LABORATORY EXAM

Paculba, Laika Jean B.

2023-06-29

1 The Fractional Distillation Data

The purity of oxygen produced by a fractional distillation process is thought to be related to the percentage of hydrocarbons in the main condensor of the processing unit. Twenty samples are shown below.

```
## Warning: package 'pander' was built under R version 4.2.3
```

observations	Purity	Hydrocarbon
1	86.91	1.02
2	89.85	1.11
3	90.28	1.43
4	86.34	1.11
5	92.58	1.01
6	87.33	0.95
7	86.29	1.11
8	91.86	0.87
9	95.61	1.43
10	89.86	1.02
11	96.73	1.46
12	99.42	1.55
13	98.66	1.55
14	96.07	1.55
15	93.65	1.40
16	87.31	1.15
17	95.00	1.01
18	96.85	0.99
19	85.20	0.95
20	90.56	0.98

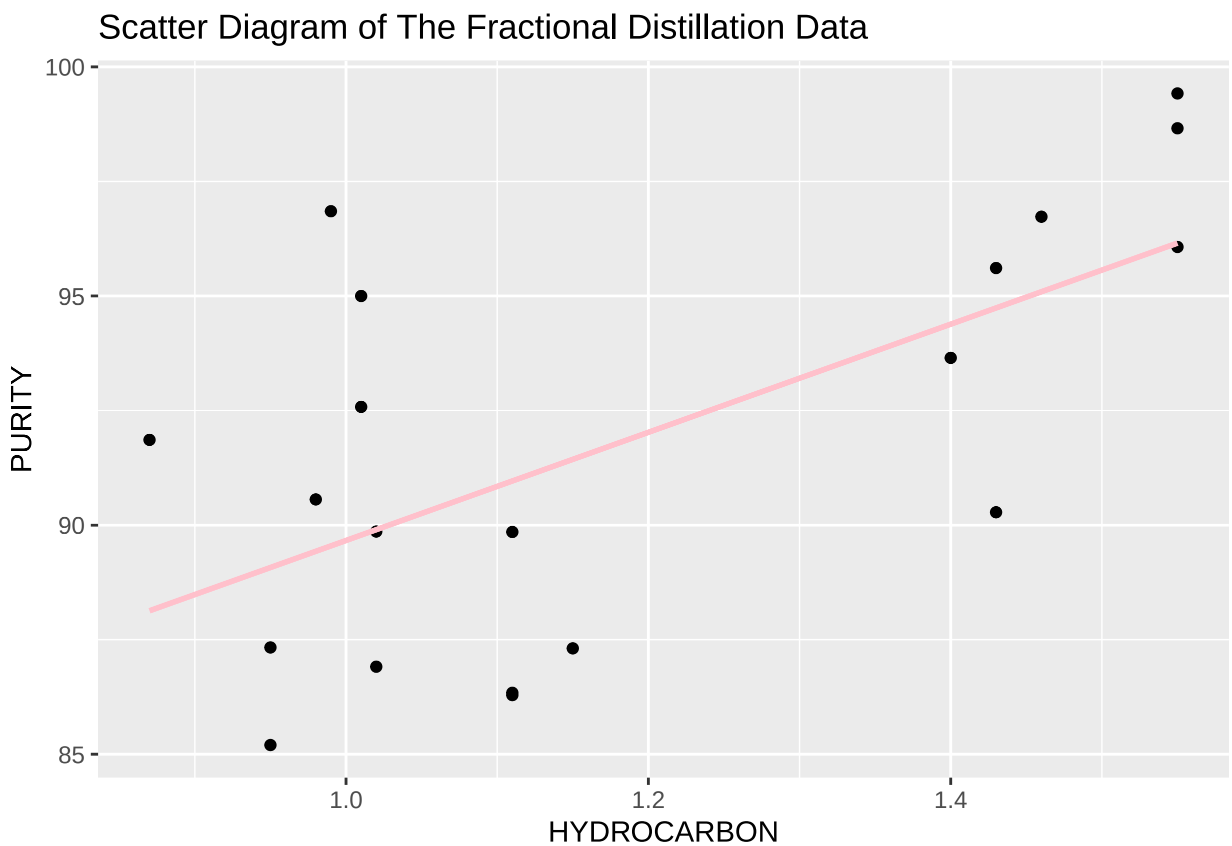
###a. Create a scatter diagram for the data.

