Homework 2

Jack Wright

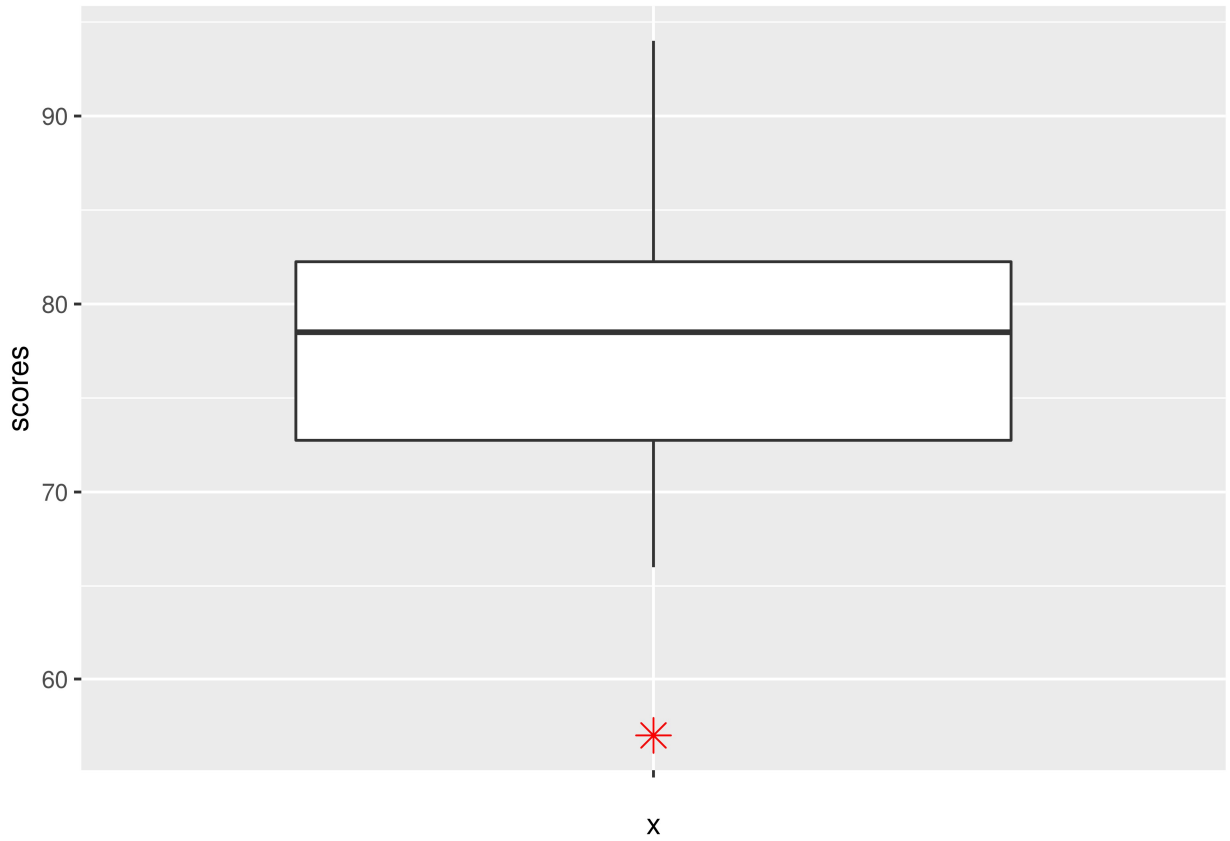
9/3/2020

```
library(tidyverse)
library(openintro)
library(dplyr)
```

Stats Scores

```
scores<-data.frame("scores"=c(57, 66, 69, 71, 72, 73, 74, 77, 78, 78, 79, 79, 81, 81, 82, 83, 83, 88, 8

ggplot(scores, aes(x="",y=scores))+geom_boxplot(outlier.color = "red", outlier.shape = 8, outlier.size =
```



Mix and Match

(a) unimodal, symmetric (2)
(b) multimodal (3)
(c) unimodal, skew right (1)

Distributions and appropriate statistics

(a) The distribution will be right skewed and the median would best represent a typical observation in the data, the mean will meaningfully change if even one 6MM house was replaced with a lower value house, while the median will not. IQR would be a better representation of the data as well because standard deviation also changes dramatically if one of the high value houses is changed in price.

(b) The distribution might be slightly right skewed, but looks fairly normal. I would not expect major differences between median and mean, or IQR and standard deviation if the data is normal.

(c) The distribution will be right skewed. for the same reasons as (a), median and IQR will be the best descriptors.

(d) The distribution will be right skewed. for the same reasons as (a), median and IQR will be the best descriptors.

Heart Transplants

(a) No, based on the mosaic plot you were more likely to survive longer getting the treatment.

(b) The boxplot suggests the treatment is effective in increasing survival from heart failure. The median and the IQR are higher for the treatment group.

(c) 88% of the control group died and 65% of the treatment group died.

(d) i. The claim being tested is whether this treatment is effective on gravely ill heart patients who seek treatement.

ii. We write alive on 28 cards representing patients who were alive at the end of the study, and dead on 75 cards representing patients who were not. Then, we shuffle these cards and split them into two groups: one group of size 69 representing treatment, and another group of size 34 representing control. We calculate the difference between the proportion of dead cards in the treatment and control groups (treatment

- control) and record this value. We repeat this 100 times to build a distribution centered at zero. Lastly, we calculate the fraction of simulations where the simulated differences in proportions are 23 % If this fraction is low, we conclude that it is unlikely to have observed such an outcome by chance and that the null hypothesis should be rejected in favor of the alternative.

iii. The results suggest that the 23% difference between our experiment and a random experiment are very low, so it is reasonable to think the transplant program is effective in this specific population.