DALITE Q2 - Boxplots, Standard Deviation and Normal Curves Solutions

EPIB607 - Inferential Statistics[a]

[a]Fall 2018, McGill University

This version was compiled on September 21, 2018

This DALITE quiz will cover more descriptives such as boxplots, standard deviation, and introduce you to normal density curves.

Boxplots | Standard deviation | Normal curves

1. Boxplots

```
library(mosaic)
library(tidyr)
dat <- data.frame(x = rep(1:6, each = 3), y = c(rep(c(1,3,4,6), each=2), rep(c(2,5), each = 5))) %>%
  gather(key = "variable", value = "count")



png("~/git_repositories/epib607/dalite/002_box_sd_curves/hist.png", res = 150)
histogram(~count | variable, data = dat, n = 6, type = "count", xlab = "integers")
dev.off()

bwplot(~count | variable, data = dat, xlab = "integers")

# arrange the two plots vertically
print(p1, position=c(0, .6, 1, 1), more=TRUE)
print(p2, position=c(0, 0, 1, .4))
```

1.1. Learning Objectives.

1. Recognize that a basic numerical description of a distribution requires both a measure of center and a measure of spread.
2. Use the quartiles and the extremes to provide information about the unequal spread in the two sides of a skewed distribution.
3. Be able to calculate the quartiles and give the five-number summary of a data set of using a computer.
4. Understand that boxplots provide less detail than stemplots or histograms but are especially useful for comparing several distributions.

1.2. Videos.

1. Against All Odds Unit 5

1.3. Required Readings.

1. Against All Odds Unit 5, pages 1-5

2. Standard Deviation

2.1. Learning Objectives.

1. Know that the sample standard deviation, s, is the measure of spread most commonly used when the mean, $\bar{x}$, is used as the measure of center.
2. Know the formula for the standard deviation s
3. Know the basic properties of the standard deviation:
   a) $s \geq 0$, and only when all data values are identical can $s = 0$
   b) s increases as the spread about x increases.
   c) s, like $\bar{x}$, is strongly influenced by outliers.
4. Know that the standard deviation is most useful for symmetric distributions and, in particular, for normal distributions.
5. Know that adding the same constant a to all the observations increases the value of $\bar{x}$ by a. However, adding the same constant a to all the observations does not change the value of s. That's because adding a constant a to all data values shifts the location of the data but does not affect its spread.
6. Know that multiplying all data values by a constant amount k changes $\bar{x}$ and s by a factor of k.

2.2. Videos.

1. Against All Odds Unit 6

2.3. Required Readings.

1. Against All Odds Unit 6, pages 1-8

3. Normal Curves

3.1. Learning Objectives.

1. Understand that the overall shape of a distribution of a large number of observations can be summarized by a smooth curve called a density curve.
2. Know that an area under a density curve over an interval represents the proportion of data that falls in that interval.
3. Recognize the characteristic bell-shapes of normal curves. Locate the mean and standard deviation on a normal density curve by eye.
4. Understand how changing the mean and standard deviation affects a normal density curve.

   - Know that changing the mean of a normal density curve shifts the curve along the horizontal axis without changing its shape.
   - Know that increasing the standard deviation produces a flatter and wider bell-shaped curve and that decreasing the standard deviation produces a taller and narrower curve.

3.2. Videos.

1. Against All Odds Unit 7

3.3. Required Readings.

1. Against All Odds Unit 7, pages 1-9