confidence interval.

b. This confidence interval is not valid since the distribution of spending in the sample is right skewed. The data distribution is not very strongly skewed. In addition, the sample is independent(<10%) and greater than 30. Therefore, we can say the sample is closed to normal distribution.

c. 95% of random samples have a sample mean between $80.31 and $89.11. True, that is the definition of confidence interval.

d. We are 95% confident that the average spending of all American adults is between $80.31 and $89.11. True, that is the definition of confidence interval.

e. A 90% confidence interval would be narrower than the 95% confidence interval since we don’t need to be as sure about our estimate. True, since we are less confident, the interval will be narrower so that it become less likely to capture the parameter.

f. In order to decrease the margin of error of a 95% confidence interval to a third of what it is now, we would need to use a sample 3 times larger. False. the formular is SE = s/(sqrt(n)). Therefore, to decrease the margin of error to a third, the sample size need to be 9 times larger.

g. The margin of error is 4.4. True, (89.11 - 80.31) / 2 = 4.4

4.24 Gifted children, Part I. (a) Are conditions for inference satisfied? Yes, sample size(36) is greater than 30, less than 10% of population. The sample is randomly collected. The figure is independent, unimodal and not very skewed. Therefore, the conditons for inference satisfied.

b. Suppose you read online that children first count to 10 successfully when they are 32 months old, on average. Perform a hypothesis test to evaluate if these data provide convincing evidence that the average age at which gifted children fist count to 10 successfully is less than the general average of 32 months. Use a significance level of 0.10. H0: = 32 HA: <32 z-score = (30.69 - 32) / (4.31 / (sqrt(36)) = -1.824 p-value =

0.0344, which is less than the alpha level of 0.1. Therefore, we can reject the null hypothesis, we can conclude that the average age of gifted children first count to 10 is younger than 32 months at alpha = 0.1

c. Interpret the p-value in context of the hypothesis test and the data. p-value is 0.0344, which means the difference we notice can happen simply due to chance, however it only happens 3.44% percent of time, 96.56% percent of time is not due to chance, instead it will trully due to the difference bewteen the population mean and our null hypothesis.

d. Calculate a 90% confidence interval for the average age at which gifted children first count to 10 successfully. 30.69 +/- 1.64 * 4.31 / sqrt(36) = (29.511, 31.868)

e. Do your results from the hypothesis test and the confidence interval agree? Explain. Yes, because the 90% confidence interval does not include 32, which is our null hypothesis. Based on the hypothesis test we perform, we are 90 % confident the actualy population mean is between 29.511 months and 31.868 months. 32 months is most likely not the population mean. Therefore, we reject the null hypothesis.

4.26 Gifted children, Part II. (a) Perform a hypothesis test to evaluate if these data provide convincing evidence that the average IQ of mothers of gifted children is di???erent than the average IQ for the population at large, which is 100. Use a significance level of 0.10. H0: = 100 HA: != 100 Z-score = (118.2 - 100) / (6.5 / (sqrt(36))) = 16.8. p-value = 0, which is less thatn 0.1. Therefore, we can reject the null hypothesis. We can conclude that the average IW of mothers of gifted children is different than the average IQ for the population at large. It will be greater than population at large.

b. Calculate a 90% confidence interval for the average IQ of mothers of gifted children. 118.2 +/- 1.64 * 6.5 / (sqrt(36)) = (116.423, 119.977)

c. Do your results from the hypothesis test and the confidence interval agree? Explain. Yes, the 90% confidence interval exclude 100. Therefore, at alpha = 0.1 level, the average IQ of mothers of gifted children is very different from 100.

4.34 CLT. Define the term “sampling distribution” of the mean, and describe how the shape, center, and spread of the sampling distribution of the mean change as sample size increases.

The sampling distributions of the mean will be nearly normal as long as the population distribution is nearly normal. The samples we collect no matter its sized will have a sample mean. If we collect numerous of samples from a population. The sample mean will be normally distributed and centered at the true population mean. The larger the sample size, the smaller the margin of error.

4.40 CFLBs. (a) What is the probability that a randomly chosen light bulb lasts more than 10,500 hours?

```
1 - pnorm(10500, mean = 9000, sd = 1000)
```

```
## [1] 0.0668072
```

6.68%

b. Describe the distribution of the mean lifespan of 15 light bulbs. The mean lifespan of 15 light bulbs shoud be normally distributed. It will be centered at 9000 hours and the standard deviation should be about 1000 / (sqrt(15)) = 258 hours.

