IS606 - Homework 4

Daniel Dittenhafer

September 24, 2015

4.4 Heights of adults (p204)

Researchers studying anthropometry collected body girth measurements and skeletal diameter measurements, as well as age, weight, height an gender, for 507 physically active individuals.

a) What is the point estimate for the average height of active individuals? What about the median? The mean would be the point estimate for the average height, which is 171.1. The median is 170.3.

b) What is the point estimate for the standard deviation of the heights of active individuals? What about the IQR? Given the distribution is mostly normal, we can use the sample standard deviation as the point estimate for the population standard deviation (p173), SD=9.4.

The IQR would be derived from the sample IQR of Q3 - Q1 = 177.8 - 163.8 = 14.

c) Is a person who is 1m 80cm (180cm) tall considered unusually tall? And is a person who is 1m 55cm (155cm) considered unusually short? Explain your reasoning. After codifying the sample mean and standard deviation, we compute the Z score for the 180cm tall scenario.

```
meanHeight <- 171.1
sdHeight <- 9.4

x <- 180
zTall <- (x - meanHeight) / sdHeight

pTall <- pnorm(zTall)
pTall
```

```
## [1] 0.8281318
```

Being 180cm tall puts one at 0.9468085 standard deviations above the mean with 17.19 % of people taller. As such, I would *not* consider being 180cm tall particularly unusual, though it is taller than 82.81 % of the sample.

```
x <- 155
zShort <- (x - meanHeight) / sdHeight
pShort <- pnorm(zShort)
pShort
```

```
## [1] 0.0433778
```

Being 155cm tall puts one at -1.712766 standard deviations below the mean with 95.66 % of people taller. As such, I would consider being 155cm tall unusual, with just 4.34 % of the sample being shorter

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. I would not expect identical mean and standard deviation in the new sample unless by some coincidence the second sample cases where identical to the first. With that said, any new sample mean and standard deviation would be normally distributed around the population mean and standard deviation.

e) We quantify the variability of the point estimates through the Standard Error (SE) which is the standard deviation of the sampling distribution. Using `R` we compute the SE for the sample below:

```
n <- 507
seHeight <- sdHeight / sqrt(n)
seHeight
```

```
## [1] 0.4174687
```

4.14 Thanksgiving spending, Part I (p208)

a) We are 95% confident that the average spending of these 436 American adults is between \$80.31 and \$89.11. More accurately, I would say we are 95% confident that the average spending of the population is between \$80.31 and \$89.11. It seems we should be 100% confident the average spending of the specific individuals in the survey is between the upper and lower bound. The confidence level of the interval is meant to measure the likelyhood that the population parameter falls within the range.

b) This confidence interval is not valid since the distribution of spending in the sample is right skewed. Although the sample's distribution is right skewed, this does not affect the *sampling* distribution's shape normal the confidence interval built from it when the sample size is sufficiently large. In our current case, n=436, which is well above 100 so we rely on the Central Limit Theorem for a normal sampling distribution.

c) 95% of random samples have a sample mean between \$80.31 and \$89.11 While this statement might be true in some particular scenario, it is not known to be true as a result of the 95% confidence internal of this data set. Rather, we expect 95% of confidence intervals at this level from random samples would contain the population's mean.

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

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. Correct. This is due to the fact that the number of standard deviations surronding the mean that encompass 90% (1.645) of the normal distribution's center is less than the number associated with 95% (1.96)

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. First we can compute the Standard Error by reversing the Margin of Error computation:

```
n <- 436
SE <- 4.4 /  1.96
SE
```

```
## [1] 2.244898
```

The Standard Error comes from the Standard Deviation divided by the square root of the number of observations.

```
stdev <- SE * sqrt(n)
stdev
```

```
## [1] 46.87485
```

Does tripling the sample size achieve a margin of error of 1.4666667 assuming standard deviation stays the same?

```
newN <- n * 3
newSE <- stdev / sqrt(newN)
newMoE <- newSE * 1.96
newMoE
```

```
## [1] 2.540341
```

The new margin of error is 2.5403412, which is not a third of the current margin of error.

In order to acheive a margin of error of 1.4666667 given a standard deviation of 46.8748456 what sample size would we need (again assuming the same standard deviation)?

```
desiredMoE <- 4.4 / 3
reqN <- ((1.96 * stdev) / desiredMoE)^2
reqN
```

```
## [1] 3924
```

```
checkMoE <- 1.96 * (stdev / sqrt(reqN))
checkMoE
```

```
## [1] 1.466667
```

Therefore, inorder to achieve a margin off error of 1.4666667, a sample of 3924 would be needed assuming current standard deviation stays the same.

g) The margin of error is 4.4. Correct. 4.4 is the result of 1.96 * SE.

4.24 Gifted children, Part I (p211)

stat	value
n	36
min	21
mean	30.69
sd	4.31
max	39

a) Are conditions for inference satisfied? Sample size is small, at 36, but above the minimum of 30. The distribution is a very rough normal shape. Maybe we can accept this as meeting the conditions for inference.

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 first count to 10 successfully is less than the general average of 32 months. Use a significance level of 0.10. Setting up the hypothesis as follows:

$H_0 : m_g = 32$ (The gifted children's average months is equal to 32 (the average for children in general).)

$H_A : m_g < 32$ (The gifted children's average months is less than 32.)

$\alpha = 0.10$

```
a <- 0.10
xbar <- 30.69
SEz <- 4.31 / sqrt(36)
theZ <- qnorm(1 - a) # ~1.285
theZ
```

```
## [1] 1.281552
```

```
lower <- xbar - (theZ * SEz)
upper <- xbar + (theZ * SEz)
ci <- c(lower, upper)
ci
```

```
## [1] 29.76942 31.61058
```

Due to the resulting 90% confidence interval of 29.7694188 - 31.6105812 being less than the 32 month value, I conclude to reject the null hypothesis in favor of the alternative.

c) Interpret the p-value in context of the hypothesis test and the data. We compute the Z score of the gifted months relative to the H_0 value of 32, then lookup the p-value using the `pnorm` function. This is a left-tailed test, therefore we don't subtract from 1, and instead keep the pval as is.

```
MgZ <- xbar - 32 / SEz
MgZ
```

```
## [1] -13.85756
```

```
pval <- pnorm(MgZ)
pval
```

```
## [1] 5.724732e-44
```

The pval of 0 is much lower than the significance level $\alpha = 0.1$. This suggests the gifted months mean of 30.69 is not even close to the 32 month average.

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

e) Do your results from the hyposthesis test and the confidence interval agree? Explain.

4.26 (p)

4.34 (p)

4.40 (p)

4.48 (p)