Problem Set 9 Solutions

MATH E-156: Mathematical Statistics

```r
rm( list = ls() )
```

```r
load( "Problem Set 9 R Objects.Rdata" )
```

Problem 1: Constructing a One-Sample t-Test

The data in `one.sample.t.test.data` comes from a normally distributed population with an unknown expected value, and we will perform a two-sided test of the null hypothesis $H_0 : \mu = 200$ using a one-sample t-test with a significance level of $\alpha = 0.05$.

Part (a): Sample size

How many observations are in the vector `one.sample.t.test.data`? Save this value in a variable, and report your answer with a `cat()` statement.

Solution

```r
sample.size <-
    length( one.sample.t.test.data )

cat( "Sample size:", sample.size )
```

```
## Sample size: 18
```

Part (b): Degrees of freedom

What are the appropriate degrees of freedom for this test? Save this value in a variable, and report your result using a `cat()` statement.

Solution

```r
degrees.of.freedom <- sample.size - 1

cat( "Degrees of freedom:", degrees.of.freedom )
```

```
## Degrees of freedom: 17
```

Part (c): Lower critical value

We wish to perform our test using a significance level of $\alpha = 0.05$. Calculate lower critical value L for this test. Store this value in a variable, and report your result using a `cat()` statement, rounding to 5 decimal places.

Solution

```
significance.level <- 0.05
```

```
lower.critical.value <-
    qt( significance.level / 2, degrees.of.freedom )

cat( "Lower critical value:",
     round( lower.critical.value, 5 ) )
```

```
## Lower critical value: -2.10982
```

Part (d): Upper critical value

We wish to perform our test using a significance level of $\alpha = 0.05$. Calculate the upper critical value U for this test. Store this value in a variable, and report your result using a `cat()` statement, rounding to 5 decimal places.

Solution

```
upper.critical.value <-
    qt( 1 - significance.level / 2, degrees.of.freedom )

cat( "Upper critical value:",
     round( upper.critical.value, 5 ) )
```

```
## Upper critical value: 2.10982
```

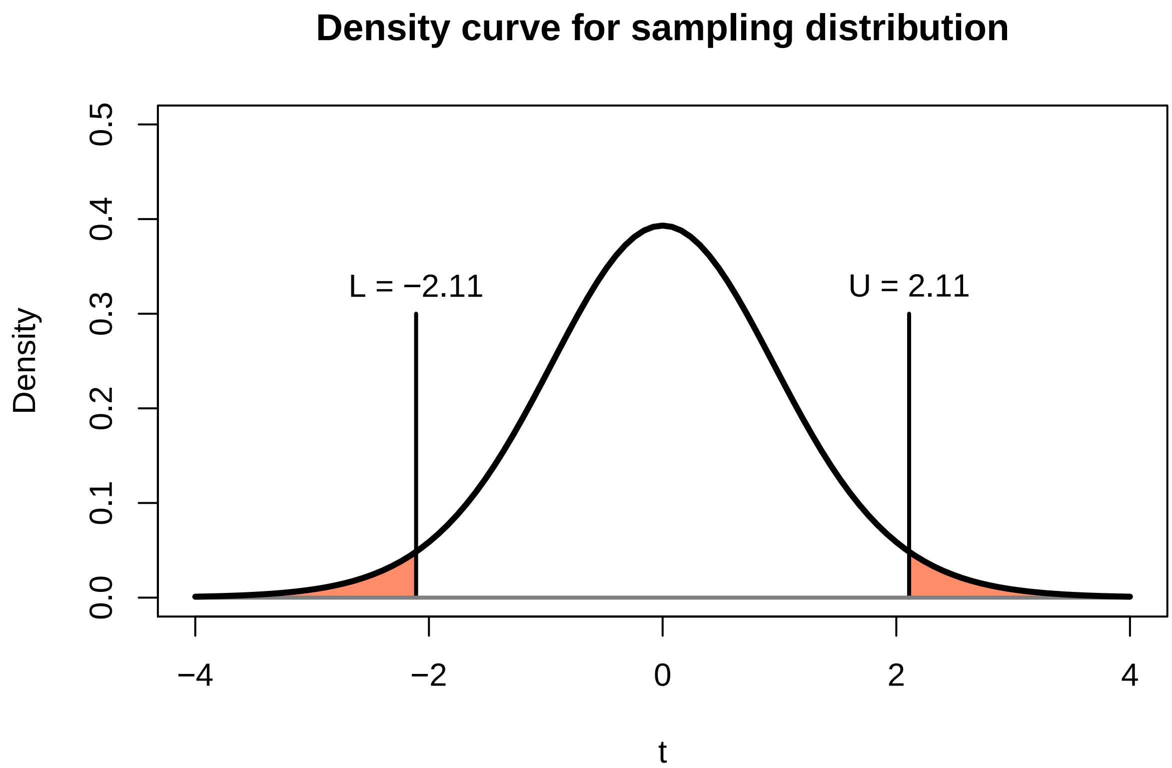
Part (e): Graphing the density curve

Draw a diagram showing the density curve for the sampling distribution of the studentized t statistic for this data. Indicate the lower and upper critical values with a vertical bar, and annotate these with test. Shade underneath the curve for the rejection region.

Solution

```
plot(
    x = NULL,
    xlim = c(-4, 4),
    ylim = c(0, 0.5),
    main = "Density curve for sampling distribution",
    xlab = "t",
    ylab = "Density"
)

shade.under.t.density.curve(
    initial.x = -4,
    final.x = lower.critical.value,
    degrees.of.freedom = degrees.of.freedom,
    fill.color = "salmon1"
)

segments(
```

```r
    lower.critical.value, 0,
    lower.critical.value, 0.3,
    lty = "solid",
    lwd = 2,
    col = "black"
)

text(
    lower.critical.value,
    0.33,
    paste( "L =", round( lower.critical.value, 2 ) )
)


shade.under.t.density.curve(
    initial.x = upper.critical.value,
    final.x = 4,
    degrees.of.freedom = degrees.of.freedom,
    fill.color = "salmon1"
)

segments(
    upper.critical.value, 0,
    upper.critical.value, 0.3,
    lty = "solid",
    lwd = 2,
    col = "black"
)

text(
    upper.critical.value,
    0.33,
    paste( "U =", round( upper.critical.value, 2 ) )
)


segments(
    -4, 0, 4, 0,
    lty = "solid",
    lwd = 2,
    col = "gray50"
)

curve(
    dt(x, df = degrees.of.freedom),
    lty = "solid",
    lwd = 3,
    col = "black",
    add = TRUE
)
```


Density curve for sampling distribution
Density
0.0
0.1
0.2
0.3
0.4
0.5
L = −2.11
U = 2.11
−4
−2
0
2
4
t

End of problem 1

Problem 2: Conducting the One-Sample t-Test

Now it's time to actually run the analysis!

Part (a)

What is the sample mean of `one.sample.t.test.data`? Store this value in a variable, and report your result using a `cat()` statement, rounding to 5 decimal places.

