ASSIGNMENT -5

Name: M Gyanada Chowdary

Reg.No: 21BCE7727

1. An urban community would like to show that the incidence of breast cancer is higher in their area than in a nearby rural area. If it is found that 20 of 200 adult women in the urban community have breast cancer and 10 of 150 adult women in the rural community have breast cancer. Can we conclude at the 0.05 level of significance that breast cancer is more prevalent in the urban community?

Code:

```
# 21BCE7727
#M Gyanada Chowdary

urban_cases <- 20
urban_sample_size <- 200
rural_cases <- 10
rural_sample_size <- 150

# Calculate the sample proportions
urban_prop <- urban_cases / urban_sample_size
rural_prop <- rural_cases / rural_sample_size

# Calculate the standard error
se <- sqrt(urban_prop*(1-urban_prop)/urban_sample_size + rural_prop*(1-rural_prop)/rural_sample_size)

# Calculate the test statistic
z <- (urban_prop - rural_prop) / se

# Calculate the p-value
pval <- 2*pnorm(-abs(z))

# Print the results
cat("Test statistic:", round(z, 3), "\n")
cat("P-value:", format(pval, scientific = FALSE), "\n")
```

if (pval < 0.05) {

  cat("Conclusion: We reject the null hypothesis and conclude that breast cancer is more prevalent in the urban community.\n")

} else {

  cat("Conclusion: We fail to reject the null hypothesis and cannot conclude that breast cancer is more prevalent in the urban community.\n")

}

```r
# 21BCE7727
#M Gyanada Chowdary

urban_cases <- 20
urban_sample_size <- 200
rural_cases <- 10
rural_sample_size <- 150

# Calculate the sample proportions
urban_prop <- urban_cases / urban_sample_size
rural_prop <- rural_cases / rural_sample_size

# Calculate the standard error
se <- sqrt(urban_prop*(1-urban_prop)/urban_sample_size + rural_prop*(1-rural_prop)/rural_sample_size)

# Calculate the test statistic
z <- (urban_prop - rural_prop) / se

# Calculate the p-value
pval <- 2*pnorm(-abs(z))

# Print the results
cat("Test statistic:", round(z, 3), "\n")
cat("P-value:", format(pval, scientific = FALSE), "\n")
if (pval < 0.05) {
  cat("Conclusion: We reject the null hypothesis and conclude that breast cancer is more prevalent in the urban community.\n")
} else {
  cat("Conclusion: We fail to reject the null hypothesis and cannot conclude that breast cancer is more prevalent in the urban community.\n")
}
```

OUTPUT:

```
R 4.2.2 · ~/
> # 21BCE7727
> #M Gyanada Chowdary
>
> urban_cases <- 20
> urban_sample_size <- 200
> rural_cases <- 10
> rural_sample_size <- 150
>
> # Calculate the sample proportions
> urban_prop <- urban_cases / urban_sample_size
> rural_prop <- rural_cases / rural_sample_size
>
> # Calculate the standard error
> se <- sqrt(urban_prop*(1-urban_prop)/urban_sample_size + rural_prop*(1-rural_prop)/rural_sample_size)
>
> # Calculate the test statistic
> z <- (urban_prop - rural_prop) / se
>
> # Calculate the p-value
> pval <- 2*pnorm(-abs(z))
>
> # Print the results
> cat("Test statistic:", round(z, 3), "\n")
Test statistic: 1.133
> cat("P-value:", format(pval, scientific = FALSE), "\n")
P-value: 0.2570091
> if (pval < 0.05) {
+   cat("Conclusion: We reject the null hypothesis and conclude that breast cancer is more prevalent in the urban community.\n")
+ } else {
+   cat("Conclusion: We fail to reject the null hypothesis and cannot conclude that breast cancer is more prevalent in the urban community.\n")
+ }
Conclusion: We fail to reject the null hypothesis and cannot conclude that breast cancer is more prevalent in the urban community.
```

2. In a study to estimate the proportion of residents in a certain city and its suburbs who favor the construction of a nuclear power plant, it is found that 63 of 100 urban residents favor the construction while only 59 of 125 suburban residents are in favor. Is there a significant difference between the proportions of urban and suburban residents who favor construction of the nuclear plant? Use an $\alpha = 0.1$ level of significance.

