Homework4_StephenJones

Stephen Jones

March 14, 2019

4.2 Heights of adults.

Researchers studying anthropometry collected body girth measurements and skeletal diameter measurements, as well as age, weight, height and gender, for 507 physically active individuals. The histogram below shows the sample distribution of heights in centimeters.

(a) What is the point estimate for the average height of active individuals? What about the median?

The point estimate is the same as the mean, which is 171.1cm. The median is 170.3cm.

(b) What is the point estimate for the standard deviation of the heights of active individuals?

The point estimate of the standard deviation is 9.4cm.

What about the IQR?

$IQR = Q3 - Q1 = 14$cm

(c) Is a person who is 1m 80cm (180cm) tall considered unusually tall?

A person who is 180cm tall is taller than 75% of the sample.

And is a person who is 1m 55cm (155cm) considered unusually short? Explain your reasoning.

A person who is 155cm is shorter than at least 75% of the sample. This assessment is directly from $Q1$ and $Q3$, which indicate the height at which 25% and 75% of the sample, respectively, fall when sorted by height.

(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.

The mean and the standard deviation would vary with another random sample due to sampling variation.

(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 $SD_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$)? Compute this quantitity using the data from the original sample under the condition that the data are a simple random sample.

The estimate is $SD_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = 0.417$.

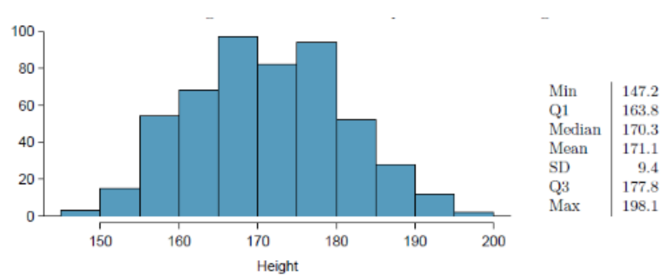


Figure 1: from the text, p. 204

4.14 Thanksgiving spending, Part I.

The 2009 holiday season, which kicked off on Nov 27, 2009 (the day after Thanksgiving)l had been marked by somewhat lower self-reported consumer spending than was seen during the comparable period in 2008. To get an estimate of consumer spending, 436 randomly sampled American adults were surveyed. Daily consumer spending for the six-day period after Thanksgiving, spanning the Black Friday weekend and Cyber Monday, averaged \$84.71. A 95% confidence interval based on this sample is (\$80.31, \$89.11). Determine whether the following statements are true or false, and explain your reasoning.

(a) We are 95% confident that the average spending of these 436 American adults is between \$80.31 and \$89.11.

This is false. The confidence interval is a range within which the mean exists with 95% probability.

(b) This confidence interval is not valid since the distribution of spending in the sample is right skewed.

This is false. The confidence interval is valid when the distribution is skewed, particularly with a sample size of 436.

(c) 95% of random samples have a sample mean between \$80.31 and \$89.11.

This is false. The confidence interval applies to the population mean, not the sample mean.

(d) We are 95% confident that the average spending of all American adults is between \$80.31 and \$89.11.

True. This aligns with the concept of what a confidence interval is, although this should reference the time period.

(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; this reasoning holds.

(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 sample would need to be $3^2 = 9$ to lower the error by $\frac{1}{3}$.

(g) The margin of error is 4.4.

True; The margin of error is $\frac{89.11-80.31}{2} = 4.4$.

4.26 Gifted children, Part II.

Exercise 4.24 describes a study on gifted children. In this study, along with variables on the children, the researchers also collected data on the mother's and father's IQ of the 36 randomly sampled gifted children. The histogram below shows the distribution of mother's IQ. Also provided are some sample statistics.

(a) Perform a hypothesis test to evaluate if these data provide convincing evidence that the average IQ of mothers of gifted children is different than the average IQ for the population at large, which is 100. Use a significance level of 0.10.

$H_0 : \mu = 100$

$H_A : \mu > 100$

$z = \frac{\bar{X}-\mu}{\sigma} = \frac{118.2-100}{6.5} = 2.8$

`pnorm((118.2-100)/6.5)` $= 0.9974449$

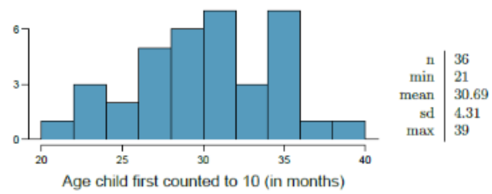


Figure 2: from the text, p. 211

Since the p-value is $0.0026 < 0.10$, we reject the null hypothesis; the IQ of these mothers are greater than the population at large.

(b) Calculate a 90% confidence interval for the average IQ of mothers of gifted children.

$SE = \frac{\sigma}{\sqrt{n}} = \frac{6.5}{\sqrt{36}} = 1.08$
The z-score for a 90% confidence interval is 1.645.
The confidence interval is $(118.2 - (1.645 \times 1.08), 118.2 + (1.645 \times 1.08)) = (116.42{,}119.98)$

(c) Do your results from the hypothesis test and the confidence interval agree? Explain.

