Homework on Chapter 4

Valerie Briot

March 10, 2016

This document pertains to the homework exercises assigned out of chapter 4 for CUNY IS606 - Probability and Statistics. This chapter is introductary chapter on inference.

The following homework exercises were assigned; 4.4, 4.14, 4.24, 4.26, 4.34, 4.40, and 4.48.

```r
library(IS606)
```

```
##
## Welcome to CUNY IS606 Statistics and Probability for Data Analytics
## This package is designed to support this course. The text book used
## is OpenIntro Statistics, 3rd Edition. You can read this by typing
## vignette('os3') or visit www.OpenIntro.org.
##
## The getLabs() function will return a list of the labs available.
##
## The demo(package='IS606') will list the demos that are available.
```

```
##
## Attaching package: 'IS606'
```

```
## The following object is masked from 'package:utils':
##
##     demo
```

Exercise 4.4 - Heights of Adults

a) What is the point estimate for the average height of active individual?

The point estimate for average height would be the mean, in tis case, Mean = 171.1

What about the median?

The median = 170.3

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

The point estimate for standard deviation would be SD = 9.4

What about the IQR? The IQR = Q3 - Q1 = 14

c) Is a person who is 1m 80cm tall considered unusually tall? Let us calculate the difference between 1m 80 and the mean; 8.9 Let us also compare this value to the standard deviation; 8.9 < 1*SD, this value is within 1 standard deviation of the mean.
This height would not be considered unusual.

And a person who is 1m 55cm considered unusually short? Again, let us calculate the difference between this measurement and the mean, in absolute value; 16.1 < 2*SD, this value is within 2 standard

deviations of the mean.
This height would not be considered unusual.

d) Were we to take another random sample, would you expect the mean and the standard deviation of this new sample to be the same as the one given for the first sample? No, we would not expect another sample to have same mean and SD, mean and SD for each sample may be different. However, over time, we would expect that the mean and SD of all the samples approach the mean and SD of population.

e) What measure do we use to quantify the vairablility of such an estimate? Standard Error, SE $= \frac{\sigma}{\sqrt{n}}$ $SE \quad = \quad \frac{9.4}{\sqrt{507}}$ = 0.42

Exercise 4.14

a) False, we cannot make inference on the population parameter, not on the point estimate. The point estimate is always in the confidence interval.

b) False, the distribution spending is right skewed but the sample size is large 436, we are also assuming that the randomly selected adults are < 10% of population so that we can assume independence, so because of the size of sample, we can be more lenient of skewness of the distribution.

c) False, the confidence interval is about sample mean, not about random samples.

d) True, for the period after Thanksgiving, we are 95% confident that the a average spending of American adults is between \$80.31 and \$89.11.

e) False, Standard Error, SE = $\frac{\sigma}{\sqrt{n}}$
Margin of error = Zˆ ˆ SE, reduce by 3, increase sample size by 3^2, by 9

f) True, The margin of error = 4.4

Exercise 4.24

a) Are conditions for inference satisfied?
Yes, sample is randomly selected we do not have number on whether sample is less than 10% population but since we are dealing with a large city we will assume so therefore assume independence. Size of sampling > 30 and the distribution does not show strong skewness.

b) Perform Hypothesis test
H0 = average age to learn to read for gifted children same as for non-gifted children, H0 $\mu = 32$ Ha = average age to learn to read for gifted children is less than for non-gifted children; Ha $\mu < 32$