Code:

```
#M Gyanada Chowdary
#21BCE7727

urban_cases <- 63
urban_sample_size <- 100
suburban_cases <- 59
suburban_sample_size <- 125
alpha <- 0.1

# Calculate the sample proportions
urban_prop <- urban_cases / urban_sample_size
suburban_prop <- suburban_cases / suburban_sample_size

# Calculate the standard error
se <- sqrt(urban_prop*(1-urban_prop)/urban_sample_size + suburban_prop*(1-
suburban_prop)/suburban_sample_size)

# Calculate the test statistic
z <- (urban_prop - suburban_prop) / se

# Calculate the p-value
pval <- 2*pnorm(-abs(z))

# Print the results
cat("Test statistic:", round(z, 3), "\n")
cat("P-value:", format(pval, scientific = FALSE), "\n")
if (pval < alpha) {
  cat("Conclusion: We reject the null hypothesis and conclude that there is a significant difference between the
proportions of urban and suburban residents who favor construction of the nuclear plant.\n")
```

} else {

cat("Conclusion: We fail to reject the null hypothesis and cannot conclude that there is a significant difference between the proportions of urban and suburban residents who favor construction of the nuclear plant.\n")

}

```r
#M Gyanada Chowdary
#21BCE7727

urban_cases <- 63
urban_sample_size <- 100
suburban_cases <- 59
suburban_sample_size <- 125
alpha <- 0.1

# Calculate the sample proportions
urban_prop <- urban_cases / urban_sample_size
suburban_prop <- suburban_cases / suburban_sample_size

# Calculate the standard error
se <- sqrt(urban_prop*(1-urban_prop)/urban_sample_size + suburban_prop*(1-suburban_prop)/suburban_sample_size)

# Calculate the test statistic
z <- (urban_prop - suburban_prop) / se

# Calculate the p-value
pval <- 2*pnorm(-abs(z))

# Print the results
cat("Test statistic:", round(z, 3), "\n")
cat("P-value:", format(pval, scientific = FALSE), "\n")
if (pval < alpha) {
  cat("Conclusion: We reject the null hypothesis and conclude that there is a significant difference between the proportions of urban and suburban residents who favor construction of the nuclear plant.\n")
} else {
  cat("Conclusion: We fail to reject the null hypothesis and cannot conclude that there is a significant difference between the proportions of urban and suburban residents who favor construction of the nuclear plant.\n")
}
```

OUTPUT:

```
Console  Terminal  Background Jobs
R 4.2.2 · ~/
> #M Gyanada Chowdary
> #21BCE7727
>
> urban_cases <- 63
> urban_sample_size <- 100
> suburban_cases <- 59
> suburban_sample_size <- 125
> alpha <- 0.1
>
> # Calculate the sample proportions
> urban_prop <- urban_cases / urban_sample_size
> suburban_prop <- suburban_cases / suburban_sample_size
>
> # Calculate the standard error
> se <- sqrt(urban_prop*(1-urban_prop)/urban_sample_size + suburban_prop*(1-suburban_prop)/suburban_sample_size)
>
> # Calculate the test statistic
> z <- (urban_prop - suburban_prop) / se
>
> # Calculate the p-value
> pval <- 2*pnorm(-abs(z))
>
> # Print the results
> cat("Test statistic:", round(z, 3), "\n")
Test statistic: 2.403
> cat("P-value:", format(pval, scientific = FALSE), "\n")
P-value: 0.0162791
> if (pval < alpha) {
+   cat("Conclusion: We reject the null hypothesis and conclude that there is a significant difference between the proportions of urban and suburban residents who favor construction of the nuclear plant.\n")
+ } else {
+   cat("Conclusion: We fail to reject the null hypothesis and cannot conclude that there is a significant difference between the proportions of urban and suburban residents who favor construction of the nuclear plant.\n")
+ }
Conclusion: We reject the null hypothesis and conclude that there is a significant difference between the proportions of urban and suburban residents who favor construction of the nuclear plant.
>
```

