

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. From the **Run** menu above, select **Run All** to run all existing code chunks.

When prompted to complete an exercise or demonstrate skills, you will see the following lines in the document:

ANSWER

These lines demarcate the region of the R Markdown document in which you are to show your work.

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
```

Be sure to remove the line **# Add code here** when you have added your own code. You should run each new code chunk you create by clicking on the dark green arrow in the upper-right corner of the code chunk.

Sometimes you will be asked to type up your thoughts. That will appear in the document with the words, “Please write up your answer here.” 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. You may need to use inline R code in these sections.

When you are finished with the assignment, knit to PDF and proofread the PDF file **carefully**. Do not download the PDF file from the PDF viewer; rather, you should export the PDF file to your computer by selecting the check box next to the PDF file in the Files pane, clicking the **More** menu, and then clicking **Export**. Submit your assignment according to your professor’s instructions.

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, 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:

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

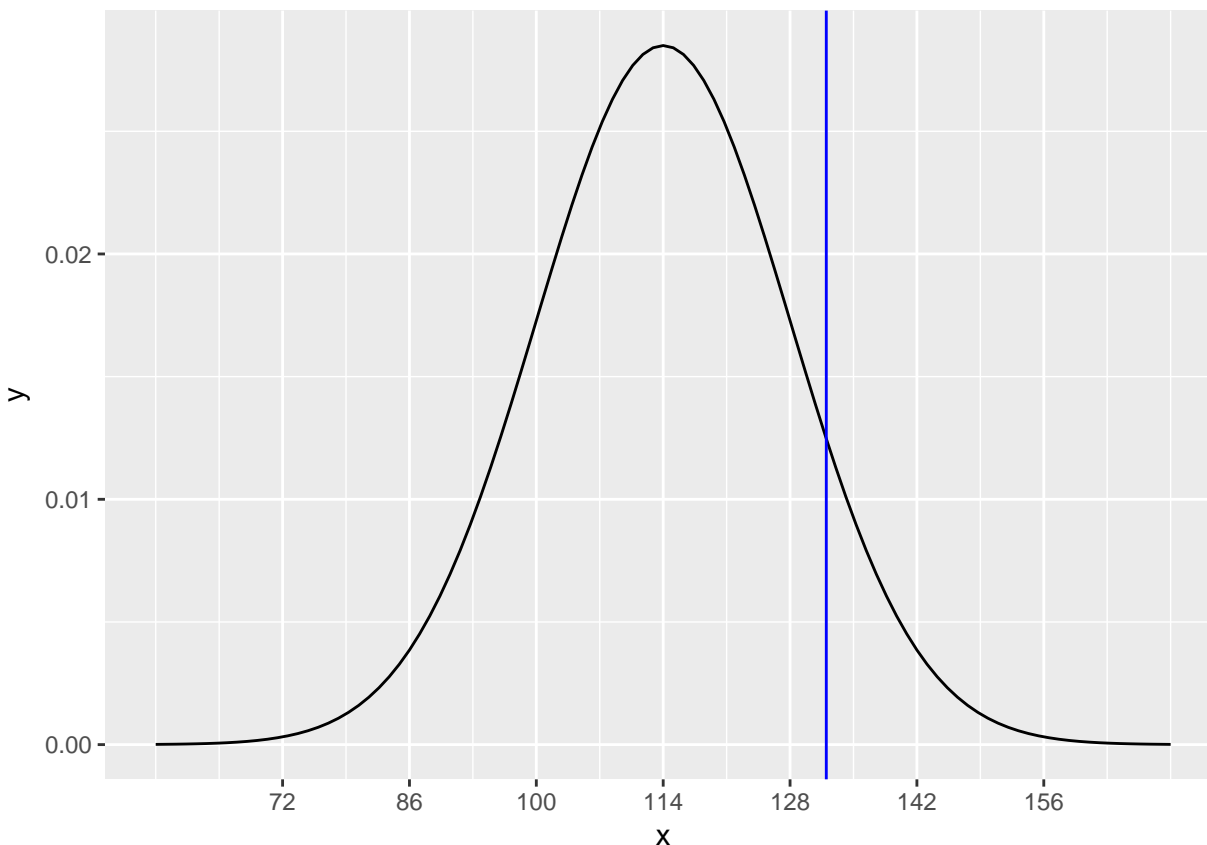
In our example, if we wanted to know the z-score for an SBP of 100, we just plug all the numbers into the formula above:

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

What is the z-score for an SBP of 132? Look at the graph of the normal model $N(\mu = 114, \sigma = 14)$:

