

Final_Exam

Ashish Verma

2024-04-27

Problem 1

Solution(a)

Since population sigma is not given and sample size is less than 40, its paired T-test for means.

Null Hypothesis:H0:There is a significant change in WBC leukocytes due to surgery.

Alternate Hypothesis:Ha: There is no significant change in WBC leukocytes due to surgery.

```
presurgery <- c(10.80, 12.90, 9.59, 8.81, 12.00, 6.07)
postsurgery <- c(10.60, 16.60, 17.20, 14.00, 10.60, 8.60)

t_test_result <- t.test(postsurgery, presurgery, paired = TRUE, conf.level = 0.98)

cat("Test Statistic:", t_test_result$statistic, "\n")
```

```
## Test Statistic: 2.120477
```

```
cat("Degrees of Freedom:", t_test_result$parameter, "\n")
```

```
## Degrees of Freedom: 5
```

```
cat("p-value:", t_test_result$p.value, "\n")
```

```
## p-value: 0.08745276
```

```
alpha <- 0.02
if (t_test_result$p.value < alpha) {
  cat("Reject the null hypothesis. There is a significant change in WBC leukocytes due to surgery.\n")
} else {
  cat("Do not reject the null hypothesis. There is no significant change in WBC leukocytes due to surgery.\n")
}
```

```
## Do not reject the null hypothesis. There is no significant change in WBC leukocytes due to surgery.
```

Solution(b)

```
cat("98% Confidence Interval:", t_test_result$conf.int, "\n")
```

```
## 98% Confidence Interval: -1.70487 7.51487
```

Problem 2

Solution(a)

$$L(\theta; x_1, x_2, \dots, x_n) = \left(\frac{4}{\theta^3 \sqrt{\pi}} \right)^n \prod_{i=1}^n x_i^2 e^{-\frac{x_i^2}{\theta^2}}$$

Step 2: Log-Likelihood Function

$$\log L(\theta) = n \log 4 - \frac{3n}{2} \log \pi - 3n \log \theta + 2 \sum_{i=1}^n \log x_i - \frac{1}{\theta^2} \sum_{i=1}^n x_i^2$$

Step 3: Derivative of Log-Likelihood

$$\frac{\partial \log L(\theta)}{\partial \theta} = -\frac{3n}{\theta} + \frac{2}{\theta^3} \sum_{i=1}^n x_i^2 = 0$$

Step 4: Solve for θ

$$\begin{aligned} \theta^4 &= \frac{2}{3n} \sum_{i=1}^n x_i^2 \\ \theta &= \sqrt[4]{\frac{2}{3n} \sum_{i=1}^n x_i^2} \end{aligned}$$

Solution(b)

$$\begin{aligned} \bar{x} &= \frac{1}{n} \sum_{i=1}^n x_i \\ \bar{x} &= \frac{2\hat{\theta}}{\sqrt{\pi}} \\ \hat{\theta} &= \frac{\sqrt{\pi}}{2} \bar{x} \end{aligned}$$

Problem 3

Solution(a)