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.2.3
```

```
data1 <- data.frame(
  observations = 1:20,
  Purity = c(86.91, 89.85, 90.28, 86.34, 92.58, 87.33, 86.29, 91.86, 95.61, 89.86, 96.73, 99.42, 98.66, 
  Hydrocarbon = c(1.02, 1.11, 1.43, 1.11, 1.01, 0.95, 1.11, 0.87, 1.43, 1.02, 1.46, 1.55, 1.55, 1.55, 1

Purity = c(86.91, 89.85, 90.28, 86.34, 92.58, 87.33, 86.29, 91.86, 95.61, 89.86, 96.73, 99.42, 98.66, 9
Hydrocarbon = c(1.02, 1.11, 1.43, 1.11, 1.01, 0.95, 1.11, 0.87, 1.43, 1.02, 1.46, 1.55, 1.55, 1.55, 1.4

ggplot(data = NULL, aes(x = Hydrocarbon, y= Purity)) +
  geom_point() + geom_smooth(method = "lm", se = FALSE, color = "pink") +
  labs(title = "Scatter Diagram of The Fractional Distillation Data",  x = "HYDROCARBON", y = "PURITY")
```

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



###b. The least-squares fit is .

```
model <- lm(data = data1,
formula = Purity ~ Hydrocarbon)
names(model)
```

```
##  [1] "coefficients"  "residuals"     "effects"       "rank"
##  [5] "fitted.values" "assign"        "qr"            "df.residual"
##  [9] "xlevels"       "call"          "terms"         "model"
```

```r
model$coefficients
```

```
## (Intercept) Hydrocarbon
##    77.86328    11.80103
```

```r
model_summary <- summary(model)
model_summary$sigma
```

```
## [1] 3.59656
```

```r
# Therefore, the least square fit is y_hat = 77.46239 + (12.02816)x
# view the fitted values
fitted.values <- fitted(model)
fitted.values
```

```
##        1        2        3        4        5        6        7        8
## 89.90033 90.96243 94.73875 90.96243 89.78232 89.07426 90.96243 88.13018
##        9       10       11       12       13       14       15       16
## 94.73875 89.90033 95.09279 96.15488 96.15488 96.15488 94.38472 91.43447
##       17       18       19       20
## 89.78232 89.54630 89.07426 89.42829
```

###c. The estimate of $\sigma^2$is.

```r
sigma_hat_squared <- ((model_summary$sigma)^2)
print(paste("Therefore the estimate of sigma squared is: ", sigma_hat_squared))
```