3. An outbreak of salmonella-related illness was attributed to ice produced at a certain factory. Scientists measured the level of Salmonella in 9 randomly sampled batches ice cream. The levels(in MPN/g) were: $0.593, 0.142, 0.329, 0.691, 0.231, 0.793, 0.519, 0.392, 0.418$ Is there evidence that the mean level pf Salmonella in ice cream greater than 0.3 MPN/g?

Code:

#21BCE7727

#M Gyanada Chowdary

salmonella_levels <- c(0.593, 0.142, 0.329, 0.691, 0.231, 0.793, 0.519, 0.392, 0.418)

mu <- 0.3

alpha <- 0.05

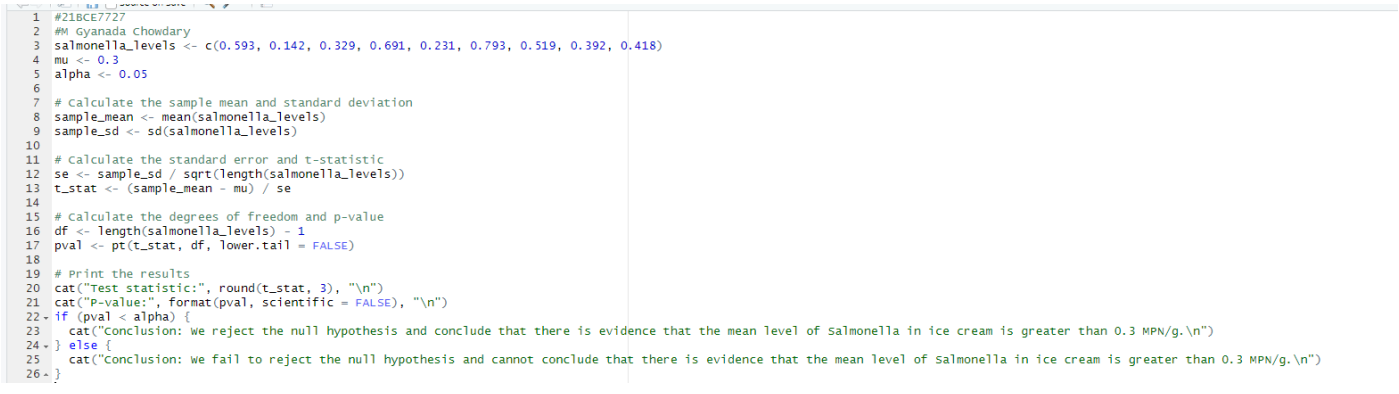
Calculate the sample mean and standard deviation

```
sample_mean <- mean(salmonella_levels)
sample_sd <- sd(salmonella_levels)

# Calculate the standard error and t-statistic
se <- sample_sd / sqrt(length(salmonella_levels))
t_stat <- (sample_mean - mu) / se

# Calculate the degrees of freedom and p-value
df <- length(salmonella_levels) - 1
pval <- pt(t_stat, df, lower.tail = FALSE)

# Print the results
cat("Test statistic:", round(t_stat, 3), "\n")
cat("P-value:", format(pval, scientific = FALSE), "\n")
if (pval < alpha) {
  cat("Conclusion: We reject the null hypothesis and conclude that there is evidence that the mean level of Salmonella in ice cream is greater than 0.3 MPN/g.\n")
} else {
  cat("Conclusion: We fail to reject the null hypothesis and cannot conclude that there is evidence that the mean level of Salmonella in ice cream is greater than 0.3 MPN/g.\n")
}
```