```
# Don't worry about the syntax here.  
# You won't need to know how to do this on your own.  
ggplot(data.frame(x = c(58, 170)), aes(x)) +  
  stat_function(fun = dnorm, args = list(mean = 114, sd = 14)) +  
  scale_x_continuous(breaks = c(72, 86, 100, 114, 128, 142, 156)) +  
  geom_vline(xintercept = 132, color = "blue")
```

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



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

$$z = \frac{132 - 114}{14} = 1.285714.$$

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 extra crap we don't care about. This goes away if you

use the command inline when you knit to PDF. Go ahead and knit to PDF so you can see the effect in the following sentence:

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

ANSWER

Please write up your answer here.

- 102

ANSWER

Please write up your answer here.

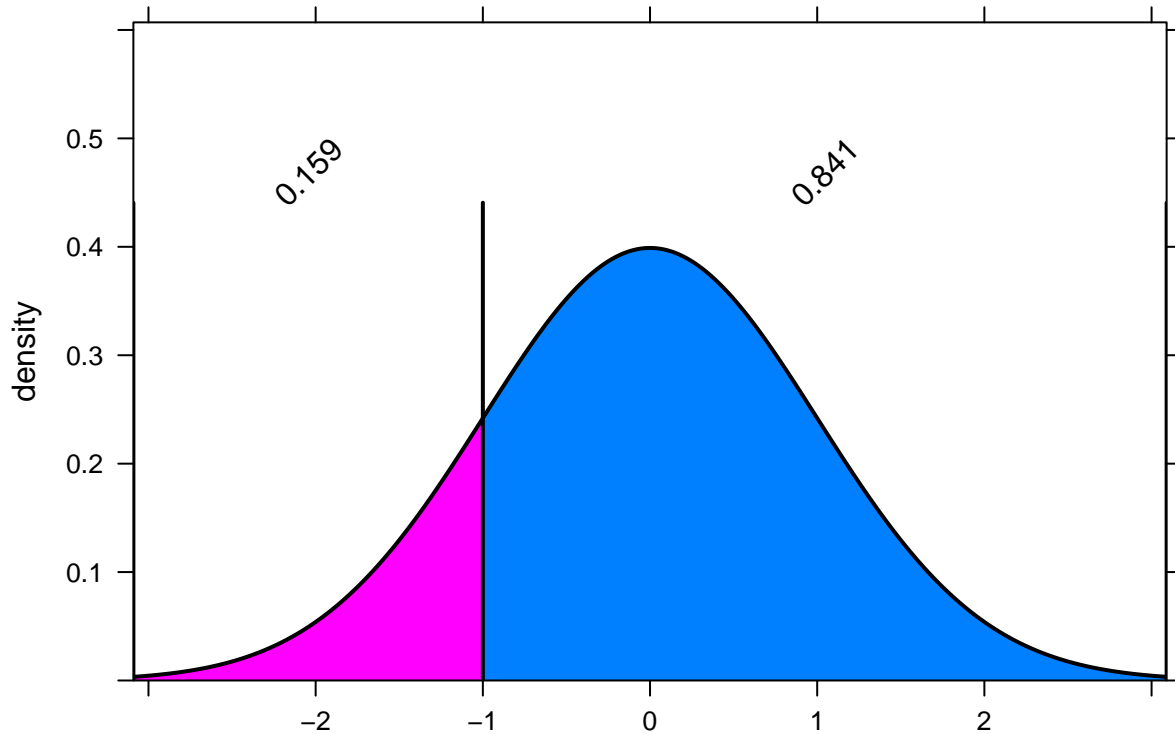
- 130

ANSWER

Please write up your answer here.

Working with z-scores also makes it easier to work with normal models. 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, invisible = TRUE)
```



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 your answer.

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 below 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.