Solution

```
sample.mean <- mean( one.sample.t.test.data )

cat( "Sample mean:",
     round( sample.mean, 5 ) )
```

```
## Sample mean: 88.936
```

Part (b)

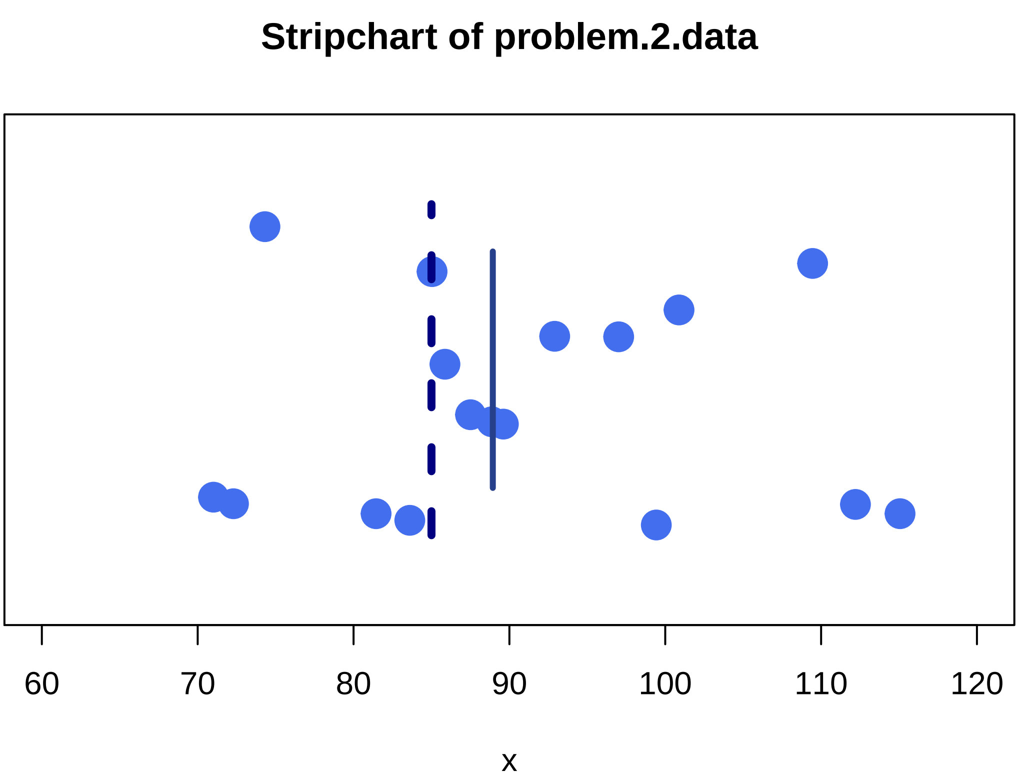
Construct a stripchart of the data in `one.sample.t.test.data`. Draw in the sample mean using a solid vertical bar, and then draw in the null hypothesis value $\mu_0 = 85$ using a dashed vertical line. Do you think this data constitutes strong evidence against the null hypothesis?

Solution

```
null.hypothesis.expected.value <- 85
```

```
stripchart(
    one.sample.t.test.data,
    xlim = c(60, 120),
    ylim = c(0, 2),
    main = "Stripchart of problem.2.data",
    xlab = "x",
    method = "jitter",
    jitter = 0.7,
    pch = 19,
    cex = 2,
    col = "royalblue2"
)

segments(
    sample.mean, 0.5, sample.mean, 1.5,
    lty = "solid",
    lwd = 3,
    col = "royalblue4"
)

segments(
    null.hypothesis.expected.value, 0.3,
    null.hypothesis.expected.value, 1.7,
    lty = "dashed",
```

```
    lwd = 4,
    col = "navy"
)
```



Part (c)

What is the sample variance of `one.sample.t.test.data`? Store this value in a variable, and report your result using a `cat()` statement, rounding to 5 decimal places.

Solution

```
sample.variance <- var( one.sample.t.test.data )

cat( "Sample variance:",
     round( sample.variance, 5 ) )
```

```
## Sample variance: 241.2274
```

Part (d)

Calculate the t statistic for this data, using the null hypothesis value of $\mu_0 = 85$. Save this value in a variable, and report your result using a `cat()` statement.

Solution

```r
t.score <-
    (sample.mean - null.hypothesis.expected.value) /
    sqrt( sample.variance/sample.size )

cat( "t score:",
     round( t.score, 5 ) )
```

```
## t score: 1.07517
```

Part (e)

Using the lower and critical values you calculate in parts (c) and (d) of problem 1, perform a two-sided test of the null hypothesis $H_0 : \mu = 85$ at the $\alpha = 0.05$ significance level.

Solution

Part (f)

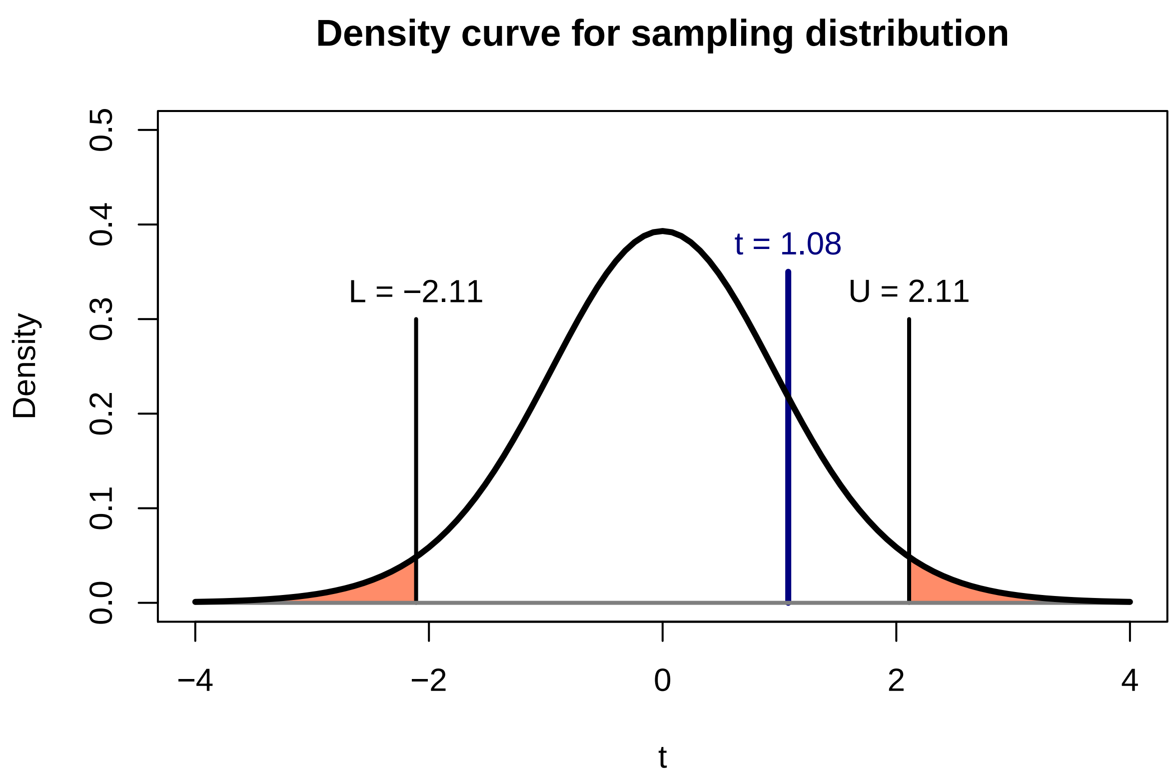
Copy your graph from part (e) of problem 1. Then add in a vertical line indicating the observed t statistic, and annotate it with text.