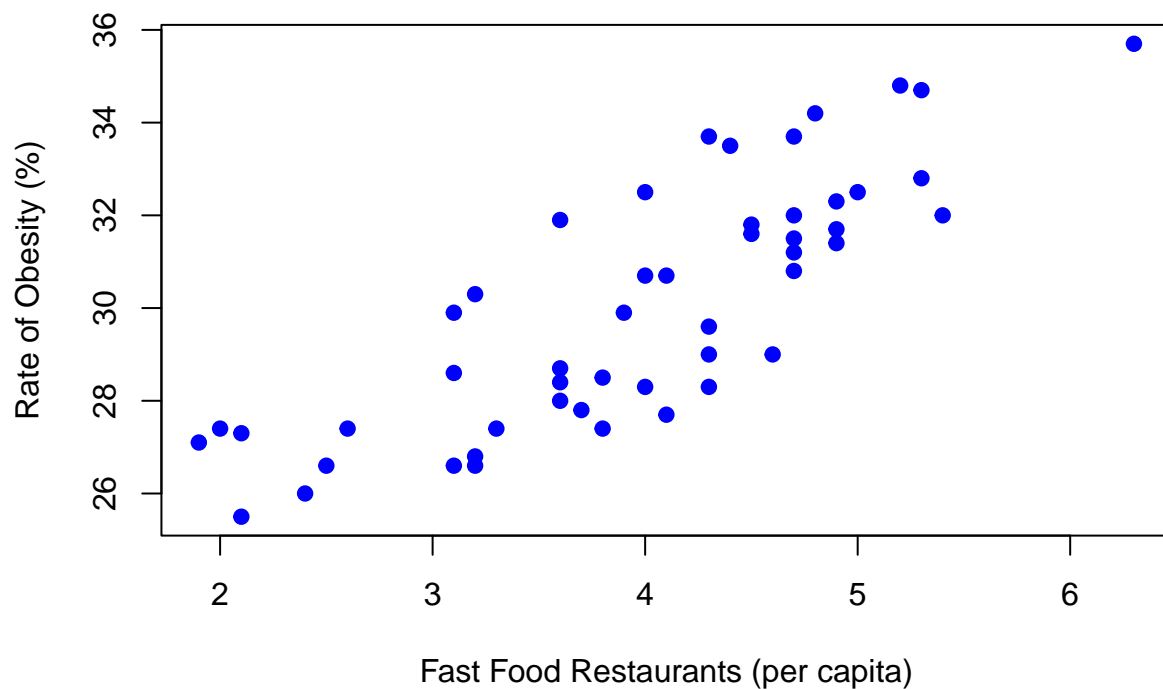
```
file_path <- "C:/Users/Ashish/Documents/ODU/STAT 604/Final Exam/fastfood.csv"
fastfood_data <- read.csv(file_path, header = TRUE)
head(fastfood_data)
```

```
##      State Restaurants Obesity
## 1   Alabama          6.3    35.7
## 2    Alaska          2.6    27.4
## 3   Arizona          4.6    29.0
## 4   Arkansas          4.7    33.7
## 5 California          3.6    28.0
## 6   Colorado          4.0    28.3
```

```
restaurants <- fastfood_data$Restaurants
obesity <- fastfood_data$Obesity

plot(restaurants, obesity,
      main = "Scatter Plot of Fast Food Restaurants vs. Obesity",
      xlab = "Fast Food Restaurants (per capita)",
      ylab = "Rate of Obesity (%)",
      pch = 19, # Solid circle for points
      col = "blue" # Color of the points
)
```

Scatter Plot of Fast Food Restaurants vs. Obesity



Direction: Positive as Fast Food Restaurants increases Obesity increases.

Strength: Somewhat strong relationship.