```
## [1] "Therefore the estimate of sigma squared is:  12.9352421041182"
```

###d. Test for significance of regression in the regression model.

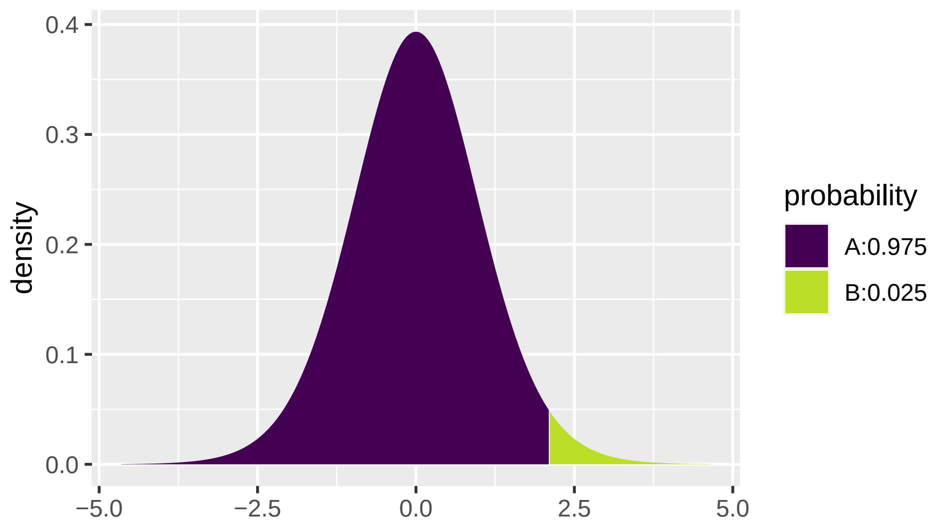
```r
library(mosaic)
```

```
## Warning: package 'mosaic' was built under R version 4.2.3
```

```r
model_summary$coefficients["Hydrocarbon",]
```

```
##     Estimate   Std. Error      t value     Pr(>|t|)
## 11.801028193  3.485118700  3.386119444  0.003291122
```

```r
mosaic::xqt(0.975, 18)
```




```
## [1] 2.100922
```

###e. Use an analysis-of-variance approach to test significance of regression. Null Hypothesis: H_0: $\beta_1 = 0$ Alternative Hypothesis: H_1: $\beta_1 ! = 0$

```
t_value <- model_summary$coefficients["Hydrocarbon","t value"]
print(t_value)
```

```
## [1] 3.386119
```

```
# Find the critical value
df <- 20
significance_level <- 0.05
critical_value <- qt(1-significance_level/2, df-2)
print(paste("The critical value is:", critical_value))
```

```
## [1] "The critical value is: 2.10092204024104"
```

When $\alpha = 0.05$, the critical value of t is $t_{0.025,18} = 2.101$ and the T Value is 3.39

Since |t value| > critical value

Thus, we would reject $H_0 : beta_1 = 0$.

Hence, there is a linear relationship between shear strength and the age of the propellant.

```
model_summary
```

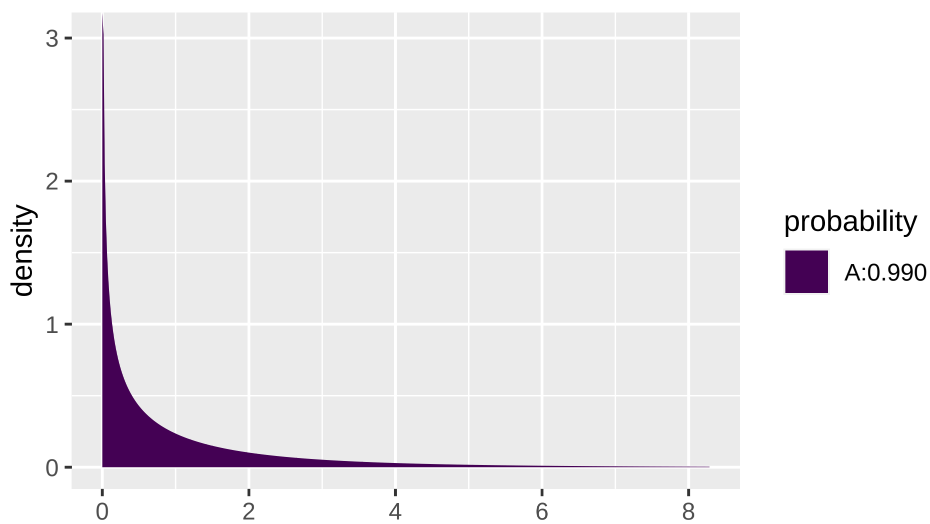
```
##
## Call:
## lm(formula = Purity ~ Hydrocarbon, data = data1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
```