Solution

```r
plot(
    x = NULL,
    xlim = c(-4, 4),
    ylim = c(0, 0.5),
    main = "Density curve for sampling distribution",
    xlab = "t",
    ylab = "Density"
)

shade.under.t.density.curve(
    initial.x = -4,
    final.x = lower.critical.value,
    degrees.of.freedom = degrees.of.freedom,
    fill.color = "salmon1"
)

segments(
    lower.critical.value, 0,
    lower.critical.value, 0.3,
    lty = "solid",
    lwd = 2,
    col = "black"
)

text(
    lower.critical.value,
    0.33,
    paste( "L =", round( lower.critical.value, 2 ) )
)
```

```r
shade.under.t.density.curve(
    initial.x = upper.critical.value,
    final.x = 4,
    degrees.of.freedom = degrees.of.freedom,
    fill.color = "salmon1"
)

segments(
    upper.critical.value, 0,
    upper.critical.value, 0.3,
    lty = "solid",
    lwd = 2,
    col = "black"
)

text(
    upper.critical.value,
    0.33,
    paste( "U =", round( upper.critical.value, 2 ) )
)


segments(
    t.score, 0, t.score, 0.35,
    lty = "solid",
    lwd = 3,
    col = "navy"
)

text(
    t.score,
    0.38,
    paste( "t =", round(t.score, 2) ),
    col = "navy"
)


segments(
    -4, 0, 4, 0,
    lty = "solid",
    lwd = 2,
    col = "gray50"
)



curve(
    dt(x, df = degrees.of.freedom),
    lty = "solid",
    lwd = 3,
    col = "black",
    add = TRUE
```

```
)
```



Part (g)

Construct a two-sided 95% confidence interval for the true population expected value. Report your result using a `cat()` statement, rounding to 5 decimal places.

Solution

```
lower.confidence.interval.endpoint <-
    sample.mean +
    qt( significance.level / 2, df = degrees.of.freedom ) *
    sqrt( sample.variance / sample.size )

upper.confidence.interval.endpoint <-
    sample.mean +
    qt( 1 - significance.level / 2, df = degrees.of.freedom ) *
    sqrt( sample.variance / sample.size )

cat( "95% confidence interval: (",
     round( lower.confidence.interval.endpoint, 5 ),
     ", ",
     round( upper.confidence.interval.endpoint, 5 ),
     ")",
```

```r
    sep = ""
)
```

```
## 95% confidence interval: (81.21237, 96.65963)
```

Part (h)

Calculate the p-value for this test statistic.

```r
2 * pt( t.score, df = sample.size - 1, lower.tail = FALSE )
```

```
## [1] 0.2973211
```

Part (i)

Conduct the hypothesis test using the built-in R function `t.test`. How do the results of this analysis compare with your work in the previous parts of this problem?

Solution

Let's run the test:

```r
t.test(
    one.sample.t.test.data,
    mu = null.hypothesis.expected.value
)
```

```
##
##  One Sample t-test
##
## data:  one.sample.t.test.data
## t = 1.0752, df = 17, p-value = 0.2973
## alternative hypothesis: true mean is not equal to 85
## 95 percent confidence interval:
##  81.21237 96.65963
## sample estimates:
## mean of x
##    88.936
```

End of problem 2

Problem 3: Constructing a Two-Sample t-Test

The next two problems will be concerned with the two-sample t test, We using a two-tailed test, with a significance level of $\alpha = 0.05$. In this problem, we will construct the test, and in the next problem we will actually perform it. As in the lecture, we will focus on the case where the variances are assumed to be equal.

Part (a)

Determine the sample size of the data for Group 1, save it in a variable, and report it using a `cat()` statement. Then determine the sample size of the data for Group 2, save it in a variable, and report it using a `cat()` statement.

```
group.1.sample.size <-
    length( group.1.data )

cat( "Group 1 sample size:", group.1.sample.size )
```

```
## Group 1 sample size: 39
```

```
group.2.sample.size <-
    length( group.2.data )

cat( "Group 2 sample size:", group.2.sample.size )
```

```
## Group 2 sample size: 48
```

Part (b)

Calculate the appropriate degrees of freedom for a two-sample t-test. Store this value in a variable, and report it using a `cat()` statement.

Solution

```
degrees.of.freedom <-
    group.1.sample.size + group.2.sample.size - 2

cat( "Degrees of freedom:", degrees.of.freedom )
```

```
## Degrees of freedom: 85
```

Part (c)

Using a significance level of $\alpha = 0.05$, calculate L, the lower critical value for the test. Store this in a variable, and report it using a `cat()` statement, rounding to 5 decimal places.

Solution

```
significance.level <- 0.05
```

```r
lower.critical.value <-
    qt(
        significance.level / 2,
        degrees.of.freedom
    )

cat( "Lower critical value:",
     round( lower.critical.value, 5 ) )
```

```
## Lower critical value: -1.98827
```

Part (d)

Using a significance level of $\alpha = 0.05$, calculate U, the upper critical value for the test. Store this in a variable, and report it using a `cat()` statement, rounding to 5 decimal places.

Solution

```r
upper.critical.value <-
    qt(
        1 - significance.level / 2,
        degrees.of.freedom
    )

cat( "Upper critical value:",
     round( upper.critical.value, 5 ) )
```

```
## Upper critical value: 1.98827
```