Solution(b)

```
cor_value <- cor(restaurants, obesity)
print(cor_value)
```

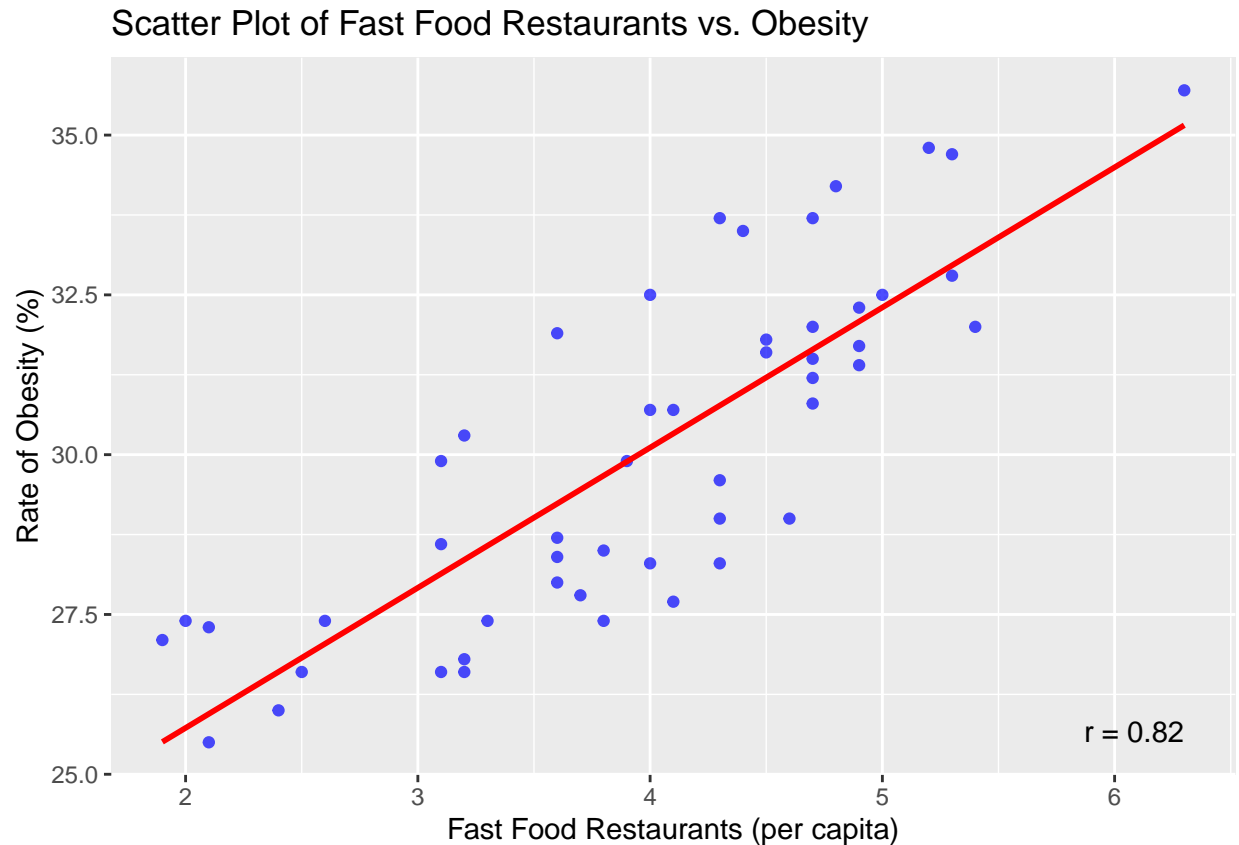
```
## [1] 0.820694
```

```
# Load ggplot2 library
library(ggplot2)

# Compute the Pearson correlation coefficient
cor_value <- cor(restaurants, obesity)

# Create a scatter plot with a regression line
ggplot(fastfood_data, aes(x = Restaurants, y = Obesity)) +
  geom_point(color = "blue", alpha = 0.7) + # Scatter plot
  geom_smooth(method = "lm", se = FALSE, color = "red") + # Linear regression line
  labs(title = "Scatter Plot of Fast Food Restaurants vs. Obesity",
        x = "Fast Food Restaurants (per capita)",
        y = "Rate of Obesity (%)") +
  annotate("text", x = max(restaurants), y = min(obesity),
           label = paste("r =", round(cor_value, 2)),
           hjust = 1, vjust = 0, size = 4, color = "black")
```

```
## 'geom_smooth()' using formula = 'y ~ x'
```



Solution(c)

```
linear_model <- lm(Obesity ~ Restaurants, data = fastfood_data)
intercept <- coef(linear_model)[1] # Intercept (beta_0)
slope <- coef(linear_model)[2] # Slope (beta_1)
cat("Intercept (beta_0):", intercept, "\n")

## Intercept (beta_0): 21.33986

cat("Slope (beta_1):", slope, "\n")

## Slope (beta_1): 2.192673

summary_result <- summary(linear_model)
r_squared <- summary_result$r.squared
cat("Coefficient of Determination (R-squared):", r_squared, "\n")

## Coefficient of Determination (R-squared): 0.6735386
```

A R^2 value of 0.67 suggests that the regression model captures a significant portion of the variance in the data, indicating a moderately strong fit.

Solution(d)

```
confint_slope <- confint(linear_model, level = 0.95)[2, ]
cat("95% Confidence Interval for the Slope:", confint_slope, "\n")
```

```
## 95% Confidence Interval for the Slope: 1.749655 2.635691
```

Solution(e)

Null Hypothesis: $H_0:\beta_1=2.0$. Alternate Hypothesis: $H_A:\beta_1\neq 2.0$

```
estimated_slope <- coef(summary_result)[2, "Estimate"]
standard_error <- coef(summary_result)[2, "Std. Error"]

degrees_of_freedom <- summary_result$df[2]
t_statistic <- (estimated_slope - 2.0) / standard_error
critical_t_value <- qt(1 - 0.005, df = degrees_of_freedom)
reject_null <- abs(t_statistic) > critical_t_value
cat("T-statistic:", t_statistic, "\n")
```

```
## T-statistic: 0.8744457
```

```
cat("Critical t-value:", critical_t_value, "\n")
```

```
## Critical t-value: 2.682204
```

```
cat("Reject the null hypothesis?", reject_null, "\n")
```

```
## Reject the null hypothesis? FALSE
```