OUTPUT:

```
> #21BCE7727
> #M Gyanada Chowdary
> salmonella_levels <- c(0.593, 0.142, 0.329, 0.691, 0.231, 0.793, 0.519, 0.392, 0.418)
> mu <- 0.3
> alpha <- 0.05
>
> # Calculate the sample mean and standard deviation
> sample_mean <- mean(salmonella_levels)
> sample_sd <- sd(salmonella_levels)
>
> # Calculate the standard error and t-statistic
> se <- sample_sd / sqrt(length(salmonella_levels))
> t_stat <- (sample_mean - mu) / se
>
> # Calculate the degrees of freedom and p-value
> df <- length(salmonella_levels) - 1
> pval <- pt(t_stat, df, lower.tail = FALSE)
>
> # Print the results
> cat("Test statistic:", round(t_stat, 3), "\n")
Test statistic: 2.205
> cat("P-value:", format(pval, scientific = FALSE), "\n")
P-value: 0.02926516
> if (pval < alpha) {
+   cat("Conclusion: We reject the null hypothesis and conclude that there is evidence that the mean level of Salmonella in ice cream is greater than 0.3 MPN/g.\n")
+ } else {
+   cat("Conclusion: We fail to reject the null hypothesis and cannot conclude that there is evidence that the mean level of Salmonella in ice cream is greater than 0.3 MPN/g.\n")
+ }
Conclusion: We reject the null hypothesis and conclude that there is evidence that the mean level of Salmonella in ice cream is greater than 0.3 MPN/g.
>
```

4. Suppose that 10 volunteers have taken an intelligence test; here are the results obtained. The average score of the entire population is 75 in the same test. Is there any significant difference (with a significance level of 95%) between the sample and population means, assuming that the variance of the population is not known.
   Scores: 65, 78, 88, 55, 48, 95, 66, 57, 79, 81.

Code:

```
#21BCE7727

#M Gyanada Chowdary

scores <- c(65, 78, 88, 55, 48, 95, 66, 57, 79, 81)

mu <- 75

alpha <- 0.05


# Calculate the t-statistic and p-value

t_test <- t.test(scores, mu = mu, alternative = "two.sided", conf.level = 1-alpha)

t_stat <- t_test$statistic

pval <- t_test$p.value


# Print the results

cat("Test statistic:", round(t_stat, 3), "\n")

cat("P-value:", format(pval, scientific = FALSE), "\n")

if (pval < alpha) {

  cat("Conclusion: We reject the null hypothesis and conclude that there is a significant difference between the sample and population means.\n")

} else {

  cat("Conclusion: We fail to reject the null hypothesis and cannot conclude that there is a significant difference between the sample and population means.\n")
```

}

```r
#21BCE7727
#M Gyanada Chowdary
scores <- c(65, 78, 88, 55, 48, 95, 66, 57, 79, 81)
mu <- 75
alpha <- 0.05

# Calculate the t-statistic and p-value
t_test <- t.test(scores, mu = mu, alternative = "two.sided", conf.level = 1-alpha)
t_stat <- t_test$statistic
pval <- t_test$p.value

# Print the results
cat("Test statistic:", round(t_stat, 3), "\n")
cat("P-value:", format(pval, scientific = FALSE), "\n")
if (pval < alpha) {
  cat("Conclusion: We reject the null hypothesis and conclude that there is a significant difference between the sample and population means.\n")
} else {
  cat("Conclusion: We fail to reject the null hypothesis and cannot conclude that there is a significant difference between the sample and population means.\n")
}

