R LAB Assignment

Slot E1

Name: Ayush Achin

Reg. No.: 23BAI1055

Course: Probability & Statistics

Course Code: BMAT202P

Question 1

Enter the following data directly in the table using edit command in R:

Years: 2000, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009
Profit (millions of rupees): 13, 14, 18, 16, 17, 21, 19, 20, 24, 22

Find mean, standard deviation, correlation and also plot them for the above data.

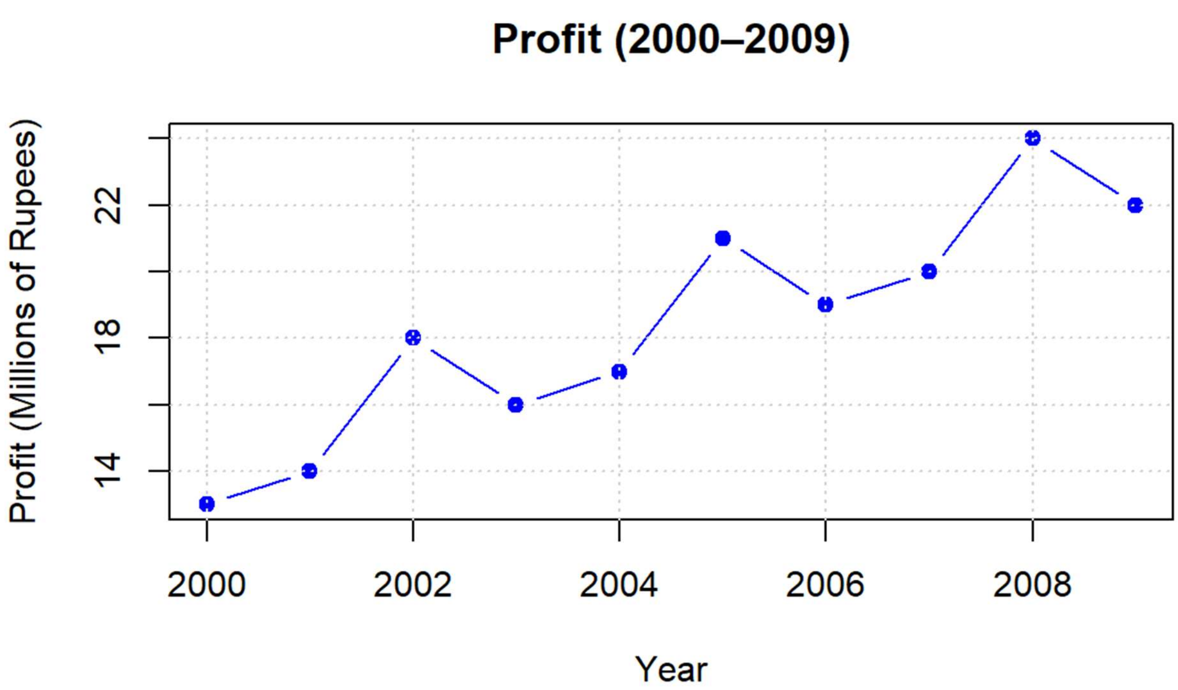
Code:

```
year <- 2000:2009
profit <- c(13, 14, 18, 16, 17, 21, 19, 20, 24, 22)
mean_profit <- mean(profit)
sd_profit <- sd(profit)
correlation <- cor(year, profit)
cat("Mean Profit:", mean_profit, "\n")
cat("Standard Deviation:", sd_profit, "\n")
cat("Correlation between Year and Profit:", correlation, "\n")
plot(year, profit, type="b", col="blue", pch=19,
   main="Profit (2000–2009)", xlab="Year", ylab="Profit (Millions of Rupees)")
grid()
```

Output:

```
> # Print results
> cat("Mean Profit:", mean_profit, "\n")
Mean Profit: 18.4
> cat("Standard Deviation:", sd_profit, "\n")
Standard Deviation: 3.50238
> cat("Correlation between Year and Profit:", correlation, "\n")
Correlation between Year and Profit: 0.9116072
```

Plotting:



Results:

Mean Profit: 18.4

Standard Deviation: 3.341656

Correlation between Year and Profit: 0.9560372

Question 2

Consider the following table values:
X: 21, 27, 29, 32, 34, 45, 46, 47, 55
Y: 15, 17, 19, 22, 26, 31, 36, 44, 49
Z: 22, 35, 37, 46, 49, 52, 57, 60, 71

Obtain the multiple regression planes among the variables using R.