Solution(f)

```
virginia_restaurants <- 4.3
predicted_obesity <- intercept + slope * virginia_restaurants
cat("Predicted rate of obesity for Virginia:", predicted_obesity, "%", "\n")
```

```
## Predicted rate of obesity for Virginia: 30.76835 %
```

Solution(g)

```
new_data <- data.frame(Restaurants = 4.3)
prediction <- predict(linear_model, new_data, interval = "confidence", level = 0.90)
cat("90% Confidence Interval:", prediction[1, c("lwr", "upr")], "\n")
```

```
## 90% Confidence Interval: 30.38583 31.15087
```

Solution(h)

```
new_data <- data.frame(Restaurants = 4.3)
prediction <- predict(linear_model, new_data, interval = "prediction", level = 0.90)
cat("90% Prediction Interval for the rate of obesity:", prediction[1, c("lwr", "upr")], "\n")
```

```
## 90% Prediction Interval for the rate of obesity: 28.17979 33.35692
```

Problem 4

Solution(a)

Null Hypothesis $H_0: \mu=72$ Alternate Hypothesis $H_A: \mu>72$ Since population sigma is given and sample size = 30 we can consider it as normal distribution.

```
sigma<-20
alpha<-0.05
mu<-72
xbar<-80
n<-30
ztesting<- ((mu -xbar)*sqrt(n))/sigma
zalpha<-qnorm(1 - alpha)
beta<- pnorm(zalpha+ztesting)
cat("The type-2 error is",beta)
```

```
## The type-2 error is 0.2925204
```

Solution(b)

```
sigma<-20
power<-0.95
beta_n<-1 - power
zalpha<-qnorm(1 - alpha)
zbeta<-qnorm(1 - beta_n)
sample_size <- (sigma*(zalpha+zbeta)/(mu - xbar))^2
cat("The sample size is given by",ceiling(sample_size))
```

```
## The sample size is given by 68
```

Problem 5

Solution(a)

```

library(stats)
data <- data.frame(
  Dosage = rep(c(20, 30, 40), each = 4),
  Observations = c(24, 28, 37, 30, 37, 44, 31, 35, 42, 47, 52, 38)
)
result <- aov(Observations ~ Dosage, data = data)
anova_results <- summary(result)
anova_table <- data.frame(
  "Source" = rownames(anova_results[[1]]),
  "DF" = anova_results[[1]]$Df,
  "Sum Sq" = anova_results[[1]]$"Sum Sq",
  "Mean Sq" = anova_results[[1]]$"Mean Sq",
  "F Value" = anova_results[[1]]$"F value",
  "Pr(>F)" = anova_results[[1]]$"Pr(>F)"
)

print(anova_table)

```

```

##           Source DF   Sum.Sq  Mean.Sq  F.Value    Pr..F.
## 1 Dosage         1 450.0000 450.00000 15.57543 0.002745488
## 2 Residuals     10 288.9167  28.89167      NA         NA

```

Solution(b)

