hypothesis_analysis

Jonathan Ndayisenga

2025-04-11

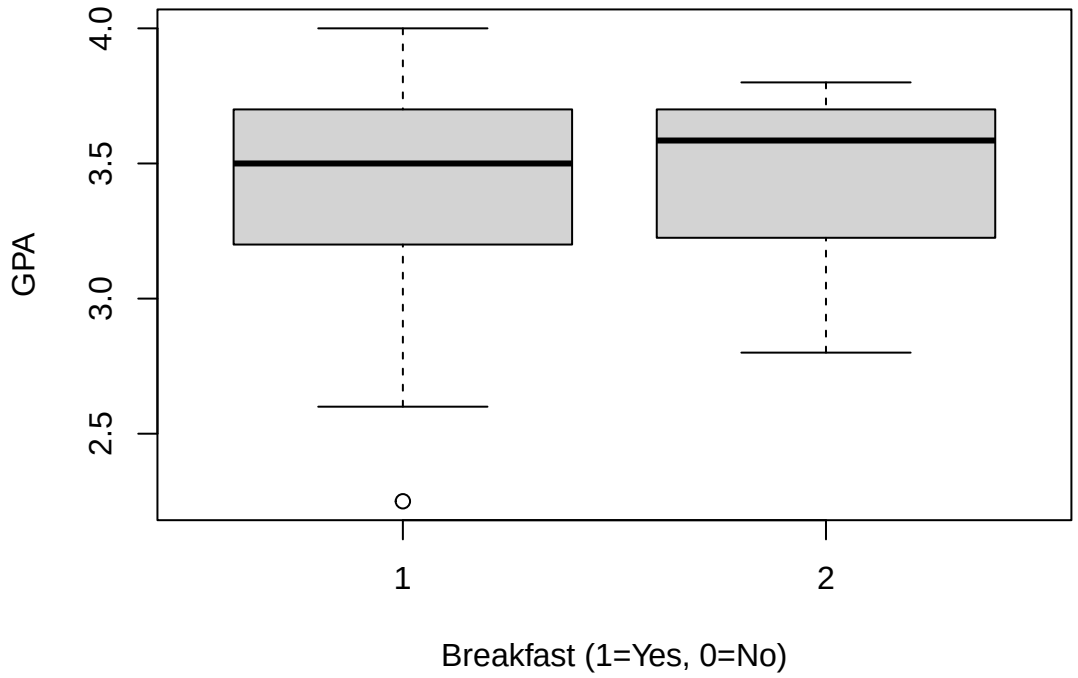
```
##LOAD DATA
data <- read_excel("cleaned_food_data.xlsx")
head(data)
```

```
## # A tibble: 6 x 61
##   GPA   Gender breakfast calories_chicken calories_day calories_scone coffee
##   <chr>  <dbl>     <dbl>            <dbl>        <dbl>          <dbl>  <dbl>
## 1 3.654      1         1              610            3            420      2
## 2 3.3        1         1              720            4            420      2
## 3 3.2        1         1              430            3            420      2
## 4 3.5        1         1              720            2            420      2
## 5 2.25       1         1              610            3            980      2
## 6 3.8        2         1              610            3            420      2
## # i 54 more variables: comfort_food <chr>, comfort_food_reasons <chr>,
## #   comfort_food_reasons_coded <chr>, cook <dbl>,
## #   comfort_food_reasons_coded.1 <dbl>, cuisine <chr>, diet_current <chr>,
## #   diet_current_coded <dbl>, drink <dbl>, eating_changes <chr>,
## #   eating_changes_coded <dbl>, eating_changes_coded1 <dbl>, eating_out <dbl>,
## #   employment <dbl>, ethnic_food <dbl>, exercise <chr>,
## #   father_education <dbl>, father_profession <chr>, fav_cuisine <chr>, ...
```

```
##Convert GPA to numeric
data$GPA <- as.numeric(as.character(data$GPA))
str(data$GPA)
```

```
##  num [1:84] 3.65 3.3 3.2 3.5 2.25 ...
```