```
## -4.6724 -3.2113 -0.0626  2.5783  7.3037
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   77.863      4.199  18.544 3.54e-13 ***
## Hydrocarbon   11.801      3.485   3.386  0.00329 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.597 on 18 degrees of freedom
## Multiple R-squared:  0.3891, Adjusted R-squared:  0.3552
## F-statistic: 11.47 on 1 and 18 DF,  p-value: 0.003291
```

```r
model_summary$fstatistic
```

```
##   value   numdf   dendf
## 11.4658  1.0000 18.0000
```

```r
mosaic::xqf(0.99,1,18)
```




```
## [1] 8.28542
```

```r
anova_result <- anova(model)
print(anova_result)
```

```
## Analysis of Variance Table
##
## Response: Purity
##             Df Sum Sq Mean Sq F value   Pr(>F)
## Hydrocarbon  1 148.31 148.313  11.466 0.003291 **
## Residuals   18 232.83  12.935
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Extracting the F value

```
f_value <- anova_result$'F value'[1]
print(f_value)
```

```
## [1] 11.4658
```

Finding the critical values

```
# Find the critical value
df1 <- 1
df2 <- 20
significance_level <- 0.05
critical_fvalue <- qf(1-significance_level, df1, df2-2)
print(paste("The Critical F value is:", critical_fvalue))
```

```
## [1] "The Critical F value is: 4.41387341917057"
```

```
result <- t.test(Hydrocarbon, Purity)
p_value <- result$p.value
print(p_value)
```

```
## [1] 1.29778e-26
```

The P value for this test is 0.003.

Using F test in testing the analysis of variance and choosing the significance level $\alpha = 0.05$, the critical value of f is computed as $f_{0.05,1,18} = 4.41$

The F value is 11.4658 and the critical F value is 4.41. Since the F value is greater than the critical value, Hence, we reject the null hypothesis.

Therefore, there is a linear relationship between shear strength and the age of the propellant.

###f. Find a 95%CI on the slope.

```
# Find a 95%CI on the slope
# Input data
y <- c(86.91, 89.85, 90.28, 86.34, 92.58, 87.33, 86.29, 91.86, 95.61, 89.86,
96.73, 99.42, 98.66, 96.07, 93.65, 87.31, 95.00, 96.85, 85.20, 90.56)
x <- c(1.02, 1.11, 1.43, 1.11, 1.01, 0.95, 1.11, 0.87, 1.43, 1.02,1.46, 1.55,
1.55, 1.55, 1.40, 1.15, 1.01, 0.99, 0.95, 0.98)
# Perform linear regression
model <- lm(y ~ x)
# Calculate 95% confidence interval
conf_interval <- confint(model, level = 0.95)
# Display the confidence interval
conf_interval
```

```
##                 2.5 %   97.5 %
## (Intercept) 69.041747 86.68482
## x            4.479066 19.12299
```

###g. Find a 95%CI on the mean purity when the hydrocarbon percentage is 1.00.

```r
# Find a 95%CI on the mean y when x percentage is 1.00.
# Input data
y <- c(86.91, 89.85, 90.28, 86.34, 92.58, 87.33, 86.29, 91.86, 95.61, 89.86,
96.73, 99.42, 98.66, 96.07, 93.65, 87.31, 95.00, 96.85, 85.20, 90.56)
x <- c(1.02, 1.11, 1.43, 1.11, 1.01, 0.95, 1.11, 0.87, 1.43, 1.02,1.46, 1.55,
1.55, 1.55, 1.40, 1.15, 1.01, 0.99, 0.95, 0.98)
# Perform linear regression
model <- lm(y ~ x)
# Set the x value for prediction

3
```

```
## [1] 3
```

```r
x_pred <- 1.00
# Predict the mean y for x_pred
y_pred <- predict(model, newdata = data.frame(x = x_pred), interval = "confidence", level = 0.95)
# Display the confidence interval
y_ci <- y_pred[, c("lwr", "upr")]
y_ci
```

```
##      lwr      upr
## 87.51017 91.81845
```

2. The Steam Consumption Data

The number of pounds of steam used per month at a plant is thought to be related to the average monthly ambient temperature. The past year's usages and temperatures follow.

