Class Activity 3

Courtney Hodge

2024-07-18

```
library(tidyverse)
```

```
## ── Attaching core tidyverse packages ──────────────────── tidyverse 2.0.0 ──
## ✓ dplyr     1.1.3     ✓ readr     2.1.4
## ✓ forcats   1.0.0     ✓ stringr   1.5.0
## ✓ ggplot2   3.4.4     ✓ tibble    3.2.1
## ✓ lubridate 1.9.3     ✓ tidyr     1.3.0
## ✓ purrr     1.0.2
## ── Conflicts ─────────────────────────────────── tidyverse_conflicts() ──
## ✗ dplyr::filter() masks stats::filter()
## ✗ dplyr::lag()    masks stats::lag()
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to becom
e errors
```

```
data <- read.csv("C:\\Users\\hodge\\Downloads\\WA_Marketing-Campaign (1).csv")
```

1

```
ggplot(data, aes(x = as.character(Promotion), y = SalesInThousands, fill = Promotion)) +
  geom_boxplot()
```
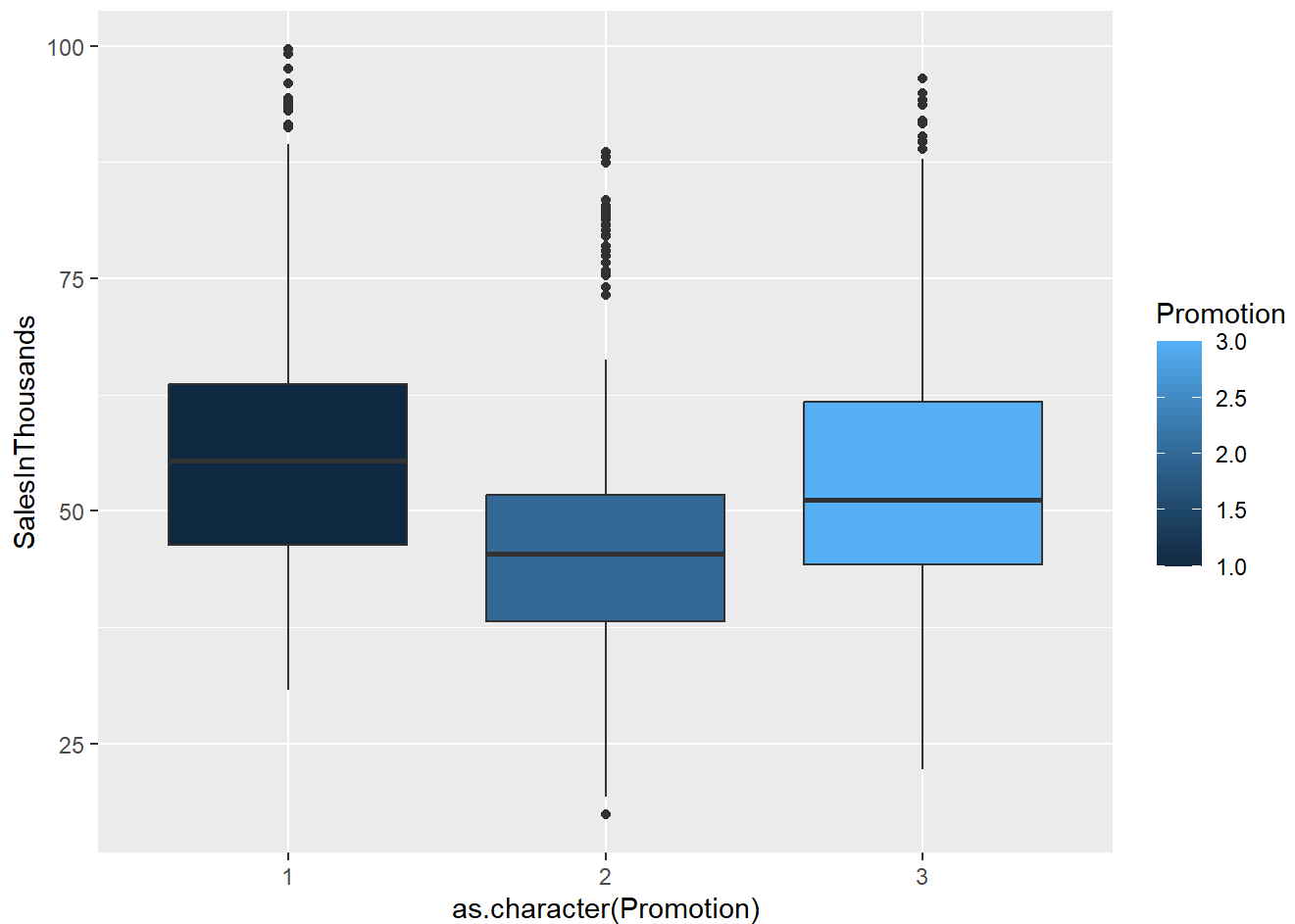
2

> The best way to conduct this investigation is to use anova method to compare more than two means.

3

$$H_0 : \mu_1 = \mu_2 = \mu_3$$

$$HA : \mu_i \neq \mu_j \text{ for some (i,j) pair}$$

> In words, the null hypothesis states that the average sales in thousands amount across the promotions should be equal to each other.

The alternative hypothesis states that at least one of the sale amounts across the promotion should not be equal to the other promotions.

4

```
anova <- aov(SalesInThousands~as.character(Promotion), data = data)

summary(anova)
```

```
##                            Df Sum Sq Mean Sq F value   Pr(>F)
## as.character(Promotion)    2  11449    5725   21.95 6.77e-10 ***
## Residuals                 545 142114     261
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Since the p-value = 6.77e-10 and it is less than the $\alpha = 0.05$, we reject the null hypothesis that all the average promotion sales in thousands will be equal to each other.

5

Based on our knowledge that the mean sales in thousands differs between the promotions, this leads to the question of which promotion yields the most sales on average. This means that we will need to do a Post Hoc test.