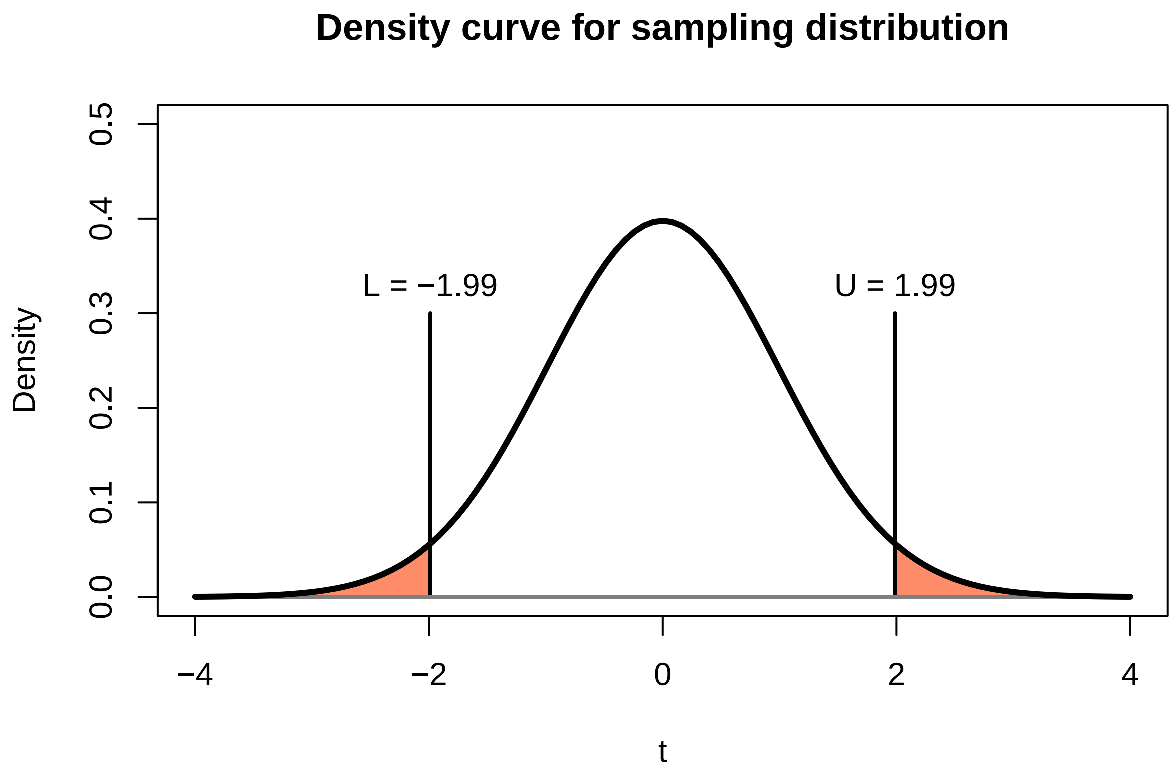
Part (e): Graphing the density curve

Draw a diagram showing the density curve for the sampling distribution of the studentized two-sample t statistic under the null hypothesis. Indicate the lower and upper critical values with a vertical bar, and annotate these with test. Shade underneath the curve for the rejection region.

Solution

```r
plot(
    x = NULL,
    xlim = c(-4, 4),
    ylim = c(0, 0.5),
    main = "Density curve for sampling distribution",
    xlab = "t",
    ylab = "Density"
)

shade.under.t.density.curve(
    initial.x = -4,
    final.x = lower.critical.value,
    degrees.of.freedom = degrees.of.freedom,
    fill.color = "salmon1"
)
```

```r
segments(
    lower.critical.value, 0,
    lower.critical.value, 0.3,
    lty = "solid",
    lwd = 2,
    col = "black"
)

text(
    lower.critical.value,
    0.33,
    paste( "L =", round( lower.critical.value, 2 ) )
)


shade.under.t.density.curve(
    initial.x = upper.critical.value,
    final.x = 4,
    degrees.of.freedom = degrees.of.freedom,
    fill.color = "salmon1"
)

segments(
    upper.critical.value, 0,
    upper.critical.value, 0.3,
    lty = "solid",
    lwd = 2,
    col = "black"
)

text(
    upper.critical.value,
    0.33,
    paste( "U =", round( upper.critical.value, 2 ) )
)


segments(
    -4, 0, 4, 0,
    lty = "solid",
    lwd = 2,
    col = "gray50"
)

curve(
    dt(x, df = degrees.of.freedom),
    lty = "solid",
    lwd = 3,
    col = "black",
    add = TRUE
)
```


Density curve for sampling distribution
Density
0.0
0.1
0.2
0.3
0.4
0.5
L = −1.99
U = 1.99
−4
−2
0
2
4
t

End of problem 3

Problem 4: Conducting the Two-Sample t-Test

Part (a)

Calculate the sample mean for Group 1. Store this value in a variable, and report it using a `cat()` statement. Then calculate the sample mean for Group 2. Store this value in a variable, and report it using a `cat()` statement.

Solution

```
group.1.sample.mean <-
    mean( group.1.data )

cat( "Group 1 sample mean:",
     round( group.1.sample.mean, 5 ) )
```

```
## Group 1 sample mean: 127.4566
```

```
group.2.sample.mean <-
    mean( group.2.data )

cat( "Group 2 sample mean:",
     round( group.2.sample.mean, 5 ) )
```

```
## Group 2 sample mean: 124.7385
```

Part (b)

Calculate the sample variance for Group 1. Store this value in a variable, and report it using a `cat()` statement. Then calculate the sample variance for Group 2. Store this value in a variable, and report it using a `cat()` statement.

Solution

```
group.1.sample.variance <-
    var( group.1.data )

cat( "Group 1 sample variance:",
     round( group.1.sample.variance, 5 ) )
```

```
## Group 1 sample variance: 31.67824
```

```
group.2.sample.variance <-
    var( group.2.data )

cat( "Group 2 sample variance:",
     round( group.2.sample.variance, 5 ) )
```

```
## Group 2 sample variance: 37.94528
```

Part (c)

Calculate S_p, the pooled estimate of the common population variance. Store this value in a variable, and report it using a `cat()` statement.

Solution

```
pooled.variance.estimate <-
    ( ( (group.1.sample.size - 1) * group.1.sample.variance ) +
          ( (group.2.sample.size - 1) * group.2.sample.variance ) ) /
    (group.1.sample.size + group.2.sample.size - 2)

cat( "Pooled variance estimate:",
     round( pooled.variance.estimate, 5 ) )
```

```
## Pooled variance estimate: 35.14355
```

Part (d)

Calculate the t statistic for this data. Store this in a variable, and report your result using a `cat()` statement.

Solution

```
t.statistic <-
    (group.1.sample.mean - group.2.sample.mean) /
    sqrt(
        pooled.variance.estimate *
            (1/group.1.sample.size + 1/group.2.sample.size)
    )

cat( "Observed t statistic:",
     round( t.statistic, 5 ) )
```

```
## Observed t statistic: 2.12681
```

Part (e)