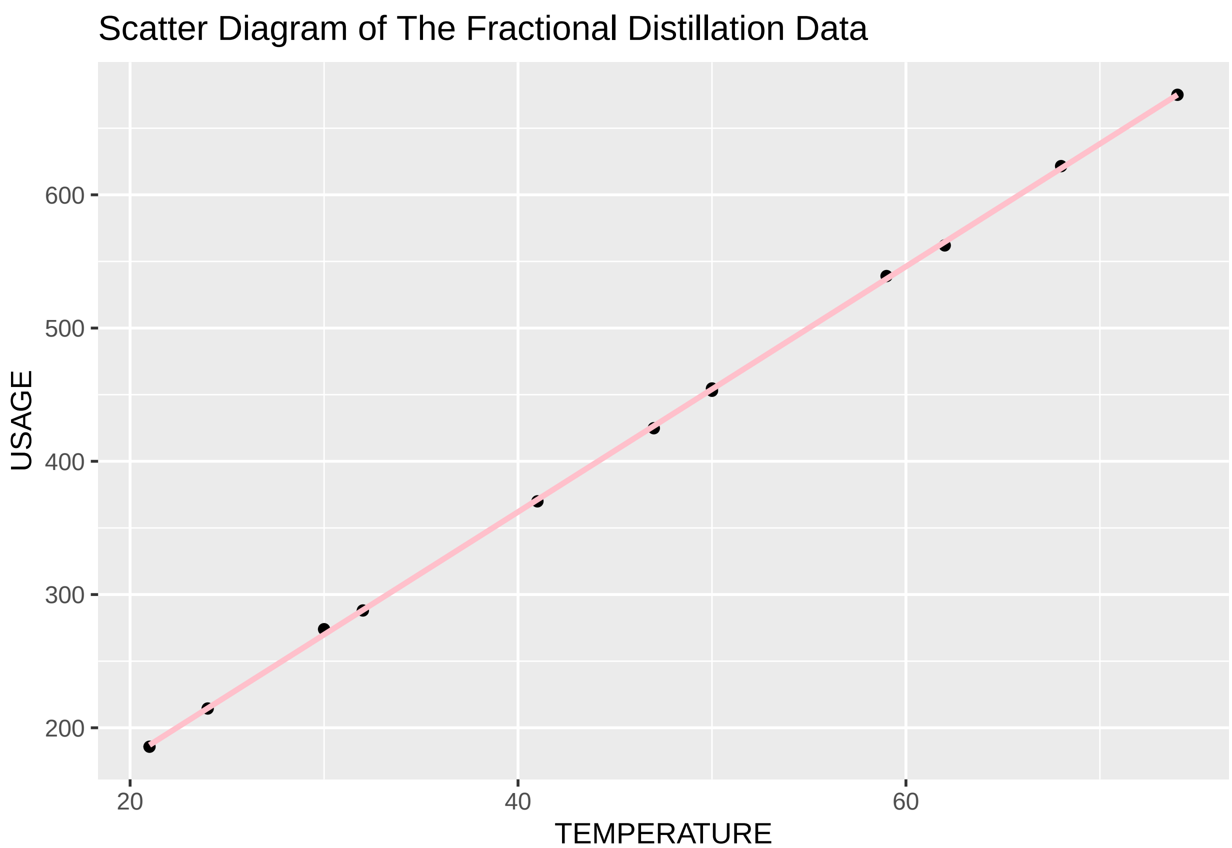
observations	Usage	Temperature
1	185.79	21.00
2	214.47	24.00
3	288.03	32.00
4	424.84	47.00
5	454.68	50.00
6	539.03	59.00
7	621.55	68.00
8	675.06	74.00
9	562.03	62.00
10	452.93	50.00
11	369.95	41.00
12	273.98	30.00