Yes, the results agree. The mean of the population at large exists outside of the range detailed by the confidence interval. Additionally, there is a miniscule probability (0.003) that the sample mean is valid for the population at large.

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 distribution of the mean refers to the tendency of the means of random samples to follow a normal distribution when sample size is nonskewed and generally greater than 30. As sample size increases, the center becomes more pronounced at the mean, the spread becomes tighter, and the shape (as it relates to frequency) becomes both smoother and more perfectly bell-shaped. This trend is addresses and described by the Central Limit Theorem.

4.40 CFLBs.

A manufacturer of compact fluorescent light bulbs advertises that the distribution of the lifespans of these light bulbs is nearly normal with a mean of 9,000 hours and a standard deviation of 1,000 hours.

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

```
greaterthan10500<-round(1 - pnorm((10500-9000)/1000),4)
cat("Probability that the bulb lasts more than 10,500 hours is",greaterthan10500,".")
```

```
## Probability that the bulb lasts more than 10,500 hours is 0.0668 .
```

(b) Describe the distribution of the mean lifespan of 15 light bulbs.

The mean of a random sample of 15 light bulbs would vary; as more random samples of 15 are tested–particularly when 30 or more random samples are analyzed–the distribution would tend toward normality.

(c) What is the probability that the mean lifespan of 15 randomly chosen light bulbs is more than 10,500 hours?

```
LB15gt10500 <-(10500-9000)/(1000/sqrt(15))
Pgt10500 <- pnorm(LB15gt10500, 9000, 1000)

Pgt10500
```

```
## [1] 1.189897e-19
```

```
cat("\n\nProbability that 15 bulbs last more than 10,500 hours is",round(Pgt10500,3),".\n")
```

```
##
##
## Probability that 15 bulbs last more than 10,500 hours is 0 .
```

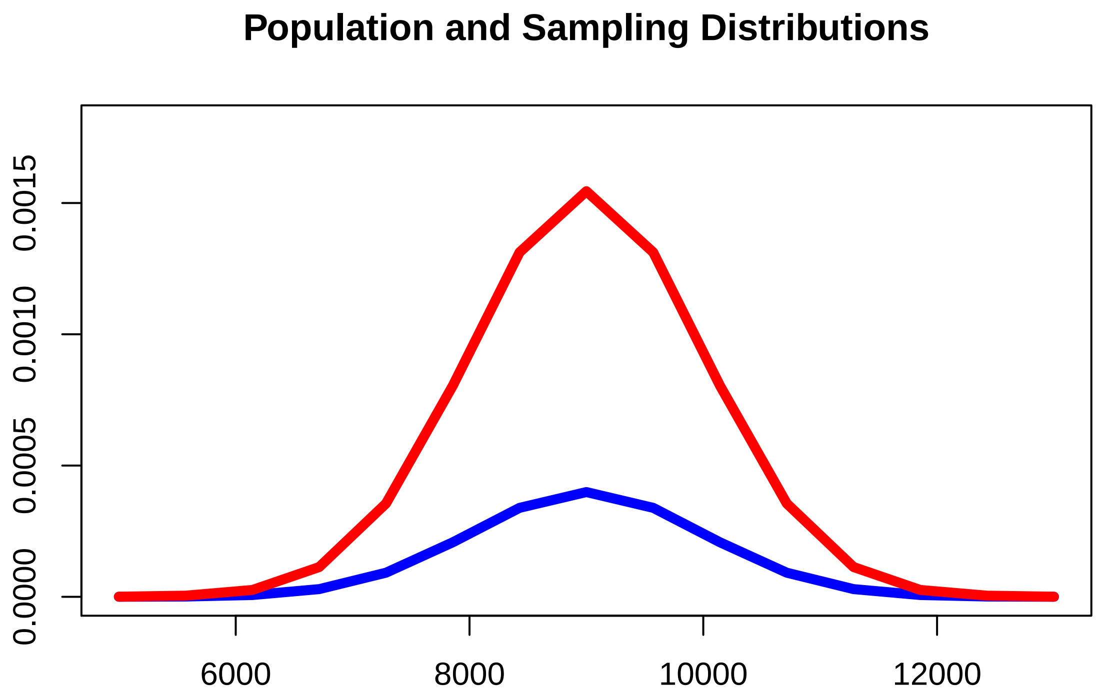
(d) Sketch the two distributions (population and sampling) on the same scale.

```
#adapted from https://stackoverflow.com/
#               questions/10543443/how-to-draw-a-standard-normal-distribution-in-r
n<-15
MN<-9000
SD<-1000
SE<-1000/sqrt(n)

Npop <- seq(MN-(4*SD),MN+(4*SD), length=15)
Nsamp <- seq(MN-(4*SE),MN+(4*SE), length=15)

hgt<-dnorm(Npop,MN,SD)
hrdm<-dnorm(Nsamp,MN,SE)

plot(Npop,hgt,type="l",col="blue",lwd=5,
xlab="",
ylab="",main="Population and Sampling Distributions", ylim=c(0,0.0018))
lines(Npop, hrdm, col="red",lwd=5)
```



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.

As sample size increases, standard error decreases; similarly, as sample size increases, standard devation decreases and z-score increases. Therefore, the p-value will decrease.
