Chapter 4

Sedo Oh

September 26, 2017

4.4 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? Mean = 171.1 Median = 170.3

(b) What is the point estimate for the standard deviation of the heights of active individuals? What about the IQR? SD = 9.4 IQR = 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. A person who is 180cm is tall but not unusally tall. This person is still within 1SD from the mean. A person who is 155cm is bit unusually short since that's outside 1SD as well as the histogram shows that there are less than 20 people (out of 507) who are shorter than 155cm.

(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.

Not exactly the same but close. Since this is sample, it wouldn't be the same people in the sample.

(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? Compute this quantity using the data from the original sample under the condition that the data are a simple random sample

9.4 / sqrt(507) = 0.417

4.14 Thanksgiving spending, Part I.

The 2009 holiday retail season, which kicked off on November 27, 2009 (the day after Thanksgiving), 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.

FALSE. We know everything about the 436 American adults. So we can 100% be sure that the average is between the range.

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

FALSE. The sample size is large enough to offset the skewness.

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

TRUE. The SD and confidence interval range is calculated based on the sample distribution so approximately 95% of sample should be between the range.

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

TRUE. This is the definition of confidence interval for the sample distribution.

(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. 90% translates into lower Z-score.

(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. For a third of what is now, we need 9 times larger since it's 1/sqrt(n).

(g) The margin of error is 4.4. TRUE. (89.11-80.31)/2 Basically half of interval.

4.24 Gifted children, Part I.

Researchers investigating characteristics of gifted children collected data from schools in a large city on a random sample of thirty-six children who were identified as gifted children soon after they reached the age of four. The following histogram shows the distribution of the ages (in months) at which these children first counted to 10 successfully. Also provided are some sample statistics.

(a) Are conditions for inference satisfied? Yes. Every observation is independent. The sample size is greater than 30. It's difficult to tell but theoretically the population distribution should be normal for the age to count 10.

(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.

Ho = Mean of gifted children's age is 32months to count 10. Ha = Mean of gifted children's age is less than 32months to count 10.

```
mean = 30.69
sd = 4.31
SE= sd/sqrt(36)
1-pnorm(32,mean = 30.69, sd=SE)
```

```
## [1] 0.0341013
```

There is only 0.034 chance that the age would be 32 or greater among gifted children if the mean is truly 32. This shows that for significance level of 0.1, we reject null hypothsis, and alternate hypothesis is plausible.

(c) Interpret the p-value in context of the hypothesis test and the data. The p-value is 0.034 which is much smaller than 0.1. This shows that null hypothesis very unlikely and shows strong evidence for alternate hypothesis.

(d) Calculate a 90% confidence interval for the average age at which gifted children first count to 10 successfully.

```
Z=qnorm(0.95)
mean + Z*SE
```

```
## [1] 31.87155
```

```
mean - Z*SE
```

```
## [1] 29.50845
```

The range is between 31.87 to 29.51

(e) Do your results from the hypothesis test and the confidence interval agree? Explain. Yes. Upper bound of 31.87 shows that there's only 5% chance that the mean is greater than 31.87. This is similar to saying that to beat 32, which is big stronger constraint, there's only 3.4% chance.

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. Ho: IQ of moms = 100 Ha: IQ of moms > 100

```
mean = 118.2
sd = 6.5
SE= sd/sqrt(36)
pnorm(100,mean = 118.2, sd=SE)
```

```
## [1] 1.22022e-63
```

The chance that the average of mom's IQ of gifted children is practically zero. This easily beats 0.1 significance level.

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

```
Z=qnorm(0.95)
mean + Z*SE
```

```
## [1] 119.9819
```

```
mean - Z*SE
```

```
## [1] 116.4181
```

The range is from 116 to 120.

(c) Do your results from the hypothesis test and the confidence interval agree? Explain. Yes. The the lower boundary in the 90% confidence interval is 116, and that is far 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.

Sampling distribution of the mean consists of number of means measured based on numerous samples of the same size. As the sample size increases, the shape becomes more normal, center remains the same (as long as the size is sufficient), and spread becomes narrower.

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?

```r
1-pnorm(10500,9000,1000)
```

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

About 6.6%.

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

```r
1000/sqrt(15)
```

```
## [1] 258.1989
```

The mean should be around 9000, and the stand deviation should be around 258. It should follow normal distribution.

