

2021101113_Results_Visualisation_Assignment

Gowlapalli Rohit

19/2/2024

Contents

Attractiveness of humans post consumption of beer to mosquitoes.	1
Part A	1
Part B	4
Part C	5
Correlation between IQ and TESTSCORES	5

Attractiveness of humans post consumption of beer to mosquitoes.

Part A

```
library(readxl)
data <- read_excel("./BRSM_Results_Visualization.xlsx", sheet = 1)

print(data)
```

```
## # A tibble: 43 x 2
##   Group 'No. of Mosquitoes'
##   <chr>                <dbl>
## 1 Beer                  27
## 2 Beer                  19
## 3 Beer                  20
## 4 Beer                  20
## 5 Beer                  23
## 6 Beer                  17
## 7 Beer                  21
## 8 Beer                  24
## 9 Beer                  31
## 10 Beer                 26
## # i 33 more rows
```

```
beer <- subset(data, Group == "Beer")
print(beer)
```

```
## # A tibble: 25 x 2
##   Group 'No. of Mosquitoes'
##   <chr>           <dbl>
## 1 Beer             27
## 2 Beer             19
## 3 Beer             20
## 4 Beer             20
## 5 Beer             23
## 6 Beer             17
## 7 Beer             21
## 8 Beer             24
## 9 Beer             31
## 10 Beer            26
## # i 15 more rows
```

```
water <- subset(data, Group == "Water")
print(water$`No. of Mosquitoes`)
```

```
## [1] 21 19 13 22 15 22 15 22 20 12 24 24 21 19 18 16 23 20
```

```
beer_mean <- mean(beer$`No. of Mosquitoes`)
print(beer_mean)
```

```
## [1] 23.6
```

```
water_mean <- mean(water$`No. of Mosquitoes`)
print(water_mean)
```

```
## [1] 19.22222
```

```
test.stat.1 <- abs(water_mean - beer_mean)

beer_median <- median(beer$`No. of Mosquitoes`)
water_median <- median(water$`No. of Mosquitoes`)
test.stat.2 <- abs(beer_median - water_median)
print(test.stat.2)
```

```
## [1] 4
```

```
cat("The difference between the group medians is", test.stat.2)
```

```
## The difference between the group medians is 4
```

```
set.seed(1979)

n <- length(data$Group)
P <- 10000
var <- data$`No. of Mosquitoes`

PermSamples <- matrix(0, nrow = n, ncol = P)
```

```

for(i in 1:P) {
  PermSamples[, i] <- sample(var, size = n, replace = FALSE)
}

Perm.test.stat1 <- rep(0, P)
Perm.test.stat2 <- rep(0, P)

for (i in 1:P) {
  Perm.test.stat1[i] <- mean(PermSamples[data$Group == "Beer", i]) - mean(PermSamples[data$Group == "Wa
  Perm.test.stat2[i] <- median(PermSamples[data$Group == "Beer", i]) - median(PermSamples[data$Group ==
}

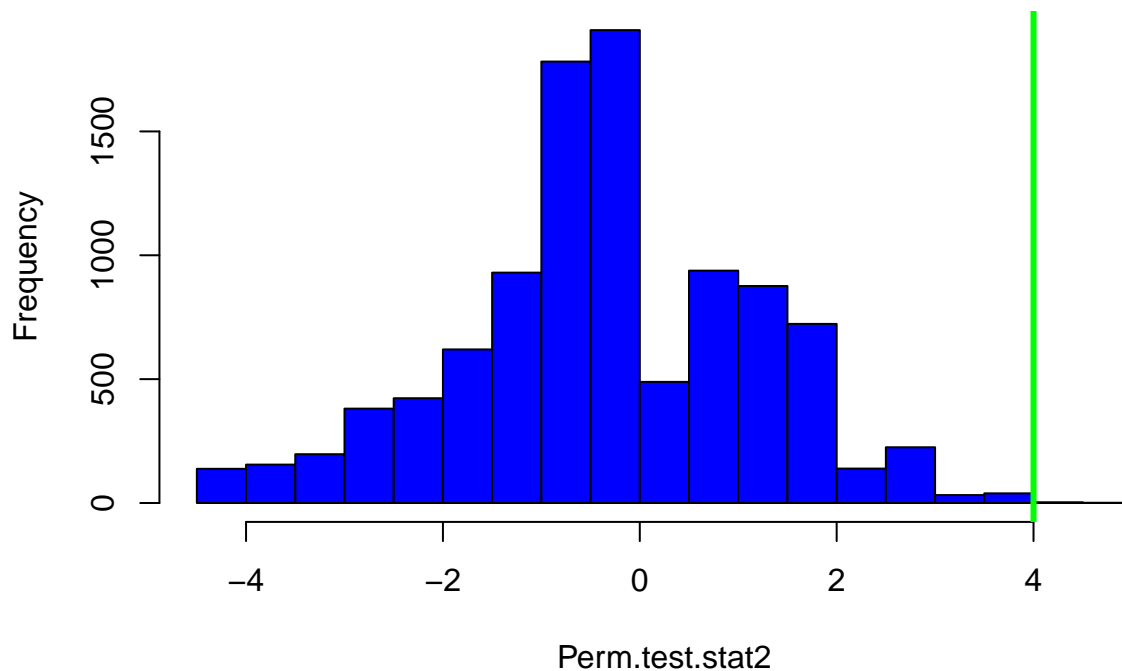
print(mean(Perm.test.stat2 >= test.stat.2))

## [1] 0.0043

hist(Perm.test.stat2, col = "blue", main = "Distribution of median differences from 10k random permutat.
abline(v = test.stat.2, lwd = 3, col = "green")

