Assignment 3 - Sampling Distributions. Due September 28, 11:59pm 2018

EPIB607 - Inferential Statistics[a]

[a]Fall 2018, McGill University

This version was compiled on September 20, 2018

Because the results of random samples include an element of chance, we can't guarantee that our inferences are correct. What we can guarantee is that our methods usually give correct answers. We will see that the reasoning of statistical inference rests on asking "How often would this method give a correct answer if I used it very many times ?" To be able to answer these questions we need to understand sampling distributions and the normal curve. In this assignment you will practice calculating quantiles and probablities from the Normal distribution. All graphs and calculations are to be completed in an R Markdown document using the provided template. You are free to choose any function from any package to complete the assignment. Concise answers will be rewarded. Be brief and to the point. Please submit both the compiled HTML report and the source file (.Rmd) to myCourses by September 28, 2018, 11:59pm. Both HTML and .Rmd files should be saved as 'IDnumber_LastName_FirstName_EPIB607_A3'.

Sampling distribution | Standard error | Normal distribution | Quantiles | Percentiles | Z-scores

Template

The `.Rmd` template for Assignment 3 is available here

The `mosaic` package (optional)

The `mosaic` package provides a consistent and user-friendly interface for descriptive statistics, plots and inference. In particular you might find the `mosaic::xpnorm` and `mosaic::xqnorm` functions useful for this assignment. Have a look at the slides on sampling distributions for some examples on how to use these functions. Remeber to install the package:

```
install.packages("mosaic", dependencies = TRUE)
```

Then you must load the library to access its functions:

```
library(mosaic)
```

In-line R code

For this and future assignments you may find it useful to include calculations from R directly in your text. For example, in the following code chunk I calculate $P(Y < 2)$ where $Y \sim \mathcal{N}(0,1)$, and store the result in an object called `prob_less_2`:

```
```{r}
prob_less_2 <- mosaic::xpnorm(2)

# round to 2 digits
prob_less_2 <- round(prob_less_2, 2)
```
```

To print this result verbatim in an inline R expression use `` `r prob_less_2` `` in the text.

You can also call the function directly without storing the result in a code chunk using `` `r round(mosaic::xpnorm(2), 2)` ``
A sample answer would be: The probability that Y is less than 2 is 0.98.

1. Normal probability calculations

Using your method of choice, calculate the following probabilities assuming Y is a standard normal distribution ($Y \sim \mathcal{N}(0,1)$):

a) $P(Y < -1.80)$
b) $P(Y > -1.80)$
c) $P(Y \geq 1.60)$
d) $P(-1.8 < Z \leq 1.6)$

2. HDL cholesterol

a. US women over the age of 19 have a mean (HDL) cholesterol measure of 55 mg/dL with a standard deviation of 15.5 mg/dL. Assume HDL follows a Normal distribution.

   i. What percent of women have low values of HDL, where low is defined to be 40 mg/dL or less?
   ii. HDL levels of 60 mg/dL or more are believed to be protective against heart disease. What percent of women have protective levels of HDL?
   iii. What proportion of women has HDL in the range of 40-60 mg/dL?
   iv. What proportion of women has HDL in the range of 35-65 mg/dL?

3. Osteoporosis

Osteoporosis is a condition in which the bones become brittle due to loss of minerals. To diagnose osteoporosis, an elaborate apparatus measures bone mineral density (BMD). BMD is usually reported in standardized form. The standardization is based on a population of healthy young adults. The Wolrd Health Organization (WHO) criterion for osteoporosis is a BMD 2.5 or more standard deviations below the mean for young adults. BMD measurements in a population of people who are similar in age and sex roughly follow a standard normal distribution.

a. What percent of healthy young adults have osteoporosis?
b. Woman aged 70 to 79 are, of course, not young adults. The mean BMD in this age is about -2 on the standard scale for young adults. Suppose that the standard deviation is the same for young adults. What percent of this older population has osteoporosis?
c. Likewise, osteopenia is low BMD, defined by the WHO as a BMD between 1 and 2.5 standard deviations below the mean of young adults. What percent of healthy young adults have osteopenia?
d. The mean BMD among women aged 70 to 79 is about -2 on the standard scale for young adults. Suppose that the standard deviation is the same as for young adults. What percent of this older population has osteopenia?

4. How deep is the ocean?

This question is based on the in-class Exercise on sampling distributions.

a. For your samples of $n = 5$ and $n = 20$ of depths of the ocean, calculate the

   i. sample mean ($\bar{y}$)
   ii. standard error of the sample mean ($SE_{\bar{y}}$)

b. Calculate the 68%, 95% and 99% confidence intervals (CI) for both samples of $n = 5$ and $n = 20$. You may use the following formulas to calculate the confidence intervals:

CI	Formula
68%	$\bar{y} \pm 1 \times SE_{\bar{y}}$
95%	$\bar{y} \pm 2 \times SE_{\bar{y}}$
99%	$\bar{y} \pm 3 \times SE_{\bar{y}}$

Here is some sample code to calculate the CI. Suppose that based on a sample of $n = 20$, my $\bar{y} = 2900$ and $SE_{\bar{y}} = 20$. Then my confidence intervals are

```
ybar <- 2900
SEybar <- 20
# 68% CI
ybar + c(-1,1) * SEybar
# 95% CI
ybar + c(-2,2) * SEybar
# 99% CI
ybar + c(-3,3) * SEybar
```

Note that I have provided this code in the template as well. Take a look at the template before starting these calculations.

c. What do you notice about the size of the three intervals for a given sample size?