```

OUTPUT:

```
> #21BCE7727
> #M Gyanada Chowdary
> scores <- c(65, 78, 88, 55, 48, 95, 66, 57, 79, 81)
> mu <- 75
> alpha <- 0.05
>
> # Calculate the t-statistic and p-value
> t_test <- t.test(scores, mu = mu, alternative = "two.sided", conf.level = 1-alpha)
> t_stat <- t_test$statistic
> pval <- t_test$p.value
>
> # Print the results
> cat("Test statistic:", round(t_stat, 3), "\n")
Test statistic: -0.783
> cat("P-value:", format(pval, scientific = FALSE), "\n")
P-value: 0.4537205
> if (pval < alpha) {
+   cat("Conclusion: We reject the null hypothesis and conclude that there is a significant difference between the sample and population means.\n")
+ } else {
+   cat("Conclusion: We fail to reject the null hypothesis and cannot conclude that there is a significant difference between the sample and population means.\n")
+ }
Conclusion: We fail to reject the null hypothesis and cannot conclude that there is a significant difference between the sample and population means.
>
```

5. Comparing two independent sample means, taken from two populations with unknown variance.The following data shows the heights of individuals of two different countries with unknown population variances. Is there any significant difference b/n the average heights of two groups.
   A: 175, 168, 168, 190, 156, 181, 182, 175, 174, 179
   B: 185, 169, 173, 173, 188, 186, 175, 174, 179, 180.

Code:

```r
# Define the data for Group A and Group B

group_a <- c(175, 168, 168, 190, 156, 181, 182, 175, 174, 179)

group_b <- c(185, 169, 173, 173, 188, 186, 175, 174, 179, 180)


# Calculate the t-statistic and p-value for two-sample t-test

t_test <- t.test(group_a, group_b, var.equal = TRUE)


# Print the results

cat("Test statistic:", round(t_test$statistic, 3), "\n")

cat("P-value:", format.pval(t_test$p.value, digits = 3), "\n")

if (t_test$p.value < 0.05) {
```

cat("Conclusion: We reject the null hypothesis and conclude that there is a significant difference between the average heights of the two groups.\n")

} else {

cat("Conclusion: We fail to reject the null hypothesis and cannot conclude that there is a significant difference between the average heights of the two groups.\n")

}

```r
#21BCE7727
#M Gyanada Chowdary
group_a <- c(175, 168, 168, 190, 156, 181, 182, 175, 174, 179)
group_b <- c(185, 169, 173, 173, 188, 186, 175, 174, 179, 180)

# Calculate the t-statistic and p-value for two-sample t-test
t_test <- t.test(group_a, group_b, var.equal = TRUE)

# Print the results
cat("Test statistic:", round(t_test$statistic, 3), "\n")
cat("P-value:", format.pval(t_test$p.value, digits = 3), "\n")
if (t_test$p.value < 0.05) {
  cat("Conclusion: We reject the null hypothesis and conclude that there is a significant difference between the average heights of the two groups.\n")
} else {
  cat("Conclusion: We fail to reject the null hypothesis and cannot conclude that there is a significant difference between the average heights of the two groups.\n")
}

```

OUTPUT:

```
R 4.2.2 · ~/
> #21BCE7727
> #M Gyanada Chowdary
> group_a <- c(175, 168, 168, 190, 156, 181, 182, 175, 174, 179)
> group_b <- c(185, 169, 173, 173, 188, 186, 175, 174, 179, 180)
>
> # Calculate the t-statistic and p-value for two-sample t-test
> t_test <- t.test(group_a, group_b, var.equal = TRUE)
>
> # Print the results
> cat("Test statistic:", round(t_test$statistic, 3), "\n")
Test statistic: -0.947
> cat("P-value:", format.pval(t_test$p.value, digits = 3), "\n")
P-value: 0.356
> if (t_test$p.value < 0.05) {
+   cat("Conclusion: We reject the null hypothesis and conclude that there is a significant difference between the average heights of the two groups.\n")
+ } else {
+   cat("Conclusion: We fail to reject the null hypothesis and cannot conclude that there is a significant difference between the average heights of the two groups.\n")
+ }
Conclusion: We fail to reject the null hypothesis and cannot conclude that there is a significant difference between the average heights of the two groups.
>
```