###a. Create a scatter diagram for the data.

```r
library(ggplot2)

data2 <- data.frame(
  observations = 1:12,
```

```r
  Usage = c(185.79, 214.47, 288.03, 424.84, 454.68, 539.03, 621.55, 675.06, 562.03, 452.93, 369.95, 273
  Temperature = c(21, 24, 32, 47, 50, 59, 68, 74, 62, 50, 41, 30))

Usage = c(185.79, 214.47, 288.03, 424.84, 454.68, 539.03, 621.55, 675.06, 562.03, 452.93, 369.95, 273.9
Temperature = c(21, 24, 32, 47, 50, 59, 68, 74, 62, 50, 41, 30)

ggplot(data = NULL, aes(x = Temperature, y= Usage)) +
  geom_point() + geom_smooth(method = "lm", se = FALSE, color = "pink") +
  labs(title = "Scatter Diagram of The Fractional Distillation Data",  x = "TEMPERATURE", y = "USAGE")
```

```
## ‘geom_smooth()‘ using formula = ’y ~ x’
```



###b. The least-squares fit is .

```r
model2 <- lm(data = data2,
formula = Usage ~ Temperature)
names(model2)
```

```
##  [1] "coefficients"  "residuals"     "effects"       "rank"
##  [5] "fitted.values" "assign"        "qr"            "df.residual"
##  [9] "xlevels"       "call"          "terms"         "model"
```

```r
model2$coefficients
```

```
## (Intercept) Temperature
##   -6.332087    9.208468
```

```r
modelsummary <- summary(model2)
modelsummary$sigma
```

```
## [1] 1.945628
```

```r
modelsummary
```

```
##
## Call:
## lm(formula = Usage ~ Temperature, data = data2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.5629 -1.2581 -0.2550  0.8681  4.0581
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -6.33209    1.67005  -3.792  0.00353 **
## Temperature  9.20847    0.03382 272.255  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.946 on 10 degrees of freedom
## Multiple R-squared:  0.9999, Adjusted R-squared:  0.9999
## F-statistic: 7.412e+04 on 1 and 10 DF,  p-value: < 2.2e-16
```

```r
# Therefore, the least square fit is y_hat = 77.46239 + (12.02816)x
# view the fitted values
fitted.values2 <- fitted(model2)
fitted.values2
```

```
##        1        2        3        4        5        6        7        8
## 187.0457 214.6711 288.3389 426.4659 454.0913 536.9675 619.8437 675.0945
##        9       10       11       12
## 564.5929 454.0913 371.2151 269.9219
```

###c. The estimate of σ^2

```r
sigma_hat_squared2 <- ((modelsummary$sigma)^2)
print(paste("Therefore the estimate of sigma squared is: ", sigma_hat_squared2))
```