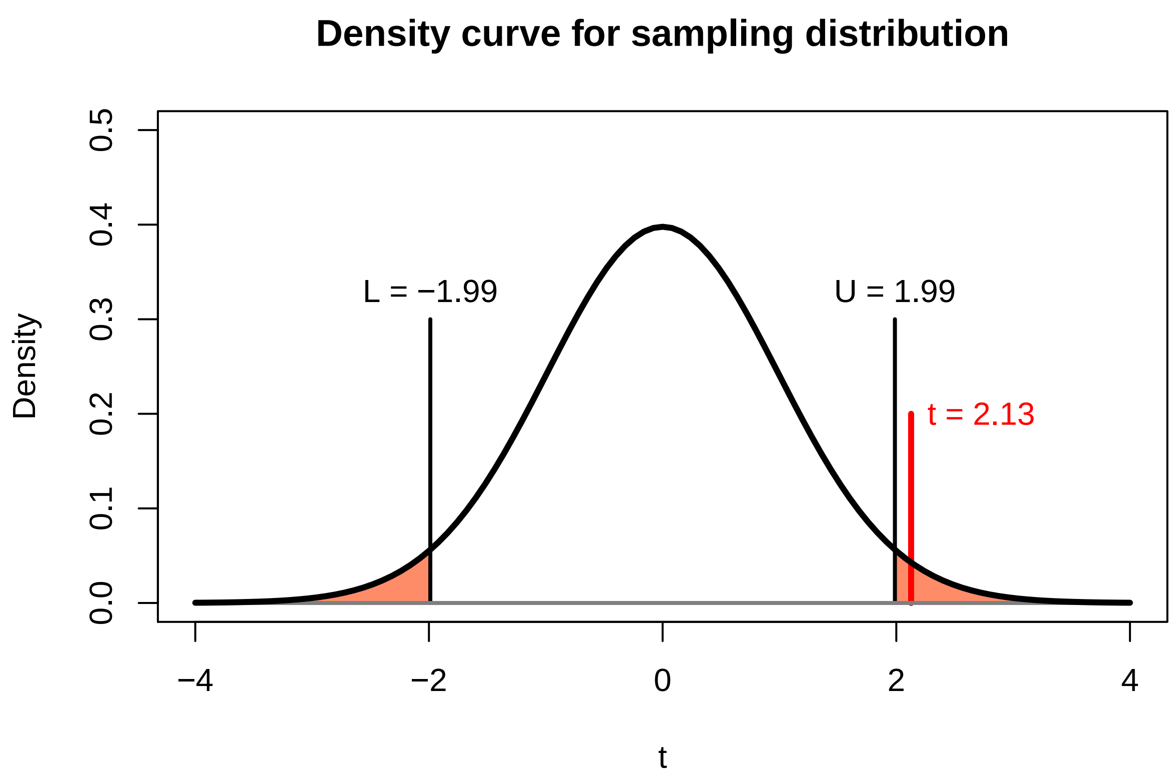
Copy the code from part (e) of problem 3 for the graph of the density curve of the test statistic under the null hypothesis. Then add a vertical line to indicate the observed t statistic, and annotate it with text.

Solution

```
plot(
    x = NULL,
    xlim = c(-4, 4),
    ylim = c(0, 0.5),
    main = "Density curve for sampling distribution",
    xlab = "t",
    ylab = "Density"
)

shade.under.t.density.curve(
```

```
    initial.x = -4,
    final.x = lower.critical.value,
    degrees.of.freedom = degrees.of.freedom,
    fill.color = "salmon1"
)

segments(
    lower.critical.value, 0,
    lower.critical.value, 0.3,
    lty = "solid",
    lwd = 2,
    col = "black"
)

text(
    lower.critical.value,
    0.33,
    paste( "L =", round( lower.critical.value, 2 ) )
)


shade.under.t.density.curve(
    initial.x = upper.critical.value,
    final.x = 4,
    degrees.of.freedom = degrees.of.freedom,
    fill.color = "salmon1"
)

segments(
    upper.critical.value, 0,
    upper.critical.value, 0.3,
    lty = "solid",
    lwd = 2,
    col = "black"
)

text(
    upper.critical.value,
    0.33,
    paste( "U =", round( upper.critical.value, 2 ) )
)


segments(
    t.statistic, 0,
    t.statistic, 0.2,
    lty = "solid",
    lwd = 3,
    col = "red"
)

text(
    t.statistic + 0.6,
```

```
    0.2,
    labels = paste( "t =", round( t.statistic, 2 ) ),
    col = "red"
)


segments(
    -4, 0, 4, 0,
    lty = "solid",
    lwd = 2,
    col = "gray50"
)

curve(
    dt(x, df = degrees.of.freedom),
    lty = "solid",
    lwd = 3,
    col = "black",
    add = TRUE
)
```



Part (f)

Using the critical values you calculated in problem 3, perform a two-sided test of the null hypothesis $h_0 : \mu_1 = \mu_2$ using a significance level of $\alpha = 0.05$. Does this data constitute strong evidence against the null hypothesis? Explain your answer with one or two sentences.

Solution

Part (g)

Calculate the lower endpoint of the 95% confidence interval for the difference of the true population expected values. Report your result using a `cat()` statement, rounding to 5 decimal places. Then calculate the upper endpoint of the 95% confidence interval for the difference of the true population expected values. Report your result using a `cat()` statement, rounding to 5 decimal places.

Solution

```
lower.confidence.interval.endpoint <-
    (group.1.sample.mean - group.2.sample.mean) +
    qt( significance.level /2, degrees.of.freedom ) *
    sqrt(
        pooled.variance.estimate *
            (1/group.1.sample.size + 1/group.2.sample.size)
    )

cat( "Lower confidence interval endpoint:",
     round( lower.confidence.interval.endpoint, 5 ) )
```

```
## Lower confidence interval endpoint: 0.17705
```

```
upper.confidence.interval.endpoint <-
    (group.1.sample.mean - group.2.sample.mean) +
    qt( 1 - significance.level /2, degrees.of.freedom ) *
    sqrt(
        pooled.variance.estimate *
            (1/group.1.sample.size + 1/group.2.sample.size)
    )

cat( "Upper confidence interval endpoint:",
     round( upper.confidence.interval.endpoint, 5 ) )
```

```
## Upper confidence interval endpoint: 5.25905
```

Part (h)

Calculate the two-sided p-value for this observed data. Report your result using a `cat()` statement, rounding to 5 decimal places.

Solution

```
p.value <-
    2 *
    pt(
        t.statistic,
        degrees.of.freedom,
        lower.tail = FALSE
```

```r
    )

p.value <-
    cat( "p-value:",
         round( p.value, 5 ) )
```

```
## p-value: 0.03633
```

Part (i)

Now use the built-in R function `t.test()` to conduct the two-sample t-test. How do the results of this analysis compare with your previous work?

Solution

```r
t.test(
    group.1.data,
    group.2.data,
    var.equal = TRUE
)
```

```
##
##  Two Sample t-test
##
## data:  group.1.data and group.2.data
## t = 2.1268, df = 85, p-value = 0.03633
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.1770516 5.2590485
## sample estimates:
## mean of x mean of y
##  127.4566  124.7385
```

End of problem 4

Problem 5: Constructing a One-Sample Variance Test

Now we're going to perform a one-sample test on a variance.

Part (a): Sample size

How many observations are contained in `one.sample.variance.test.data`? Save this value in a variable, and report your result using a `cat()` statement.

Solution

```
sample.size <- length( one.sample.variance.test.data )

cat( "Sample size:", sample.size )
```

```
## Sample size: 53
```

Part (b): Degrees of freedom

The data in `one.sample.variance.test.data` come from a normally distributed population with an unknown variance, and we will perform a two-sided test on the population variance. What are the appropriate degrees of freedom for this test? Save this value in a variable, and report your result using a `cat()` statement.

Solution

```
degrees.of.freedom <- sample.size - 1

cat( "Degrees of freedom:", degrees.of.freedom )
```

```
## Degrees of freedom: 52
```

Part (c): Lower critical value

Suppose we wish to perform our two-sided test using the standardized variance test statistic, with a significance level of $\alpha = 0.10$. Calculate the lower critical value L for this test. Store this value in a variable, and report your result using a `cat()` statement, rounding to 5 decimal places.

Solution

```
significance.level <- 0.10
```

```
lower.critical.value <-
    qchisq( significance.level / 2, degrees.of.freedom )

cat( "Lower critical value:",
     round( lower.critical.value, 5 ) )
```

```
## Lower critical value: 36.43709
```

Part (d): Upper critical value

Suppose we wish to perform our two-sided test using a significance level of $\alpha = 0.10$. Calculate the upper critical value U for this test. Store this value in a variable, and report your result using a `cat()` statement, rounding to 5 decimal places.