```

Distribution of median differences from 10k random permutations



```

p_value = sum(Perm.test.stat2 >= test.stat.2)/P
print(p_value)

```

```
## [1] 0.0043
```

```
cat("The p_value obtained is ", p_value)
```

```
## The p_value obtained is 0.0043
```

We note that the p-value is below 0.05, indicating a strong likelihood that the alternative directional hypothesis is true and that the observed statistic is statistically significant.

Part B

```
t_initial <- t.test(beer$`No. of Mosquitoes`, water$`No. of Mosquitoes`, var.equal = TRUE)
t_initial <- t_initial$statistic
cat("The initial t-score is ", t_initial, "\n")
```

```
## The initial t-score is 3.586984
```

```
P <- 10000
p2_perm <- rep(0, P)
for(i in 1:P) {
  temp_t <- t.test(head(PermSamples[, i], 25), tail(PermSamples[,i], 18), var.equal = TRUE)
  p2_perm[i] <- temp_t$statistic
}
p_value_2 <- sum(p2_perm >= t_initial) / P
cat("The observed p-value is ", p_value_2, "\n")
```

```
## The observed p-value is 3e-04
```

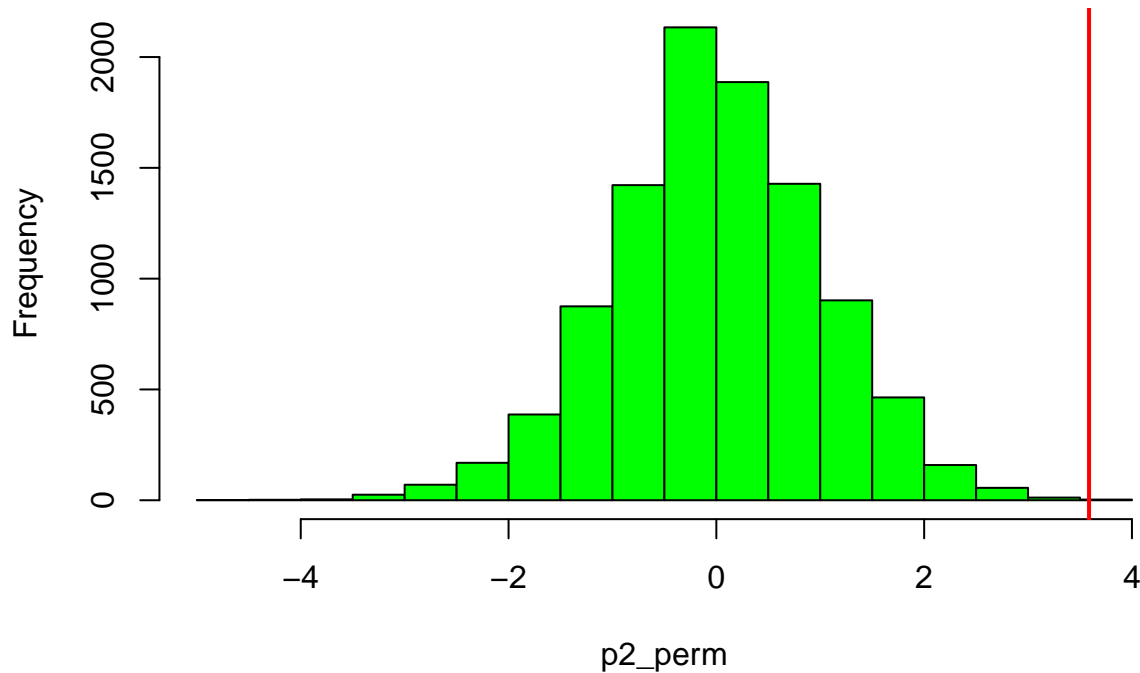
```
if(p_value_2 < 0.05) {
  cat("The obtained p-value is below 0.05, affirming the validity of the alternative hypothesis and indicating its significance")
} else {
  cat("The obtained p-value is not below 0.05, suggesting that the observed statistic may not hold significance")
}
```

```
## The obtained p-value is below 0.05, affirming the validity of the alternative hypothesis and indicating its significance
```