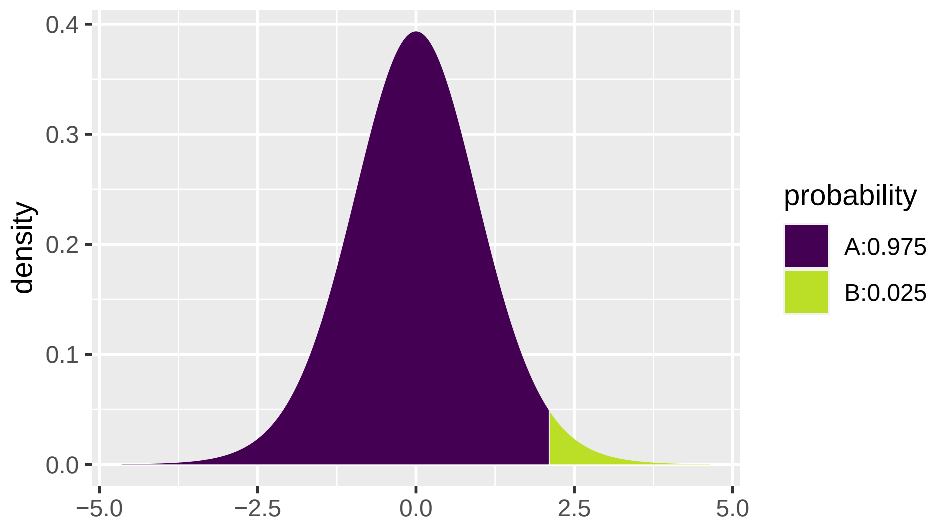
```
## [1] "Therefore the estimate of sigma squared is:  3.78546984889691"
```

###d. Test for significance of regression in the regression model.

```r
library(mosaic)
modelsummary$coefficients["Temperature",]
```

```
##     Estimate   Std. Error      t value     Pr(>|t|)
## 9.208468e+00 3.382295e-02 2.722550e+02 1.099192e-20
```

```
mosaic::xqt(0.975, 18)
```




```
## [1] 2.100922
```

###e. Use an analysis-of-variance approach to test significance of regression.

```
t_value2 <- modelsummary$coefficients["Temperature","t value"]
print(t_value2)
```

```
## [1] 272.255
```

```
# Find the critical value
df <- 20
significance_level <- 0.05
critical_value <- qt(1-significance_level/2, df-2)
print(paste("The critical value is:", critical_value))
```

```
## [1] "The critical value is: 2.10092204024104"
```

When $\alpha = 0.05$, the critical value of t is $t_{0.025,18} = 2.101$ and the T Value is 272.255

Since $|\text{t value}| > \text{critical value}$

Thus, we would reject $H_0 : beta_1 = 0$.

Hence, there is a linear relationship between shear strength and the age of the propellant.

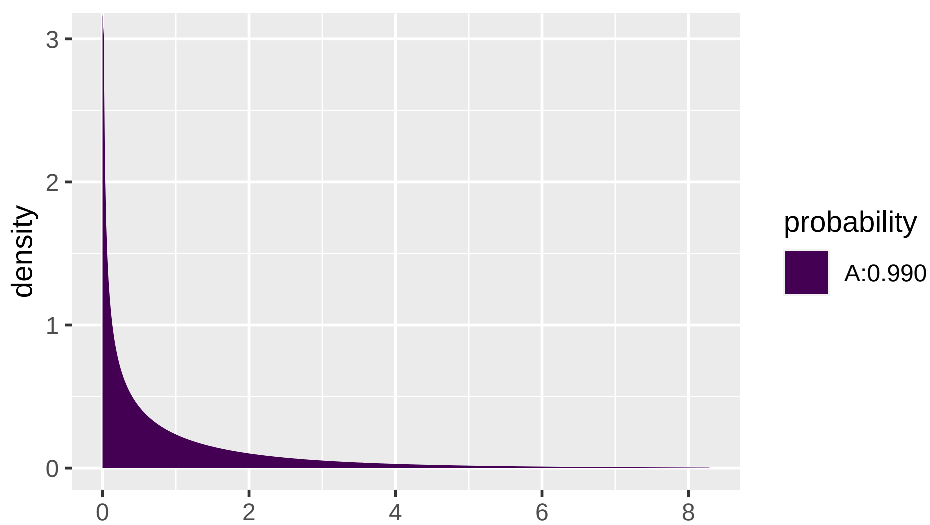
```
modelsummary
```

```
##
## Call:
## lm(formula = Usage ~ Temperature, data = data2)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.5629 -1.2581 -0.2550  0.8681  4.0581
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -6.33209    1.67005  -3.792  0.00353 **
## Temperature  9.20847    0.03382 272.255  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.946 on 10 degrees of freedom
## Multiple R-squared:  0.9999, Adjusted R-squared:  0.9999
## F-statistic: 7.412e+04 on 1 and 10 DF,  p-value: < 2.2e-16
```

```
modelsummary$fstatistic
```

```
##    value    numdf    dendf
## 74122.78     1.00    10.00
```

```
mosaic::xqf(0.99,1,18)
```




```
## [1] 8.28542
```

```
anova_result2 <- anova(model2)
print(anova_result2)
```

```
## Analysis of Variance Table
```

```
##
## Response: Usage
##             Df Sum Sq Mean Sq F value    Pr(>F)
## Temperature  1 280590  280590   74123 < 2.2e-16 ***
## Residuals   10     38       4
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Extracting the F value