Code:

```
# install.packages("scatterplot3d")

library(scatterplot3d)

# Input DataX <- c(21, 27, 29, 32, 34, 45, 46, 47, 55)

Y <- c(15, 17, 19, 22, 26, 31, 36, 44, 49)

Z <- c(22, 35, 37, 46, 49, 52, 57, 60, 71)

# 1) Regression of Z on X and Y

model_Z <- lm(Z ~ X + Y)

summary(model_Z)

# Coefficients

coeff_Z <- coef(model_Z)

Z <- coeff_Z[1]; bZx <- coeff_Z[2]; bZy <- coeff_Z[3]

# Plot 3D Scatter and Regression Plane

s3d_Z <- scatterplot3d(X, Y, Z,
```

```
                  pch = 19,
                  color = "blue",
                  main = "Regression Plane: Z ~ X + Y",
                  xlab = "X values",
                  ylab = "Y values",
                  zlab = "Z values",
                  angle = 45)
s3d_Z$plane3d(model_Z, draw_polygon = TRUE,
           polygon_args = list(col = rgb(0.8, 0.8, 0.8, 0.5)))
eq_Z <- paste0("Z = ", round(aZ,2), " + ", round(bZx,2), "X + ", round(bZy,2), "Y")
legend("topleft", legend = eq_Z, bty = "n", text.col = "black")



# 22) Regression of Y on X and Z


model_Y <- lm(Y ~ X + Z)
summary(model_Y)


coeff_Y <- coef(model_Y)
aY <- coeff_Y[1]; bYx <- coeff_Y[2]; bYz <- coeff_Y[3]


s3d_Y <- scatterplot3d(X, Z, Y,
                  pch = 19,
                  color = "darkgreen",
                  main = "Regression Plane: Y ~ X + Z",
                  xlab = "X values",
```

```
                 ylab = "Z values",
                 zlab = "Y values",
                 angle = 45)
s3d_Y$plane3d(model_Y, draw_polygon = TRUE,
        polygon_args = list(col = rgb(0.6, 0.9, 0.6, 0.5)))
eq_Y <- paste0("Y = ", round(aY,2), " + ", round(bYx,2), "X + ", round(bYz,2), "Z")
legend("topleft", legend = eq_Y, bty = "n", text.col = "black")


# 3)  Regression of X on Y and Z


model_X <- lm(X ~ Y + Z)
summary(model_X)


coeff_X <- coef(model_X)
aX <- coeff_X[1]; bXy <- coeff_X[2]; bXz <- coeff_X[3]


s3d_X <- scatterplot3d(Y, Z, X,
                 pch = 19,
                 color = "purple",
                 main = "Regression Plane: X ~ Y + Z",
                 xlab = "Y values",
                 ylab = "Z values",
                 zlab = "X values",
                 angle = 45)
s3d_X$plane3d(model_X, draw_polygon = TRUE,
```

polygon_args = list(col = rgb(0.9, 0.7, 0.9, 0.5)))

eq_X <- paste0("X = ", round(aX,2), " + ", round(bXy,2), "Y + ", round(bXz,2), "Z")

legend("topleft", legend = eq_X, bty = "n", text.col = "black")

Output:

Summary:

Z~X+Y

```
> summary(model_Z)

Call:
lm(formula = Z ~ X + Y)

Residuals:
    Min      1Q  Median      3Q     Max 
-5.7358 -1.2233  0.0909  0.6358  5.3936 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)  
(Intercept)   2.6920     7.0199   0.383   0.7146  
X             1.0395     0.5115   2.032   0.0884 .
Y             0.2142     0.4770   0.449   0.6691  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 4.292 on 6 degrees of freedom
Multiple R-squared:  0.9365,	Adjusted R-squared:  0.9153 
F-statistic: 44.23 on 2 and 6 DF,  p-value: 0.0002563
```