The observed p-value, being less than 0.05, validates the alternate hypothesis and signifies the significance of the observed statistic.

```
hist(p2_perm, col = "green", main = "Distribution of t-scores from 10k random permutations")
abline(v = t_initial, lwd = 2, col = "red")
```

Distribution of t-scores from 10k random permutations



Part C

As we compute the p-value under the assumption of a non-directional hypothesis, it's essential to calculate the p-value using the absolute differences of the medians.

```
p_value_significance = sum(abs(Perm.test.stat2) >= test.stat.2)/P
cat("The p-value is this case is ", p_value_significance)
```

```
## The p-value is this case is 0.0181
```

The p-value is noted to be below 0.05, suggesting the validity of the non-directional hypothesis and indicating significance in the observed statistic.

The statistical analyses performed consistently indicate that the observed data support the alternative hypothesis over the null hypothesis. The significance of the observed statistic, as indicated by the p-values, reinforces the validity of the conclusions drawn from the analyses. Moreover, the consideration of absolute differences of medians underscores the importance of appropriately framing hypotheses to accurately interpret the results of statistical tests.

Correlation between IQ and TESTSCORES

Following is the code for the third question.

```

set.seed(42)
library(readxl)
data <- read_excel("./IQ.xlsx", sheet = 1)
p_value <- cor(data$IQ, data$TESTSCORE)
print("The initial correlation value is ")

```

```
## [1] "The initial correlation value is "
```

```
print(p_value)
```

```
## [1] 0.4931479
```

```

n <- length(data$IQ)
P <- 10000
iq <- data$IQ
test_scores <- data$TESTSCORE

```

```

library(reshape2)
library(ggplot2)

