STAT 400 - Discussion 6

Colin Gibbons-Fly

One sample - Function

```
set.seed(100)
df <- rnorm(50,3,2)

t.test(df, mu=3)
```

```
    One Sample t-test

data:  df
t = 0.70372, df = 49, p-value = 0.4849
alternative hypothesis: true mean is not equal to 3
95 percent confidence interval:
 2.697497 3.628539
sample estimates:
mean of x
 3.163018
```

One Sample - By hand

```
# Sample statistics
sample_mean <- mean(df)
sample_sd <- sd(df)
n <- length(df)
mu <- 3  # Null hypothesis mean
```

```r
# Calculate t-statistic
t_stat <- (sample_mean - mu) / (sample_sd / sqrt(n))

# Degrees of freedom
df_degrees <- n - 1

# One-sided p-value
p_value <- pt(t_stat, df = df_degrees, lower.tail = FALSE)

# Print results
cat("One-Sided t-Test (greater):\n")
```

```
One-Sided t-Test (greater):
```

```r
cat("Sample Mean:", sample_mean, "\n")
```

```
Sample Mean: 3.163018
```

```r
cat("Sample SD:", sample_sd, "\n")
```

```
Sample SD: 1.638023
```

```r
cat("t-Statistic:", t_stat, "\n")
```

```
t-Statistic: 0.7037208
```

```r
cat("p-Value:", p_value, "\n")
```

```
p-Value: 0.2424687
```

Two Sample - Function

```r
df2 <- rnorm(50,8,2)
df3 <- rnorm(50,10,1)

t.test(df2, df3, var.equal=TRUE)
```

```
    Two Sample t-test

data:  df2 and df3
t = -5.7723, df = 98, p-value = 9.182e-08
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -2.756892 -1.346271
sample estimates:
mean of x mean of y
 7.848632  9.900213
```

Two Sample - By hand

```r
# Sample statistics
mean_df2 <- mean(df2)
mean_df3 <- mean(df3)
sd_df2 <- sd(df2)
sd_df3 <- sd(df3)
n_df2 <- length(df2)
n_df3 <- length(df3)

# Pooled variance
sp_squared <- (((n_df2 - 1) * sd_df2^2) + ((n_df3 - 1) * sd_df3^2)) / (n_df2 + n_df3 - 2)

# t-statistic
t_stat <- (mean_df2 - mean_df3) / sqrt(sp_squared * (1 / n_df2 + 1 / n_df3))

# Degrees of freedom
df_degrees <- n_df2 + n_df3 - 2

# Two-sided p-value
p_value <- 2 * pt(-abs(t_stat), df = df_degrees)

# 95% confidence interval
alpha <- 0.05
t_critical <- qt(1 - alpha / 2, df = df_degrees)
ci_lower <- (mean_df2 - mean_df3) - t_critical * sqrt(sp_squared * (1 / n_df2 + 1 / n_df3))
ci_upper <- (mean_df2 - mean_df3) + t_critical * sqrt(sp_squared * (1 / n_df2 + 1 / n_df3))
```

```r
# Print results
cat("Two-Sample t-Test by Hand:\n")
```

```
Two-Sample t-Test by Hand:
```

```r
cat("Mean (df2):", mean_df2, "\n")
```

```
Mean (df2): 7.848632
```

```r
cat("Mean (df3):", mean_df3, "\n")
```

```
Mean (df3): 9.900213
```

```r
cat("Pooled Variance:", sp_squared, "\n")
```

```
Pooled Variance: 3.158004
```

```r
cat("t-Statistic:", t_stat, "\n")
```

```
t-Statistic: -5.772346
```

```r
cat("p-Value:", p_value, "\n")
```

```
p-Value: 9.182013e-08
```

```r
cat("95% Confidence Interval:", ci_lower, "to", ci_upper, "\n")
```

```
95% Confidence Interval: -2.756892 to -1.346271
```