

Wondr Health Case Study - Statistical Analysis

Nicholas Edwards

2024-03-02

Reads in the relevant spreadsheet and puts either tab of data into its own data frame.

```
fct_weights <- read_xlsx(path = "CaseStudyDataSet2024.xlsx", sheet = "fct_weights")
dim_project_dates <- read_xlsx(path = "CaseStudyDataSet2024.xlsx", sheet = "dim_project_dates")
```

Changing the new feature column to be a character to prevent it from being understood as a continuous variable.

```
fct_weights$new_feature <- fct_weights$new_feature %>%
  as.character()

# Verified that the column was now a character instead of numeric
#is.character(fct_weights$new_feature)
```

Displays summary statistics for max_weight_loss for the the relevant groupings: new_feature and mobile.

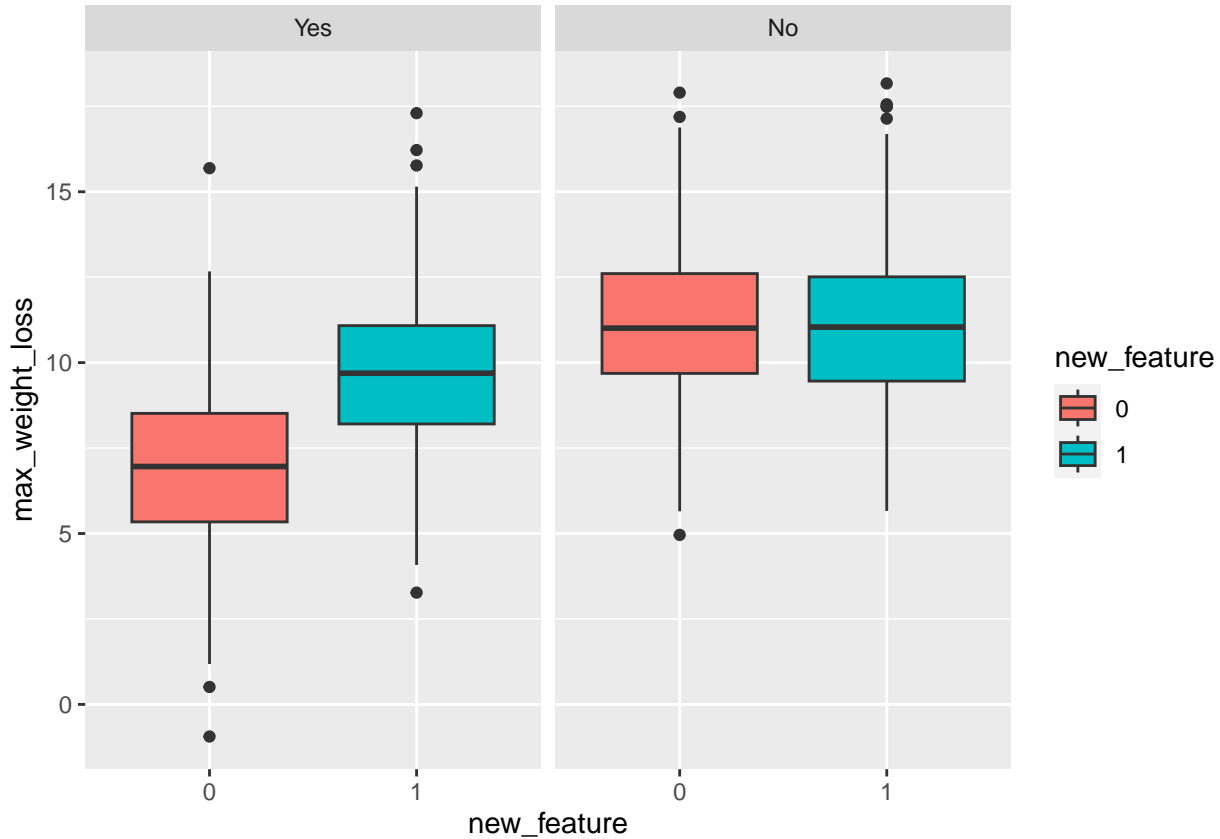
```
group_by(fct_weights, mobile, new_feature) %>%
  summarise(
    count = n(),
    mean = mean(max_weight_loss),
    max = max(max_weight_loss),
    min = min(max_weight_loss),
    sd = sd(max_weight_loss)
  )
```

```
## # A tibble: 4 x 7
## # Groups:   mobile [2]
##   mobile new_feature count  mean   max   min   sd
##   <chr>   <chr>     <int> <dbl> <dbl> <dbl> <dbl>
## 1 No     0           399  11.1  17.9  4.96  2.18
## 2 No     1           399  11.0  18.2  5.66  2.22
## 3 Yes    0           399   6.93  15.7 -0.94  2.32
## 4 Yes    1           400   9.68  17.3  3.27  2.21
```

Creates simple box plots to visualize the summary statistics on the relevant groupings: new_feature and mobile.

```
fct_weights %>%
  #Assigns new_feature as the discrete x var, max_weight_loss as the y var, and also adds the color (fi
  ggplot(aes(x = new_feature, y = max_weight_loss, group = new_feature, fill = new_feature)) +
```

```
#Specifies the plot type (i.e., box plot)
geom_boxplot() +
#Creates additional set of box plots by group for mobile and reorders (levels) to have Yes first and No second
facet_wrap(~ factor(mobile, levels = c("Yes", "No")))
```



Checking for normality in the two groups to be tested (needs to be $p > 0.05$).

```
fct_weights %>%
  with(shapiro.test(max_weight_loss[mobile == "Yes" & new_feature == "1"]))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  max_weight_loss[mobile == "Yes" & new_feature == "1"]
## W = 0.99732, p-value = 0.7646
```

```
fct_weights %>%
  with(shapiro.test(max_weight_loss[mobile == "Yes" & new_feature == "0"]))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  max_weight_loss[mobile == "Yes" & new_feature == "0"]
## W = 0.99766, p-value = 0.8531
```

As both of the p-values are > 0.05 , it can be assumed that the distribution of the data is not significantly different from a normal distribution.

Checking for difference in variance between the groups (of those using mobile, if they had the new feature or not) to perform t-test

```
#Creating new data frame to perform the test by filtering on only users on mobile
mobile_yes <- fct_weights %>%
  filter(mobile == "Yes")

#Executes the variance test
var.test(max_weight_loss ~ new_feature, mobile_yes)
```

```
##
## F test to compare two variances
##
## data: max_weight_loss by new_feature
## F = 1.1071, num df = 398, denom df = 399, p-value = 0.3102
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
## 0.9094772 1.3477105
## sample estimates:
## ratio of variances
## 1.107105
```

As the p-value is > 0.05 , it can be assumed that the variances of the two groups of data are not significantly different.

Determining if there is a statistically significant difference between the means of the new_feature groups of users on mobile

```
t.test(max_weight_loss ~ new_feature, data = mobile_yes, var.equal = TRUE)
```

```
##
## Two Sample t-test
##
## data: max_weight_loss by new_feature
## t = -17.16, df = 797, p-value < 2.2e-16
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
## -3.066387 -2.436854
## sample estimates:
## mean in group 0 mean in group 1
## 6.931955 9.683575
```

As the p-value is < 0.05 , the difference in the average weights between the two feature groups is considered to be a significant difference.