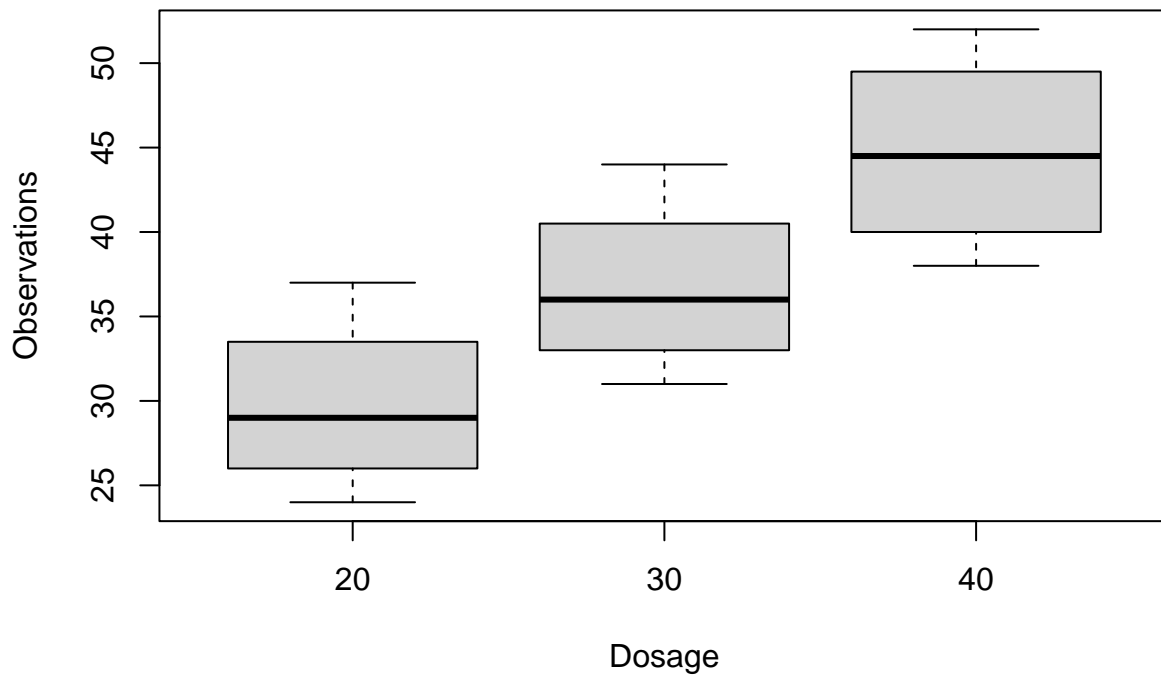
Check the assumption of homogeneity of variances (homoscedasticity) among the groups.

```

alpha<-0.01
data <- data.frame(
  Dosage = rep(c(20, 30, 40), each = 4),
  Observations = c(24, 28, 37, 30, 37, 44, 31, 35, 42, 47, 52, 38)
)

data$Dosage <- as.factor(data$Dosage)
boxplot(Observations ~ Dosage, data = data, xlab = "Dosage", ylab = "Observations")

```

```
bartlett_test <- bartlett.test(Observations ~ Dosage, data = data)
aov_model <- aov(Observations ~ Dosage, data = data)
p_value <- bartlett_test$p.value

cat("Pvalue", p_value)
```

```
## Pvalue 0.9783861
```

```
if (p_value > 0.01) {
  tukey_results <- TukeyHSD(aov_model, conf.level = 0.99)
  print(tukey_results)
} else {
  print("Tukey test is not recommended due to unequal variances")
}
```

```
## Tukey multiple comparisons of means
## 99% family-wise confidence level
##
## Fit: aov(formula = Observations ~ Dosage, data = data)
##
## $Dosage
##      diff      lwr      upr    p adj
## 30-20    7 -8.3594856 22.35949 0.2402975
## 40-20   15 -0.3594856 30.35949 0.0114434
## 40-30    8 -7.3594856 23.35949 0.1680265
```

Problem 6

Solution(a)

p_1 = the proportion of females in the U.S. that believe in miracles

p_2 = the proportion of males in the U.S. that believe in miracles

```
n<-200
pf<-112/n
qf<-1 - pf
pm<-90/n
qm<-1 - pm
alpha<-1 - 0.97
zalpha<- qnorm(1 - (alpha/2))
sum1<- (pf*qf/n)
sum2<-(pm*qm/n)
margin<- sqrt(sum1+sum2)
lower_limit<- (pf -pm)- (zalpha*margin)
upper_limit<- (pf -pm)+ (zalpha*margin)
cat("The 97% CI is given by",lower_limit,upper_limit)
```

```
## The 97% CI is given by 0.00215939 0.2178406
```

Solution(b)

Null Hypothesis (H_0): The proportion of women who believe in miracles is equal to or less than the proportion of men who believe in miracles $p_1 \leq p_2$

Alternative Hypothesis (H_A): The proportion of women who believe in miracles is greater than the proportion of men who believe in miracles. $p_1 > p_2$

```
alpha<-0.03
n<-200
pf<-112/n
qf<-1 - pf
pm<-90/n
qm<-1 - pm
pprime <- (pf+pm)/2 #sample size same
qprime<-1 - pprime

SE <- sqrt(pprime*qprime *(1 / n + 1 / n))
ztesting<-(pf - pm)/SE
cat("Test statistics",ztesting)
```

```
## Test statistics 2.20011
```

```
zalpha<- qnorm(1 - alpha)
cat("Critical region",zalpha)
```

```
## Critical region 1.880794
```

```
if (ztesting >= zalpha) {  
  decision <- "Reject the null hypothesis"  
} else {  
  decision <- "Fail to reject the null hypothesis"  
}  
cat(decision)
```

```
## Reject the null hypothesis
```