Solution

```
upper.critical.value <-
    qchisq( 1 - significance.level / 2, degrees.of.freedom )

cat( "Upper critical value:",
     round( upper.critical.value, 5 ) )
```

```
## Upper critical value: 69.83216
```

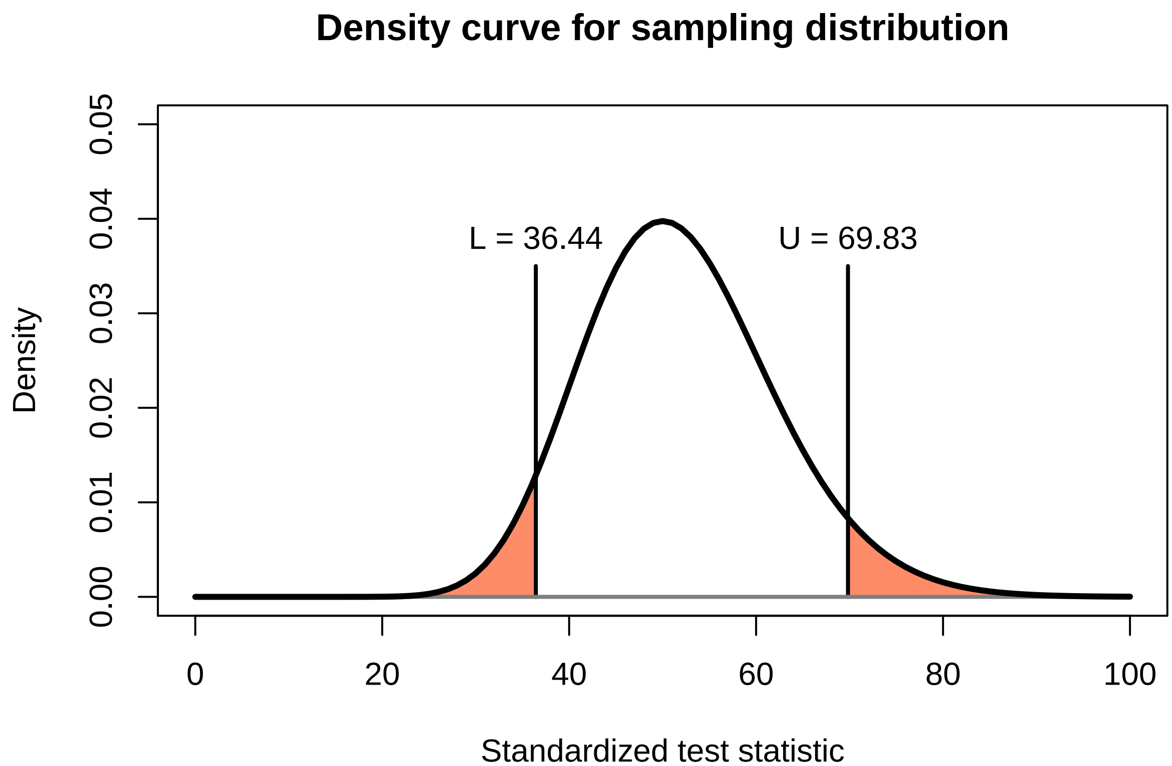
Part (e): Graphing the density curve

Draw a diagram showing the density curve for the sampling distribution of the standardized variance test statistic for this data. Indicate the lower and upper critical values with a vertical bar, and annotate these with test. Shade underneath the curve for the rejection region.

Solution

```
plot(
    x = NULL,
    xlim = c(0, 100),
    ylim = c(0, 0.05),
    main = "Density curve for sampling distribution",
    xlab = "Standardized test statistic",
    ylab = "Density"
)

shade.under.chisq.density.curve(
    initial.x = 0,
    final.x = lower.critical.value,
    degrees.of.freedom = degrees.of.freedom,
    fill.color = "salmon1"
)

segments(
    lower.critical.value, 0,
    lower.critical.value, 0.035,
    lty = "solid",
    lwd = 2,
    col = "black"
)

text(
    lower.critical.value,
    0.038,
    paste( "L =", round( lower.critical.value, 2 ) )
)
```

```r
shade.under.chisq.density.curve(
    initial.x = upper.critical.value,
    final.x = 100,
    degrees.of.freedom = degrees.of.freedom,
    fill.color = "salmon1"
)

segments(
    upper.critical.value, 0,
    upper.critical.value, 0.035,
    lty = "solid",
    lwd = 2,
    col = "black"
)

text(
    upper.critical.value,
    0.038,
    paste( "U =", round( upper.critical.value, 2 ) )
)


segments(
    0, 0, 100, 0,
    lty = "solid",
    lwd = 2,
    col = "gray50"
)

curve(
    dchisq(x, df = degrees.of.freedom),
    lty = "solid",
    lwd = 3,
    col = "black",
    add = TRUE
)
```


Density curve for sampling distribution
L = 36.44
U = 69.83
Density
0.00
0.01
0.02
0.03
0.04
0.05
0
20
40
60
80
100
Standardized test statistic

End of problem 5

Problem 6

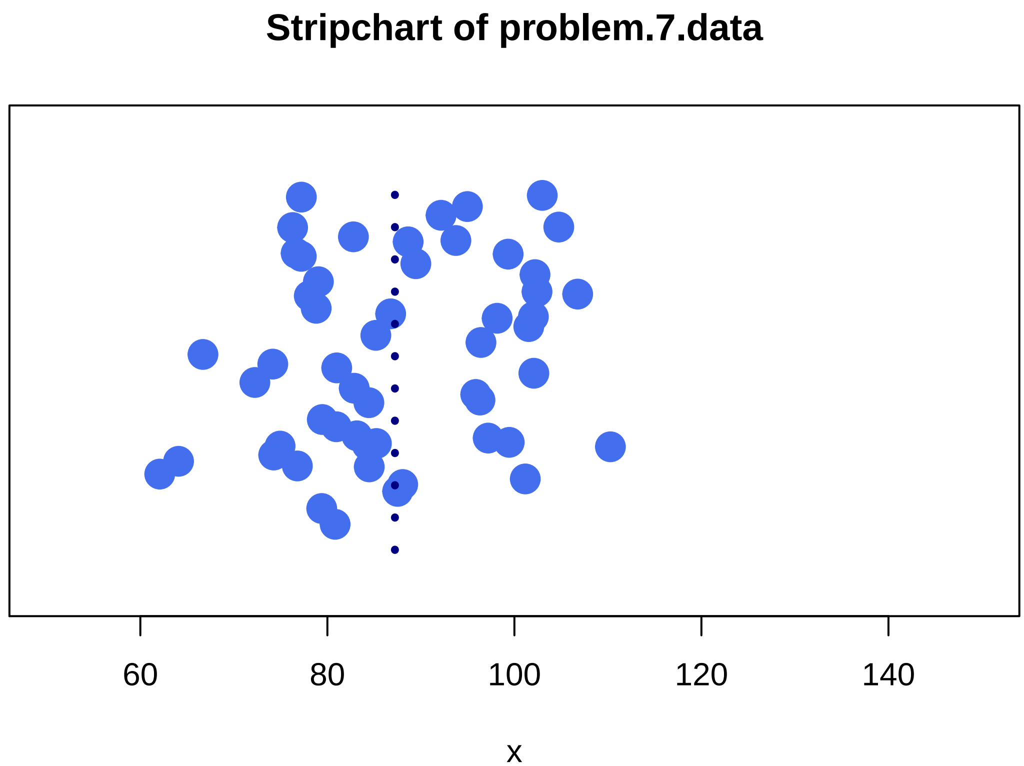
Now it's time to actually run the analysis!

