# Z-scores

*Put your name here*

*Put the date here*

## Introduction

In this assignment we will learn about z-scores which give us a way to put quantitative data on a meaningful scale, assuming that such data is normally distributed.

## Instructions

Presumably, you have already created a new project and downloaded this file into it. Please knit the document and work back and forth between this R Markdown file and the PDF output as you work through this module.

When you are finished with the assignment, knit to PDF one last time, proofread the PDF file **carefully**, export the PDF file to your computer, and then submit your assignment.

Sometimes you will be asked to add your own R code. That will appear in this document as a code chunk with a request for you to add your own code, like so:

```
## Add code here to [do some task]...
```

Be sure to remove the line `## Add code here to [do some task]...` when you have added your own code.

Sometimes you will be asked to type up your thoughts. That will appear in the document as follows:

Please write up your answer here.

Again, please be sure to remove the line "Please write up your answer here" when you have written up your answer. In these areas of the assignment, please use contextually meaningful full sentences/paragraphs (unless otherwise indicated) and proper spelling, grammar, punctuation, etc. This is not R code, but rather a free response section where you talk about your analysis and conclusions. If you need to use some R code as well, you can use inline R code inside the block between `\begin{answer}` and `\end{answer}`, or if you need an R code chunk, please go outside the `answer` block and start a new code chunk.

## Load Packages

We load the standard `mosaic` package.

```
library(mosaic)
```

## Z-scores

Sometimes it is easier to refer to a value in terms of how many standard deviations it lies from the mean. As an example of this, systolic blood pressure (SBP, measured in millimeters of mercury, or mmHg) is more-or-less normally distributed in women ages 30–44 in the U.S. and Canada, with a mean of 114 and a standard deviation of 14.[1] Therefore, a systolic blood pressure of 100 is 14 mmHg below the mean (114

[1]Statistics from the World Health Organization: http://www.who.int/publications/cra/chapters/volume1/0281-0390.pdf

mmHg), but since the standard deviation is 14 mmHg, this means that 100 is one standard deviation below the mean. This distance from the mean in terms of standard deviations is called a *z-score*.

We calculate z-scores using the following formula:

$$z = \frac{x - \mu}{\sigma}.$$

In our example, if we wanted to know the z-score for an SBP of 100, we just plug all the numbers into the formula above:

$$z = \frac{100 - 114}{14} = -1.$$

What is the z-score for an SBP of 132? In the graph of the normal model $N(\mu = 114, \sigma = 14)$, we can see that 132 lies between 128 and 142, which are 1 and 2 standard deviations above the mean, respectively. The exact z-score is

$$z = \frac{132 - 114}{14}$$

which comes out to 1.2857143.

The `scale` function from R also computes z-scores. Just note that the function takes arguments `center` and `scale`, not `mean` and `sd`.

```
scale(x = 100, center = 114, scale = 14)
```

```
##      [,1]
## [1,]   -1
## attr(,"scaled:center")
## [1] 114
## attr(,"scaled:scale")
## [1] 14
```

```
scale(x = 132, center = 114, scale = 14)
```

```
##          [,1]
## [1,] 1.285714
## attr(,"scaled:center")
## [1] 114
## attr(,"scaled:scale")
## [1] 14
```

Also note that the function spits about a bunch of crap we don't care about. This goes away if you use the command inline:

The z-score for 100 is -1 and the z-score for 132 is 1.2857143.

**Your turn**

If IQ scores have a mean of 100 and a standard deviation of 16, what are the z-scores for the following IQ scores? Write up your answers as full sentences using inline R code.

- 80

Please write up your answer here.

- 102

Please write up your answer here.

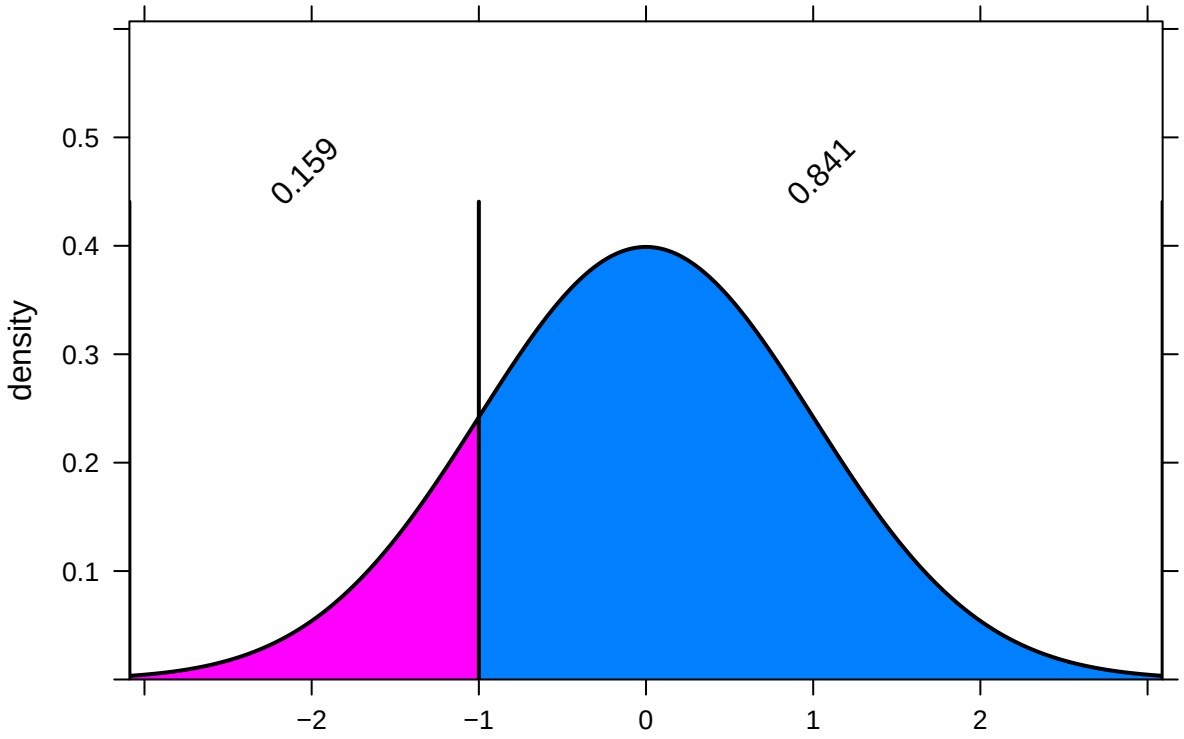
- 130

Please write up your answer here.

---

If you are working with z-scores, that also makes the commands easier when working with the normal model. The default settings for `pdist` and `qdist` are `mean = 0` and `sd = 1`. That saves you some typing. So, for example, we calculated above that an SBP of 100 has a z-score of -1. What percentage of women are expected to have SBP lower than 100?

```
pdist("norm", q = -1)
```




```
## [1] 0.1586553
```

The model predicts that 15.8655254% of women ages 30–44 in the U.S. and Canada will have SBP less than 100.

**Exercise**

Albert Einstein supposedly had an IQ of 160. Calculate the z-score for his IQ and then use that z-score to figure out what percentage of the population is smarter than Einstein. Use full sentences and inline R code to express you answer.

Please write up your answer here.

---

## Conclusion

The use of z-scores allows us to measure a distance from the mean in terms of standard deviations. So even though different real-life problems will have different units and different orders of magnitude, there is a scale in which all these problems are comparable. An effect that has a z-score of -2 is two standard deviations from the mean, and this has the same significance whether we're talking about blood pressure, IQ, or any other quantitative measurement, as long as such measurements are normally distributed.