$SE \quad = \quad \frac{\sigma}{\sqrt{n}}$

$SE \quad = \quad \frac{4.31}{\sqrt{36}}$

SE = 0.72

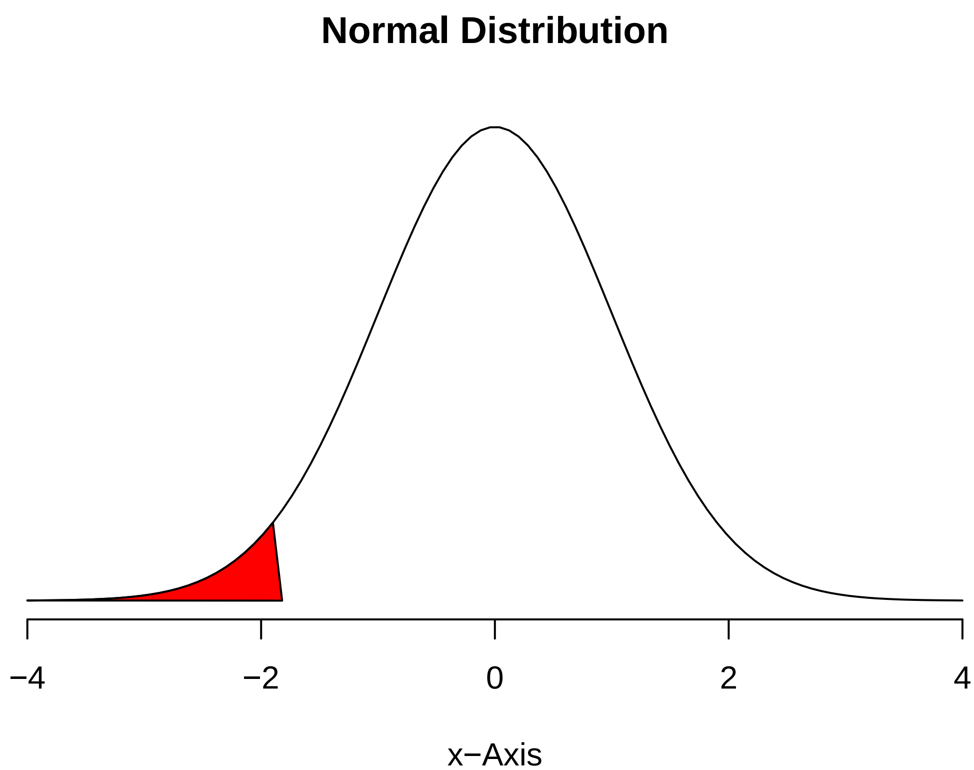
$Z \quad = \quad \frac{\bar{x} - \quad \mu_0}{SE}$

$Z \quad = \quad \frac{30.69 \quad - \quad 32}{0.72}$

Z = -1.82

We will draw a normal distribution and marked what area of the curve we are looking at

```r
# draw the normal distribution
normalPlot(bounds = c(-1.82), tails = TRUE)
```




```r
# finding the corresponding p-value

round(pnorm(-1.82, lower.tail = TRUE),4)
```

```
## [1] 0.0344
```

c) Interpret the p-value in context of H0 If the null hypothesis is true, the probability of observing a sample mean at least as low as 30.69 for a sample size of 32 children is 0.0344.
Since p-value is smaller than the level of significance 0.10, there is strong evidence to reject H0 and accept Ha.

d) Calculate 90% Interval

Interval = $\overline{x} \quad \pm \quad 1.65 \quad \times \quad SE$

30.69 + 1.65 * 0.72 = 31.878 30.69 - 1.65 * 0.72 = 29.502

Interval = (29.502, 31.878)

e) Do you result from d) agree with result from b)

Yes, The 90% confidence interval does not contain the value of 32. We are 90% confident that the average age gifted children learn to read is between 29.502 and 31.878 months old. It would be unsual to have a mean of 32.

Exercise 4.26 - Gifted children Part II

a)Hypothesis Test

H0: $\mu = 100$
Ha: $\mu > 100$

n = 36
$\bar{x} \quad = \quad 118.2 \quad \sigma \quad = \quad 6.5$

```
sample_n <- 36
sample_mean <- 118.2
sample_sd <- 6.5

null_Hypothesis_mean <- 100

# calculate SE
sample_se <- round(sample_sd/sqrt(sample_n), 2)

z_score <- round((sample_mean - null_Hypothesis_mean)/sample_se, 2)


# finding the corresponding p-value

round(pnorm(z_score, lower.tail = FALSE),4)
```

```
## [1] 0
```

This would mean that if H0 is true the probability of observing a sample mean = 100 is 0 which is less than the level of significance and we would reject the H0 Hypothesis.

The sample size is > 30 and distribution of the Mother's IQ is symetric with no skewness and no outliers. However, do we have independence? the sample of the mother's IQ is based on the sample of the gifted children (see above).

b) 90% Confidence Interval

```
interval_lower <- sample_mean - (1.65 * sample_se)
interval_higher <- sample_mean + (1.65 * sample_se)

interval_lower
```

```
## [1] 116.418
```

```
interval_higher
```

```
## [1] 119.982
```

Exercise 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.