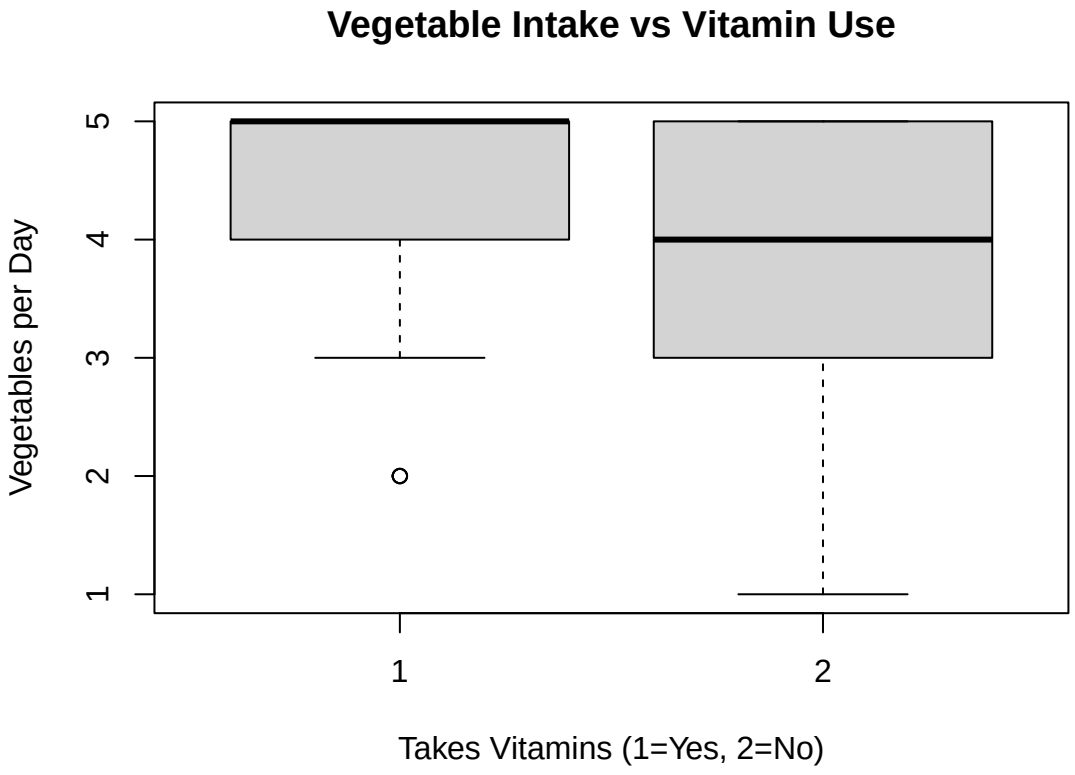
```
##hypothesis 1 GPA vs Breakfast

boxplot(GPA ~ breakfast, data = data,
        main = "GPA vs Breakfast",
        xlab = "Breakfast (1=Yes, 0=No)",
        ylab = "GPA")
```




```r
t.test(GPA ~ breakfast, data = data)
```

```
##
##  Welch Two Sample t-test
##
## data:  GPA by breakfast
## t = -0.051489, df = 8.6602, p-value = 0.9601
## alternative hypothesis: true difference in means between group 1 and group 2 is not equal to 0
## 95 percent confidence interval:
##  -0.3030780  0.2896671
## sample estimates:
## mean in group 1 mean in group 2
##        3.445795        3.452500
```

```r
##Hypothesis 2: Vegetables vs Vitamin use
data$veggies_day <- as.numeric(data$veggies_day)
boxplot(veggies_day ~ vitamins, data = data,
        main = "Vegetable Intake vs Vitamin Use",
        xlab = "Takes Vitamins (1=Yes, 2=No)",
        ylab = "Vegetables per Day")
```




```
t.test(veggies_day ~ vitamins, data = data)
```

```
##
##  Welch Two Sample t-test
##
## data:  veggies_day by vitamins
## t = 2.6539, df = 68.25, p-value = 0.00989
## alternative hypothesis: true difference in means between group 1 and group 2 is not equal to 0
## 95 percent confidence interval:
##  0.155028 1.094400
## sample estimates:
## mean in group 1 mean in group 2
##        4.282609        3.657895
```

