M04_activity

Hannah Valenty

2024-07-18

Task 1

```
library(dplyr)
```

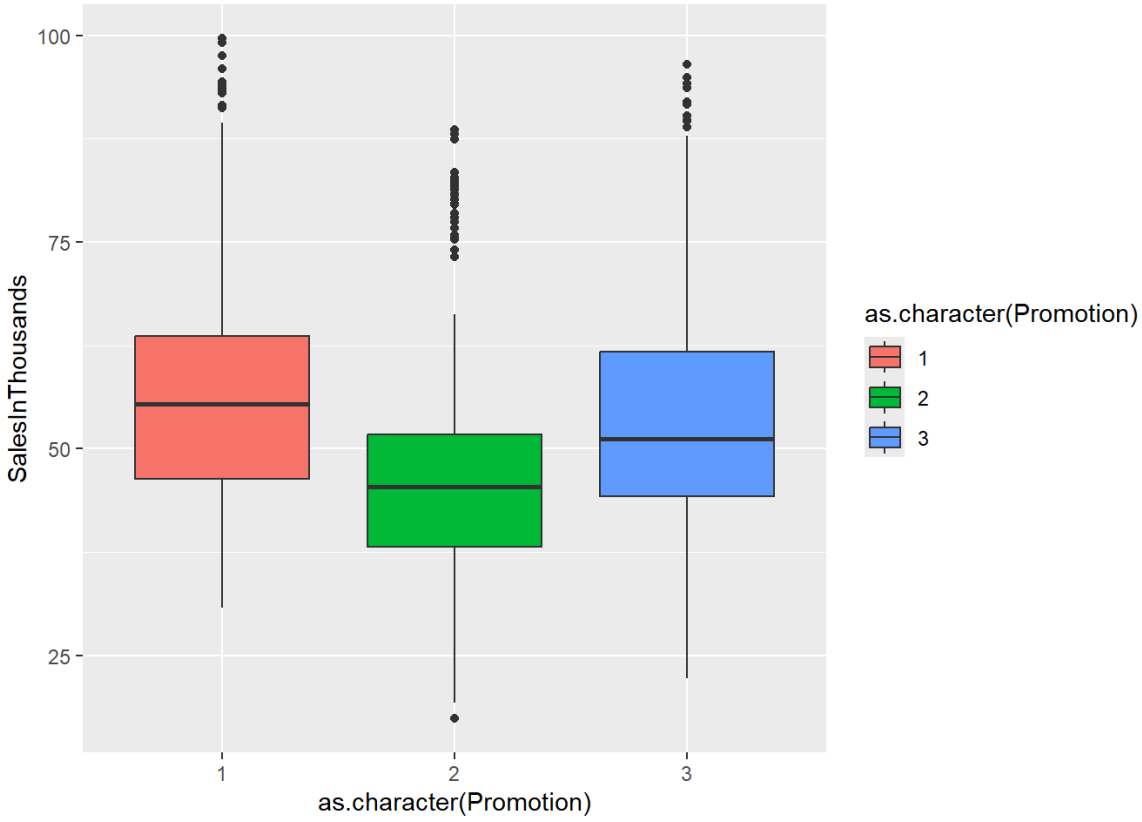
```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(ggplot2)

df <- read.csv('WA_Marketing-Campaign.csv')
ggplot(df, aes(x=as.character(Promotion), y=SalesInThousands, fill=as.character(Promotion)))+
  geom_boxplot()
```



Task 2

An ANOVA analysis should be used to compare the variance within the data and across the subsets of values.

Task 3

$H_o : \mu_1 = \mu_2 = \mu_3$ There is no significant difference between the means of the three promotions.

$H_a : \mu_i \neq \mu_j$ (for some i, j) At least one pair of mean sale values across the three promotions are not equal.

Task 4

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

p-value = 6.77e-10 < α = 0.05

We are 95% confident that one of the promotions has a different mean sales value than another and therefore we reject the null hypothesis.

Task 5

Yes, a post-hoc test would be useful in determining the promotion which yielded the most sales on average, and differed from another mean.