```
f_value2 <- anova_result2$'F value'[1]
print(f_value2)
```

```
## [1] 74122.78
```

Finding the critical values

```
# Find the critical value
df1 <- 1
df2 <- 20
significance_level2 <- 0.05
critical_fvalue2 <- qf(1-significance_level, df1, df2-2)
print(paste("The Critical F value is:", critical_fvalue2))
```

```
## [1] "The Critical F value is: 4.41387341917057"
```

```
result <- t.test(Temperature, Usage)
p_value2 <- result$p.value
print(p_value2)
```

```
## [1] 5.04028e-06
```

The P value for this test is 0.003.

Using F test in testing the analysis of variance and choosing the significance level $\alpha = 0.05$, the critical value of f is computed as $f_{0.05,1,18} = 4.41$

The F value is 74122.78 and the critical F value is 4.41. Since the F value is greater than the critical value, Hence, we reject the null hypothesis.

Therefore, there is a linear relationship between shear strength and the age of the propellant.

###f. Find a 99%CI on the slope.

```
# Find a 95%CI on the slope
# Input data
y2 <- c(185.79, 214.47, 288.03, 424.84, 454.68, 539.03, 621.55, 675.06, 562.03, 452.93, 369.95, 273.98)
x2 <- c(21, 24, 32, 47, 50, 59, 68, 74, 62, 50, 41, 30)
# Perform linear regression
model2 <- lm(y2 ~ x2)
# Calculate 95% confidence interval
conf_interval2 <- confint(model2, level = 0.99)
# Display the confidence interval
conf_interval2
```

```
##                  0.5 %    99.5 %
## (Intercept) -11.624917 -1.039256
## x2            9.101274  9.315662
```

###g. Construct a 99% prediction interval on steam usage in a month with average ambient temperature of 58 ∘

```
# Find a 95%CI on the mean y when x percentage is 1.00.
# Input data
y2 <- c(185.79, 214.47, 288.03, 424.84, 454.68, 539.03, 621.55, 675.06, 562.03, 452.93, 369.95, 273.98)
x2 <- c(21, 24, 32, 47, 50, 59, 68, 74, 62, 50, 41, 30)
# Perform linear regression
model2 <- lm(y2 ~ x2)
# Set the x value for prediction

3
```

```
## [1] 3
```

```
x_pred2 <- 58
# Predict the mean y for x_pred2
y_pred2 <- predict(model2, newdata = data.frame(x = x_pred2), interval = "confidence", level = 0.99)
# Display the confidence interval
y_ci2 <- y_pred2[, c("lwr", "upr")]
y_ci2
```

```
##         lwr      upr
## 1  183.7838 190.3077
## 2  211.6735 217.6687
## 3  285.9757 290.7020
## 4  424.6851 428.2467
## 5  452.2722 455.9104
## 6  534.7395 539.1955
## 7  616.9317 622.7558
## 8  671.6509 678.5381
## 9  562.1579 567.0279
## 10 452.2722 455.9104
## 11 369.3400 373.0902
## 12 267.4126 272.4313
```