Given a type of distribution, if we take random samples of this distribution of a given size (n), and look at the mean of such samples, we will be looking at the sampling distribution of the mean. As the sample size increase, the sampling distribution of the mean approaches a normal distribution. More symmetrical, less skewed, with less outlierss specially when the n>30.

Exercise 4.40 - CFLB's

a)

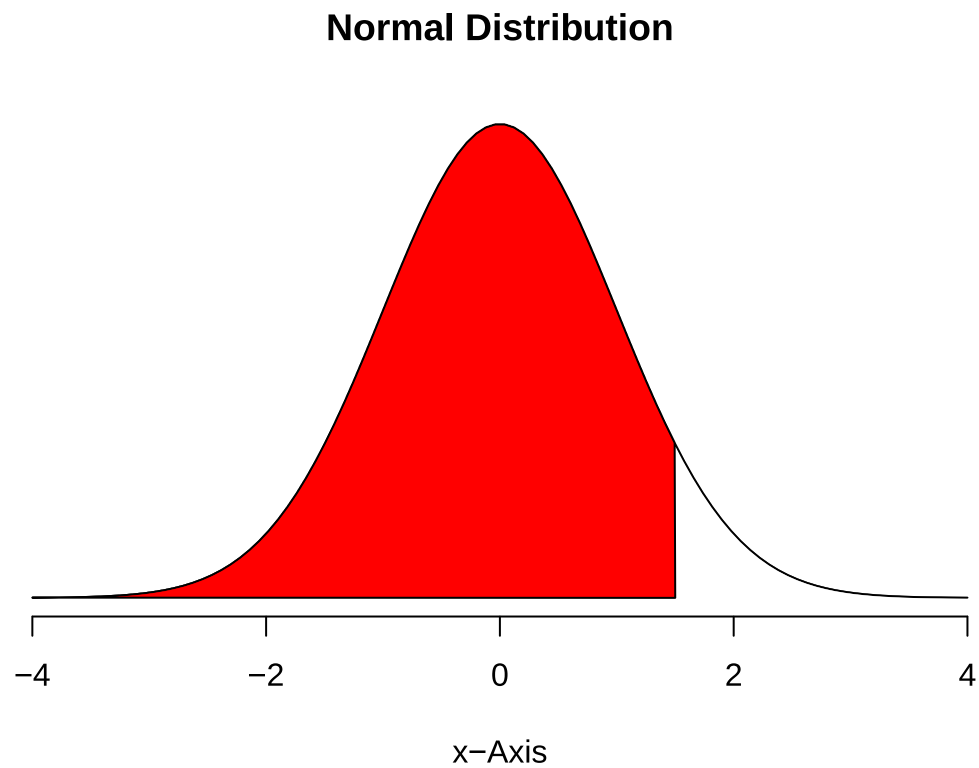
```
distribution_mean <- 9000
distribution_sd <- 1000

x<- 10500

z_score2<- (x - distribution_mean)/distribution_sd
z_score2
```

```
## [1] 1.5
```

```
normalPlot(bounds = c(z_score2), tails = TRUE)
```




```
round(pnorm(z_score2, lower.tail = FALSE),4)
```

```
## [1] 0.0668
```

Since we want probability that light bulb last more than 10,500, we need to take the upper tail.

b)

```
n_sample <-15
se <- distribution_sd/sqrt(n_sample)

z_score3 <- (x - distribution_mean)/se

z_score3
```

```
## [1] 5.809475
```

Probability that the mean lifespan than 15 randomly selected light bulb is more than 10500 is 0.

e) We would not be able to estimate the probability in a) without knowing that the distribution is nearly normal. Also, a sample size of 15 is too small to insure that he mean distribution is normal, hence we would not be able to estimate the probability in c).

Exercise 4.48

n=50 p-value = 0.08

however, n is actually 500

All other things being equal, SE = sd/sqrt(n)

Hence, Actual_SE would be smaller by a factor of sqrt(10) (approximately 3.16)

In turn, the z-score = (x - mu)/SE in absolute value, would be greater by the same factor, hence the z-score would be further away in the tail, the p-value should decrease.