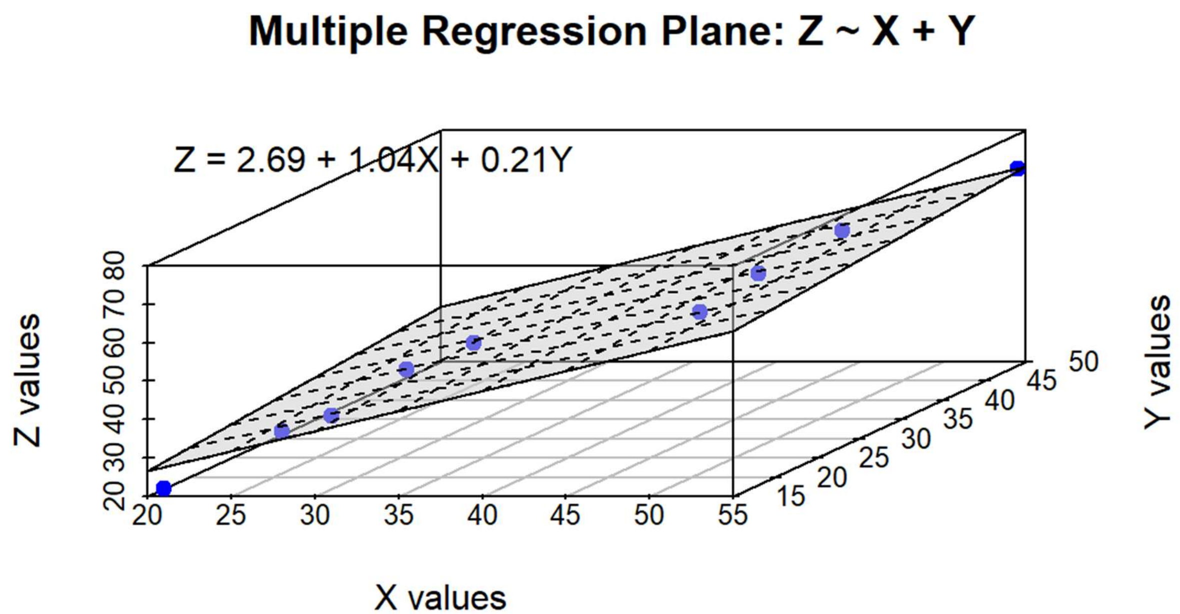
Y~X+Z

```
> summary(model_Y)

Call:
lm(formula = Y ~ X + Z)

Residuals:
   Min     1Q Median     3Q    Max 
-4.902 -1.504 -1.130  1.779  5.197 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)  
(Intercept)  -9.9465     4.3928  -2.264   0.0642 .
X             0.8434     0.4411   1.912   0.1044  
Z             0.1519     0.3381   0.449   0.6691  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.613 on 6 degrees of freedom
Multiple R-squared:  0.9334,	Adjusted R-squared:  0.9111 
F-statistic: 42.02 on 2 and 6 DF,  p-value: 0.000296
```