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

```r
1-pnorm(10500,9000,258)
```

```
## [1] 3.050719e-09
```

Practically zero.

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

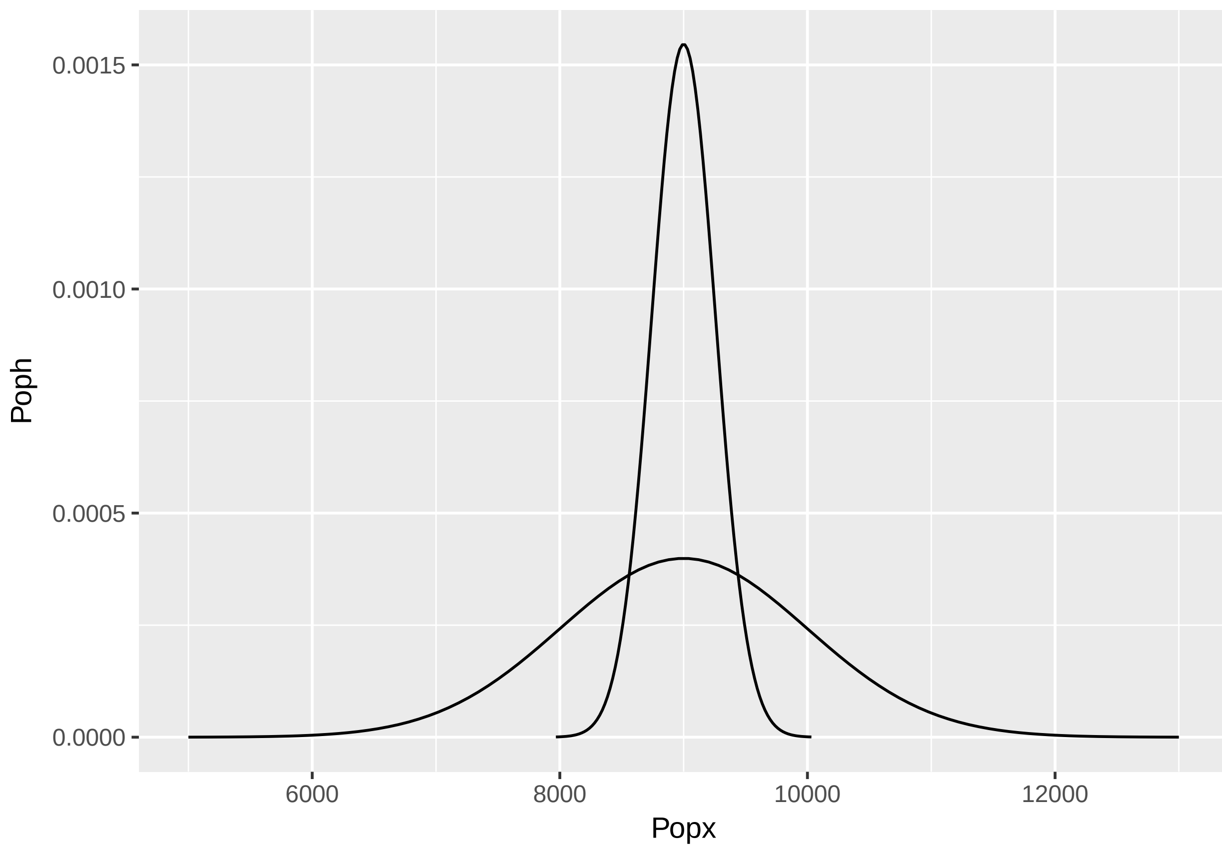
```r
##http://www.statmethods.net/advgraphs/probability.html - took some code from here
mean=9000; sd=1000

Popx <- seq(-4,4,length=100)*sd + mean
Poph <- dnorm(Popx,mean,sd)

mean=9000; sd=258
Samx <- seq(-4,4,length=100)*sd + mean
Samh <- dnorm(Samx,mean,sd)

library(data.table)
bulbdata = as.data.table(cbind(Popx,Poph,Samx,Samh))

library(ggplot2)
ggplot(data=bulbdata, aes(Popx,Poph)) + geom_line() +
geom_line(data=bulbdata, aes(Samx,Samh))
```



(e) Could you estimate the probabilities from parts (a) and (c) if the lifespans of light bulbs had a skewed distribution? It's going to be difficult with skewed distribution.

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 is going to decrease since larger sample size will result in smaller Standard Error.