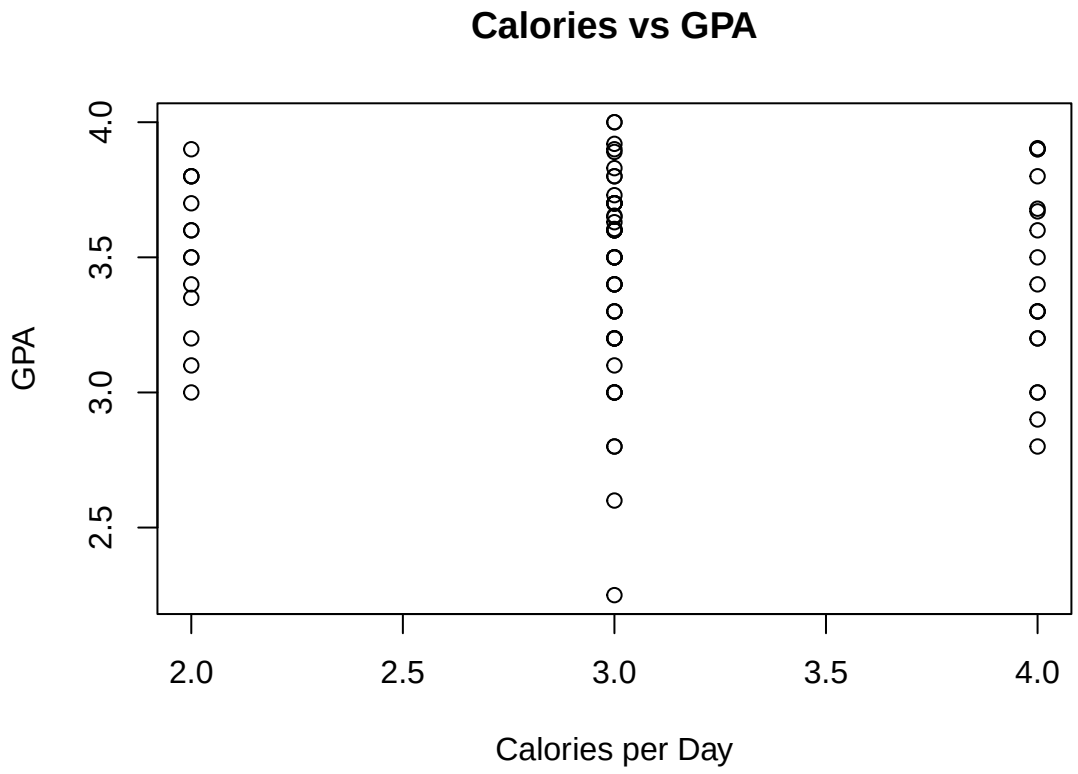
```
##Hypothesis 3:Gender vs Breakfast (Chi_square)
table(data$Gender, data$breakfast)
```

```
##
##      1  2
##   1 49  6
##   2 27  2
```

```
chisq.test(table(data$Gender, data$breakfast))
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  table(data$Gender, data$breakfast)
## X-squared = 0.041924, df = 1, p-value = 0.8378
```

```r
##Hypothesis 4: Calories vs GPA Corelation
plot(data$calories_day, data$GPA,
     main = "Calories vs GPA",
     xlab = "Calories per Day", ylab = "GPA")
```




```r
cor.test(data$calories_day, data$GPA, use = "complete.obs")
```

```
##
##  Pearson's product-moment correlation
##
## data:  data$calories_day and data$GPA
## t = -0.84619, df = 79, p-value = 0.4
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.3067762  0.1261854
## sample estimates:
##        cor
## -0.0947753
```

```r
##Hypothesis 5: Fruit vs Vegetable Intake (Correlation)
data$fruit_day <- as.numeric(data$fruit_day)
cor.test(data$fruit_day, data$veggies_day, use = "complete.obs")
```

```
##
##  Pearson's product-moment correlation
##
## data:  data$fruit_day and data$veggies_day
## t = 8.0714, df = 82, p-value = 5.046e-12
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.5260282 0.7699420
## sample estimates:
##       cor
## 0.6653829
```