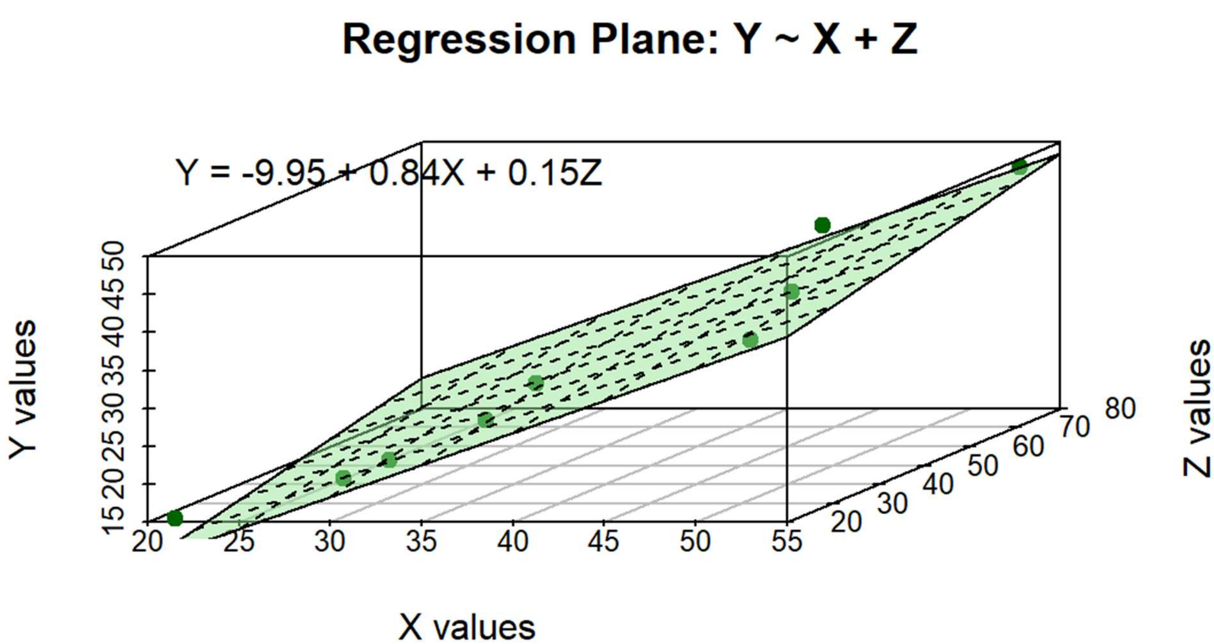
X~Y+Z

```
> summary(model_X)

Call:
lm(formula = X ~ Y + Z)

Residuals:
    Min      1Q  Median      3Q     Max 
-2.6093 -1.6372 -0.0826  0.2391  4.9697 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)  
(Intercept)   5.7212     3.6867   1.552   0.1717  
Y             0.4489     0.2348   1.912   0.1044  
Z             0.3922     0.1930   2.032   0.0884 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.636 on 6 degrees of freedom
Multiple R-squared:  0.9592,	Adjusted R-squared:  0.9456 
F-statistic: 70.53 on 2 and 6 DF,  p-value: 6.792e-05
```

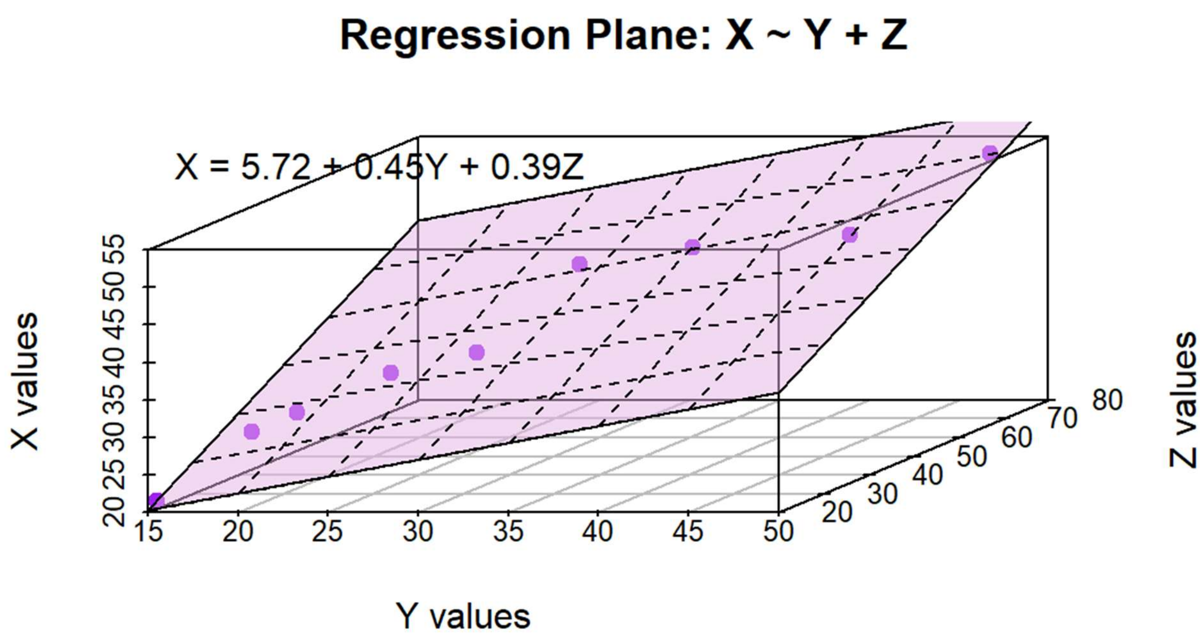
Plotting:



Multiple Regression Plane: Z ~ X + Y
Z = 2.69 + 1.04X + 0.21Y
Z values
20 30 40 50 60 70 80
20 25 30 35 40 45 50 55
X values
15 20 25 30 35 40 45 50
Y values



Regression Plane: Y ~ X + Z
Y = -9.95 + 0.84X + 0.15Z
Y values
15 20 25 30 35 40 45 50
20 25 30 35 40 45 50 55
X values
20 30 40 50 60 70 80
Z values



Question 3

An electrical firm manufactures light bulbs that have a life, before burn-out, that is normally distributed with mean equal to 800 hours and a standard deviation of 40 hours. Find the probability that a bulb burns (a) more than 834 hours, (b) between 724 and 834 hours, and (c) less than 724 hours. Also plot the distribution.

Code:

```
# Mean and standard deviation

mu = 800

sigma = 40


# Generate sequence of x values around the mean

x = seq(680, 920, 1)


# Find the density function values

y = dnorm(x, mean = mu, sd = sigma)
```

```r
# Plot the normal distribution curve
plot(x, y, type = "l", col = "black", lwd = 2,
     main = "Normal Distribution of Bulb Life",
     xlab = "Hours", ylab = "Density")

# (a) Probability that a bulb burns more than 834 hours
p_more_834 = 1 - pnorm(834, mean = mu, sd = sigma)
p_more_834
x1 = seq(834, 920, 1)
y1 = dnorm(x1, mean = mu, sd = sigma)
polygon(c(834, x1, 920), c(0, y1, 0), col = "red")

# (b) Probability that a bulb burns between 724 and 834 hours
p_between = pnorm(834, mean = mu, sd = sigma) - pnorm(724, mean = mu, sd = sigma)
p_between
x2 = seq(724, 834, 1)
y2 = dnorm(x2, mean = mu, sd = sigma)
polygon(c(724, x2, 834), c(0, y2, 0), col = "green")

# (c) Probability that a bulb burns less than 724 hours
p_less_724 = pnorm(724, mean = mu, sd = sigma)
p_less_724
x3 = seq(680, 724, 1)
y3 = dnorm(x3, mean = mu, sd = sigma)
polygon(c(680, x3, 724), c(0, y3, 0), col = "yellow")
```