```

```

df <- data.frame(iq = iq, score = test_scores)
corr <- cor(df)

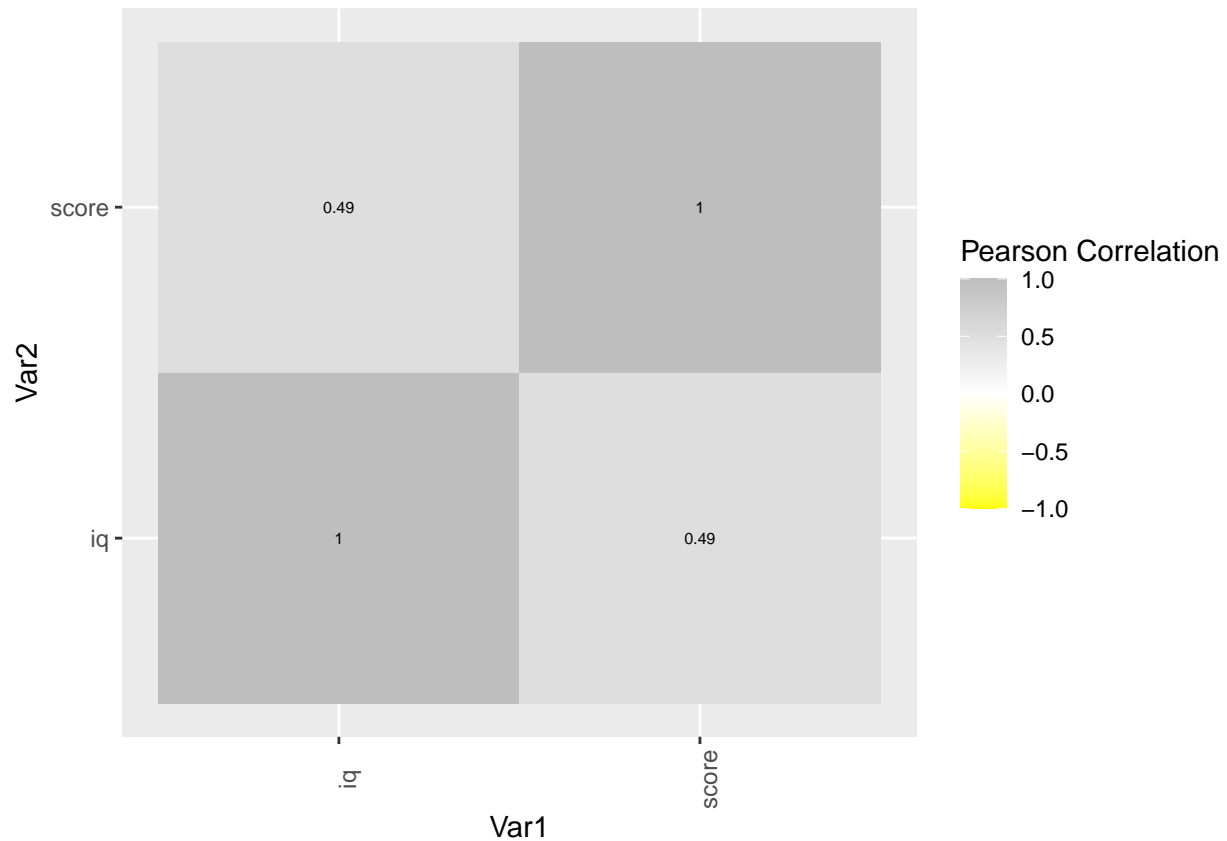
```

```
corr_melted <- melt(corr)
```

```

ggplot(data = corr_melted, aes(x = Var1, y = Var2, fill = value)) +
  geom_tile() +
  geom_text(aes(label = round(value, 2)), size = 2) +
  scale_fill_gradient2(low = "yellow", high = "gray", limit = c(-1, 1), name = "Pearson Correlation") +
  theme(axis.text.x = element_text(angle = 90))

```



```
iq_samples <- matrix(0, nrow = n, ncol = P)
test_scores_samples <- matrix(0, nrow = n, ncol = P)

for(i in 1:P) {
  iq_samples[, i] <- sample(iq, size = n, replace = FALSE)
  test_scores_samples[, i] <- sample(test_scores, size = n, replace = FALSE)
}
```

```
p3_perm <- rep(0, P)

for(i in 1:P) {
  p3_perm[i] <- cor(iq_samples[, i], test_scores_samples[, i])
}

p_value_statistic <- sum(p3_perm >= p_value) / P
print(p_value_statistic)
```

```
## [1] 0
```

```
print(paste("The observed p-value is ", p_value_statistic))
```

```
## [1] "The observed p-value is 0"
```

The obtained p-value from running the aforementioned code fluctuates between 0 and 1e-04. Currently, the p-value is 1e-04, yet it appears as 0 in the PDF document for unknown reasons.

Upon examination, we note that the p-value significantly falls below 0.05, based on our observed statistic. Hence, we confidently conclude that we can reject the null hypothesis, indicating a correlation between IQ and Test scores.

```
hist(p3_perm, col="green", main = "Distribution of correlation values from 10k random permutations", xlab="p3_perm", ylab="Frequency", lwd=2, col="red")
```