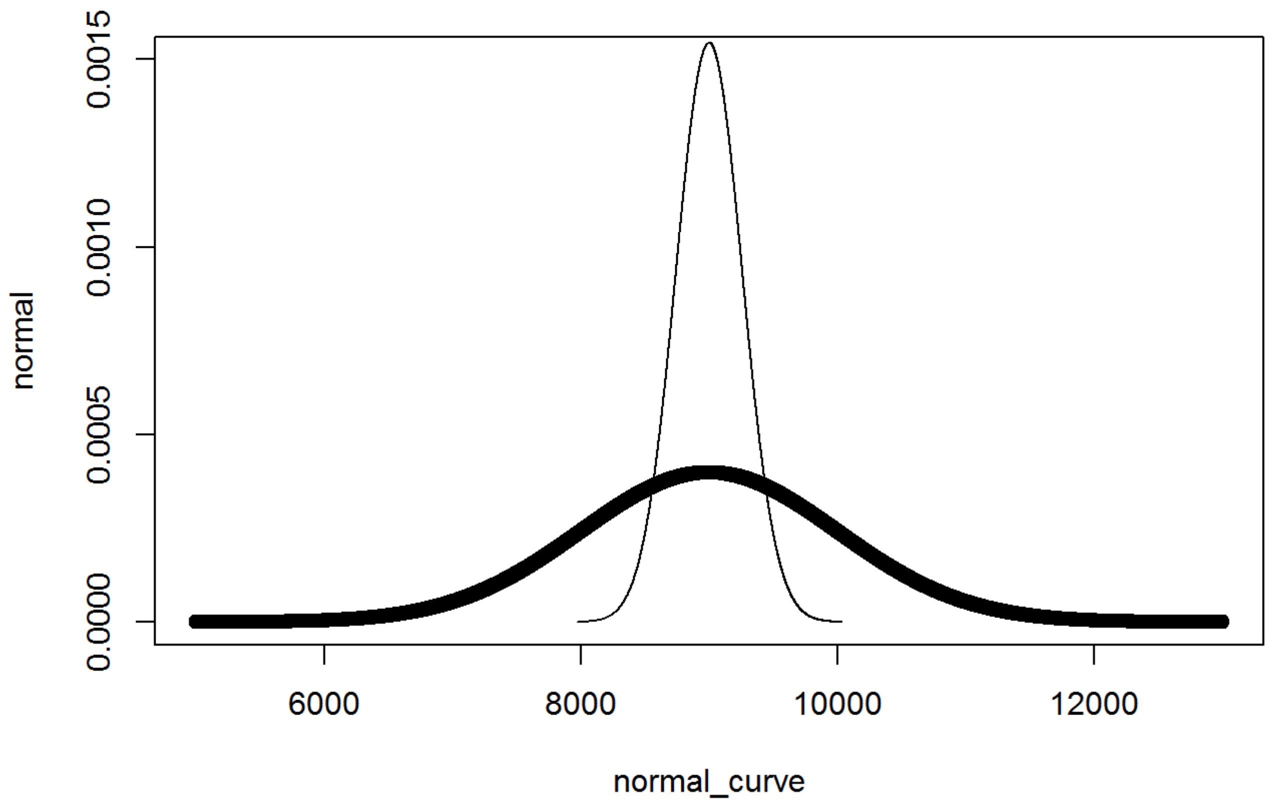
c. What is the probability that the mean lifespan of 15 randomly chosen light bulbs is more than 10,500 hours? z-score = (10500 - 9000) / 258 = 5.814 p-value = 100%, therefore, the mean lifespan of 15 ramdonly chosen light bulbs higher than 10500 hours is closed to 0.

d. Sketch the two distributions (population and sampling) on the same scale.

```
normal_curve <- seq(9000 - (4 * 1000), 9000 + (4 * 1000), 1)
normal <- dnorm(normal_curve, 9000, 1000)

sample_curve<- seq(9000 - (4 * 1000/(sqrt(15))),  9000+ (4 * 1000/(sqrt(15))), 1)
sample<- dnorm(sample_curve, 9000, 1000/(sqrt(15)))
```

```
plot(normal_curve, normal, lines(sample_curve, sample), ylim = c(0, 0.0015))
```



e. Could you estimate the probabilities from parts (a) and (c) if the lifespans of light bulbs had a skewed distribution? We can not, because the probability we got from part a and b is based on normal distribution. If the distribution is skewed, then we can not estimate.

4.48 Same observation, different sample size. Suppose you conduct a hypothesis test based on a sample where the sample size is n = 50, and arrive at a p-value of 0.08. You then refer back to your notes and discover that you made a careless mistake, the sample size should have been n = 500. Will your p-value increase, decrease, or stay the same? Explain. p-value will decrease, because SE will decrease because of larger sample size(inverse relationship). If SE is smaller, larger the z-score of the sample mean. Therefore, p-value will be smaller.