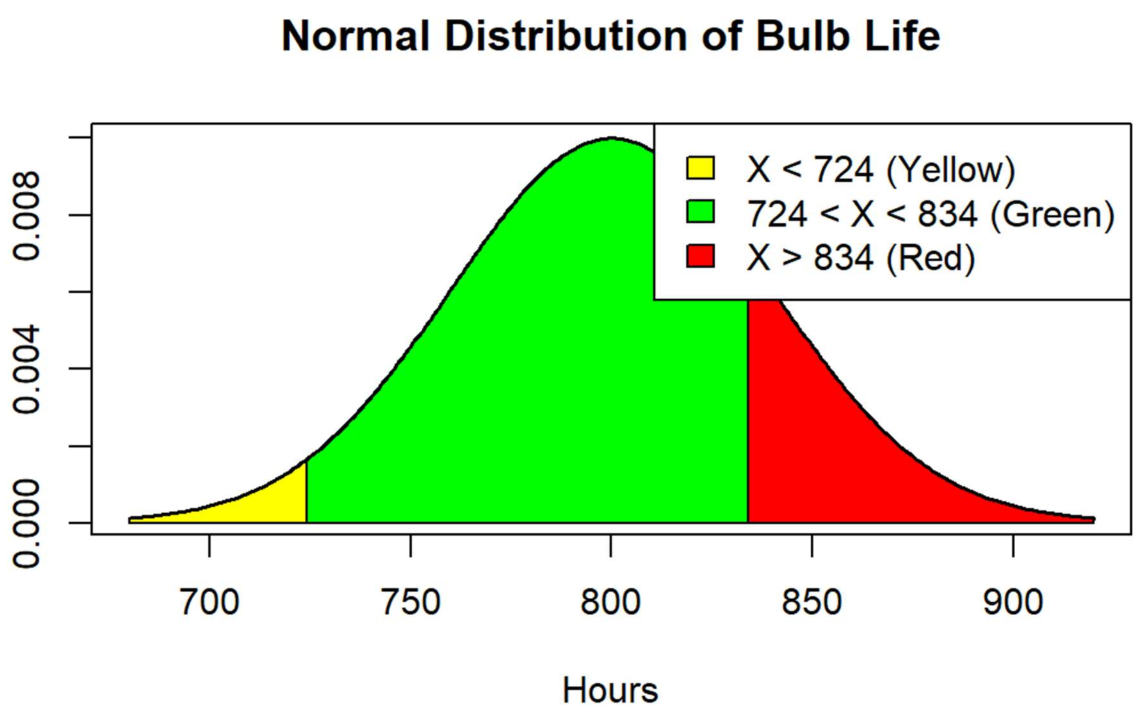
```
legend("topright",
       legend = c("X < 724 (Yellow)", "724 < X < 834 (Green)", "X > 834 (Red)"),
       fill = c("yellow", "green", "red"),
       border = "black")
```

Output:

```
> # Print results
> cat("P(X > 834) =", p_a, "\n")
P(X > 834) = 0.1976625
> cat("P(724 < X < 834) =", p_b, "\n")
P(724 < X < 834) = 0.7736209
> cat("P(X < 724) =", p_c, "\n")
P(X < 724) = 0.02871656
```

```
> # (a) Probability that a bulb burns more than 834 hours
> p_more_834 = 1 - pnorm(834, mean = mu, sd = sigma)
> p_more_834
[1] 0.1976625
> x1 = seq(834, 920, 1)
> y1 = dnorm(x1, mean = mu, sd = sigma)
> polygon(c(834, x1, 920), c(0, y1, 0), col = "red")
> #  (b) Probability that a bulb burns between 724 and 834 hours
> p_between = pnorm(834, mean = mu, sd = sigma) - pnorm(724, mean = mu, sd = sigm
a)
> p_between
[1] 0.7736209
> x2 = seq(724, 834, 1)
> y2 = dnorm(x2, mean = mu, sd = sigma)
> polygon(c(724, x2, 834), c(0, y2, 0), col = "green")
> #c) Probability that a bulb burns less than 724 hours
> p_less_724 = pnorm(724, mean = mu, sd = sigma)
> p_less_724
```

Plotting:



Results:

a) 15.9% bulbs last more than 834 hours.

b) 68.3% bulbs last between 724 and 834 hours.

c) 15.9% bulbs last less than 724 hours.

Question 4

In 54 randomly selected hours of production, the mean and the standard deviation of the number of acceptable pieces produced by an automatic stamping machine are $\bar{x} = 1238$ and $s = 124$. At the 0.05 level of significance, does this enable us to reject the null hypothesis $H_0: \mu = 1200$ against the alternative hypothesis $H_1: \mu > 1200$?