Part (a)

Construct a stripchart of the data in `one.sample.variance.test.data`. Apply jitter to points, and include a vertical line indicating the sample mean. Do you think this data constitutes strong evidence against the null hypothesis?

Solution

```
stripchart(
    one.sample.variance.test.data,
    xlim = c(50, 150),
    ylim = c(0, 2),
    main = "Stripchart of problem.7.data",
    xlab = "x",
    method = "jitter",
    jitter = 0.7,
    pch = 19,
    cex = 2,
    col = "royalblue2"
)

sample.mean <-
    mean( one.sample.variance.test.data )

segments(
    sample.mean, 0.2,
    sample.mean, 1.8,
    lty = "dotted",
    lwd = 4,
    col = "navy"
)
```



Part (b)

What is the sample variance of `one.sample.variance.test.data`? Store this value in a variable, and report your result using a `cat()` statement, rounding to 5 decimal places.

Solution

```
sample.variance <- var( one.sample.variance.test.data )

cat( "Sample variance:",
     round( sample.variance, 5 ) )
```

```
## Sample variance: 135.5365
```

Part (c)

Calculate the standardized variance test statistic for this data, using the null hypothesis value of $\sigma^2 = 120$. Save this value in a variable, and report your result using a `cat()` statement.

Solution

```
null.hypothesis.variance <- 120
```

```
standardized.variance.test.statistic <-
    (sample.size - 1) * sample.variance /
    null.hypothesis.variance

cat( "Standardized variance test statistic:",
     round( standardized.variance.test.statistic, 5 ) )
```

```
## Standardized variance test statistic: 58.73248
```

Part (d)

Using the lower and critical values you calculated in parts (c) and (d) of problem 5, perform a two-sided test of the null hypothesis $H_0 : \mu = 200$ at the $\alpha = 0.10$ significance level.

Solution

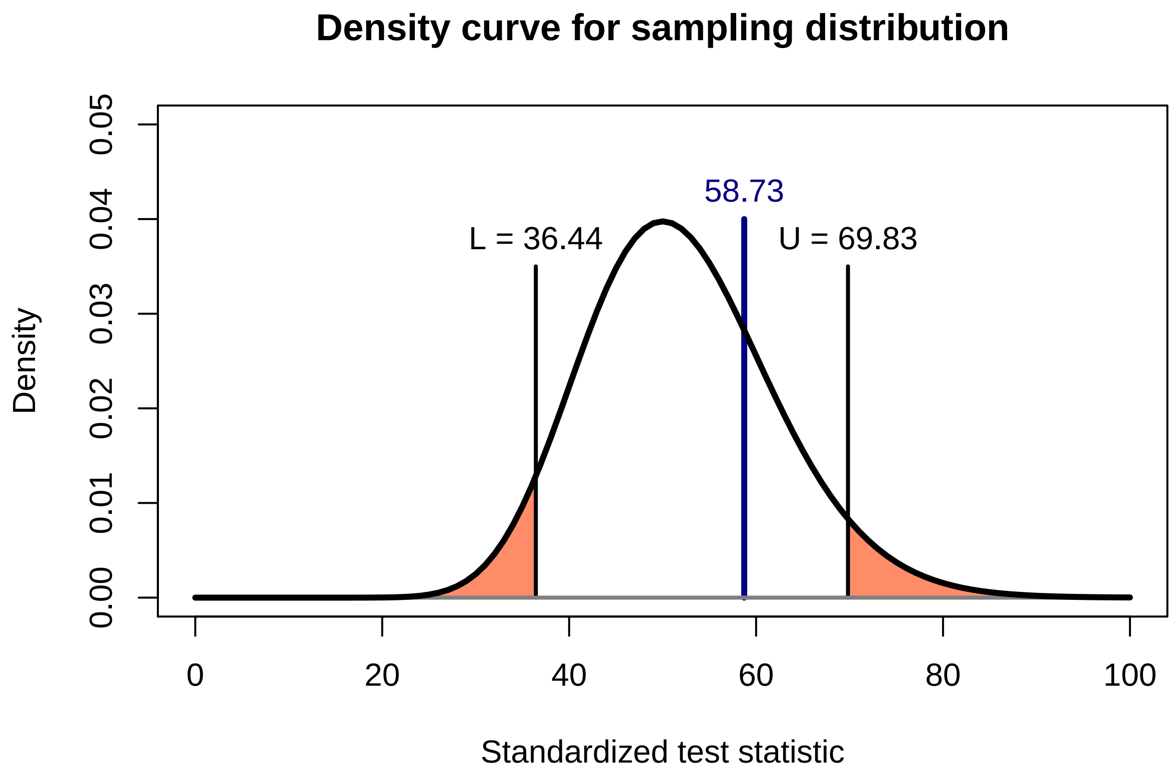
Part (e)

Copy your graph from part (e) of problem 5. Then add in a vertical line indicating the observed standardized variance test statistic, and annotate it with text.

Solution

```
plot(
    x = NULL,
    xlim = c(0, 100),
    ylim = c(0, 0.05),
    main = "Density curve for sampling distribution",
    xlab = "Standardized test statistic",
    ylab = "Density"
)

shade.under.chisq.density.curve(
    initial.x = 0,
    final.x = lower.critical.value,
    degrees.of.freedom = degrees.of.freedom,
    fill.color = "salmon1"
)

segments(
    lower.critical.value, 0,
    lower.critical.value, 0.035,
    lty = "solid",
    lwd = 2,
    col = "black"
)

text(
    lower.critical.value,
    0.038,
    paste( "L =", round( lower.critical.value, 2 ) )
)
```

```r
shade.under.chisq.density.curve(
    initial.x = upper.critical.value,
    final.x = 100,
    degrees.of.freedom = degrees.of.freedom,
    fill.color = "salmon1"
)

segments(
    upper.critical.value, 0,
    upper.critical.value, 0.035,
    lty = "solid",
    lwd = 2,
    col = "black"
)

text(
    upper.critical.value,
    0.038,
    paste( "U =", round( upper.critical.value, 2 ) )
)


segments(
    standardized.variance.test.statistic, 0,
    standardized.variance.test.statistic, 0.04,
    lty = "solid",
    lwd = 3,
    col = "navy"
)

text(
    standardized.variance.test.statistic,
    0.043,
    paste( round( standardized.variance.test.statistic, 2 ) ),
    col = "navy"
)


segments(
    0, 0, 100, 0,
    lty = "solid",
    lwd = 2,
    col = "gray50"
)

curve(
    dchisq(x, df = degrees.of.freedom),
    lty = "solid",
    lwd = 3,
    col = "black",
    add = TRUE
)
```



Part (f)

Construct a two-sided 90% confidence interval for the true population variance. Report your result using a `cat()` statement, rounding to 5 decimal places.