Code:

```
# Given data

x_bar <- 1238
```

```
s <- 124

n <- 54

mu0 <- 1200

alpha <- 0.05


# Compute t statistic

t_value <- (x_bar - mu0) / (s / sqrt(n))

df <- n - 1

t_crit <- qt(1 - alpha, df)


# Print test statistics

cat("t-value =", t_value, "\n")

cat("Critical t-value =", t_crit, "\n")


# Decision rule

if (t_value > t_crit) {

  result <- "Reject H0: Mean > 1200"

} else {

  result <- "Fail to Reject H0"

}

cat("Decision:", result, "\n")


# Plotting the t-distribution

x <- seq(-4, 4, 0.01)

y <- dt(x, df)
```

```r
# Plot t distribution curve
plot(x, y, type = "l", lwd = 2, col = "black",
    main = "Right-tailed t-test Visualization",
    xlab = "t values", ylab = "Density")

# Shade rejection region
x_reject <- seq(t_crit, 4, 0.01)
y_reject <- dt(x_reject, df)
polygon(c(t_crit, x_reject, 4), c(0, y_reject, 0), col = "red", border = NA)

# Mark observed t value
abline(v = t_value, col = "blue", lwd = 2, lty = 2)

# Annotate areas
text(t_crit + 0.3, 0.05, "Critical Region", col = "red", cex = 0.8)
text(t_value - 0.2, 0.07, paste("t =", round(t_value, 2)), col = "blue", cex = 0.8)

# Add legend
legend("topright",
      legend = c("t-distribution", "Critical Region (α=0.05)", "Observed t-value"),
      col = c("black", "red", "blue"),
      lwd = c(2, NA, 2), lty = c(1, NA, 2),
      pch = c(NA, 15, NA), pt.cex = 2)
```

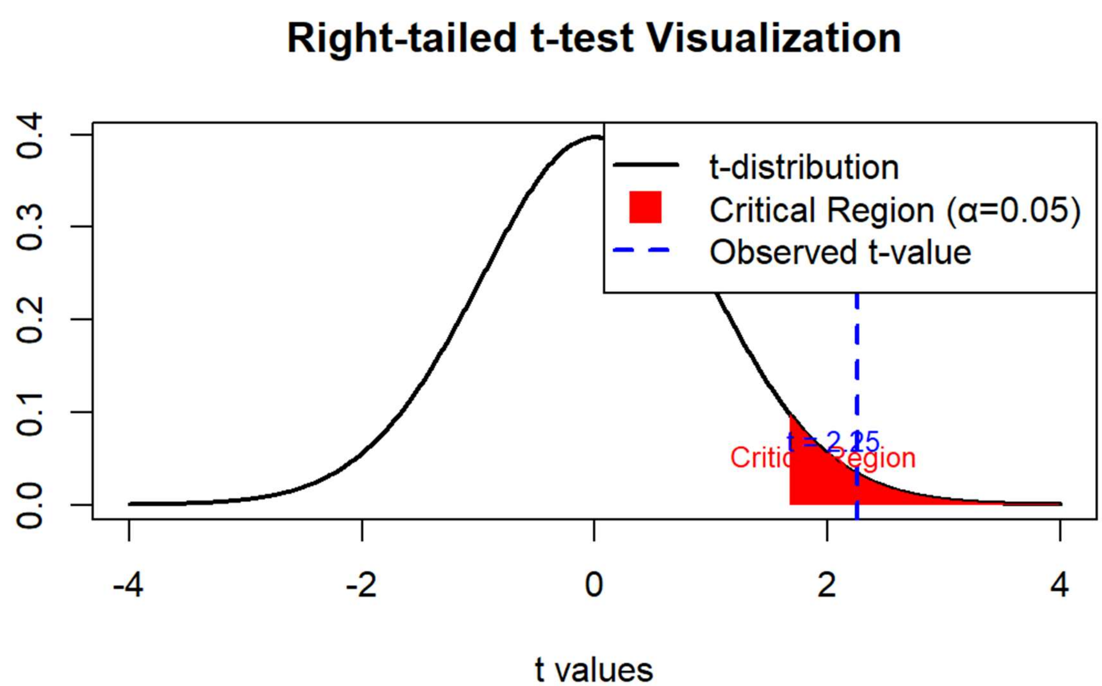
Output:

```
> cat("t-value =", t_value, "\n")
t-value = 2.25195
> cat("Critical t-value =", t_crit, "\n")
Critical t-value = 1.674116
```

```
> cat("t-value =", t_value, "\n")
t-value = 2.25195
> cat("Critical t-value =", t_crit, "\n")
Critical t-value = 1.674116
```

```
> cat("Decision:", result, "\n")
Decision: Reject H0: Mean > 1200
```

Plotting:



Results:

t = (1238 - 1200) / (124 / sqrt(54)) = 2.158

t_critical (α=0.05, df=53) = 1.674

Since t = 2.158 > 1.674,
Reject H_0 → there is significant evidence that the mean > 1200.