Solution

```
lower.confidence.interval.endpoint <-
    sample.variance * degrees.of.freedom /
    qchisq( 1 - significance.level / 2, df = degrees.of.freedom )

upper.confidence.interval.endpoint <-
    sample.variance * degrees.of.freedom /
    qchisq( significance.level / 2, df = degrees.of.freedom )

cat( "90% confidence interval: (",
     round( lower.confidence.interval.endpoint, 5 ),
     ", ",
     round( upper.confidence.interval.endpoint, 5 ),
     ")",
     sep = ""
)
```

```
## 90% confidence interval: (100.9262, 193.4265)
```

Part (g)

Calculate the p-value for this test statistic.

```r
2 * pchisq( standardized.variance.test.statistic,
            df = degrees.of.freedom,
            lower.tail = FALSE )
```

```
## [1] 0.4849181
```

End of problem 6

Problem 7: Constructing a Two-Sample Test on Variances

The next two problems will be concerned with the two-sample test on variances, We using a two-tailed test, with a significance level of $\alpha = 0.10$. In this problem, we will construct the test, and in the next problem we will actually perform it.

Part (a)

Determine the sample size of the data for Group A, save it in a variable, and report it using a `cat()` statement. Then determine the sample size of the data for Group 2, save it in a variable, and report it using a `cat()` statement.

```
group.a.sample.size <-
    length( group.a.data )

cat( "Group A sample size:", group.a.sample.size )
```

```
## Group A sample size: 54
```

```
group.b.sample.size <-
    length( group.b.data )

cat( "Group B sample size:", group.b.sample.size )
```

```
## Group B sample size: 73
```

Part (b)

Calculate the appropriate numerator and denominator degrees of freedom for a two-sample test on variances. Store these values in variable, and report each one separately using a `cat()` statement, rounding to 5 decimal places.

Solution

```
numerator.degrees.of.freedom <-
    group.a.sample.size - 1

cat( "Numerator degrees of freedom:",
     numerator.degrees.of.freedom )
```

```
## Numerator degrees of freedom: 53
```

```
denominator.degrees.of.freedom <-
    group.b.sample.size - 1

cat( "Denominator degrees of freedom:",
     denominator.degrees.of.freedom )
```

```
## Denominator degrees of freedom: 72
```

Part (c)

Using a significance level of $\alpha = 0.05$, calculate L, the lower critical value for the test. Store this in a variable, and report it using a `cat()` statement, rounding to 5 decimal places.

Solution

```
significance.level <- 0.05
```

```
lower.critical.value <-
    qf(
        significance.level / 2,
        df1 = numerator.degrees.of.freedom,
        df2 = denominator.degrees.of.freedom
    )

cat( "Lower critical value:",
     round( lower.critical.value, 5 ) )
```

```
## Lower critical value: 0.59672
```

Part (d)

Using a significance level of $\alpha = 0.05$, calculate U, the upper critical value for the test. Store this in a variable, and report it using a `cat()` statement, rounding to 5 decimal places.

Solution

```
upper.critical.value <-
    qf(
        1 - significance.level / 2,
        df1 = numerator.degrees.of.freedom,
        df2 = denominator.degrees.of.freedom
    )

cat( "Upper critical value:",
     round( upper.critical.value, 5 ) )
```

```
## Upper critical value: 1.64314
```

Part (e): Graphing the density curve

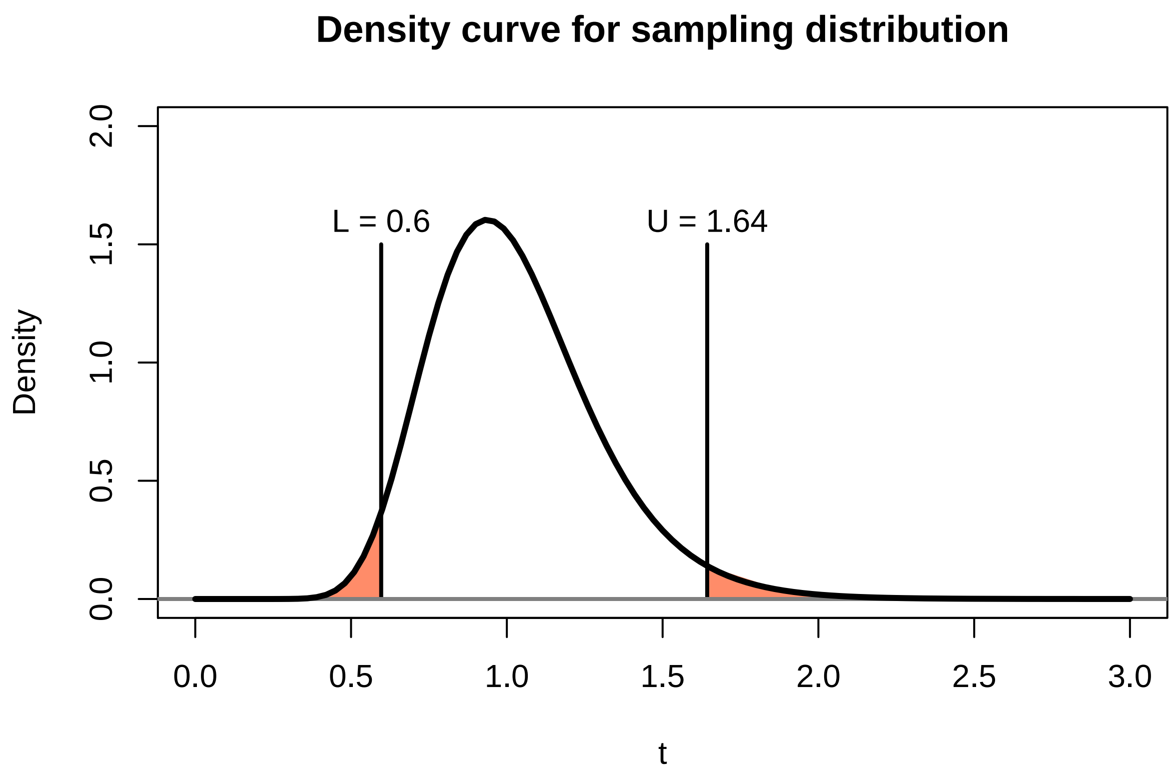
Draw a diagram showing the density curve for the sampling distribution of the F statistic under the null hypothesis. Indicate the lower and upper critical values with a vertical bar, and annotate these with test. Shade underneath the curve for the rejection region.

Solution

```
plot(
    x = NULL,
    xlim = c(0, 3),
    ylim = c(0, 2),
```

```r
    main = "Density curve for sampling distribution",
    xlab = "t",
    ylab = "Density"
)

shade.under.f.density.curve(
    initial.x = 0,
    final.x = lower.critical.value,
    df1 = numerator.degrees.of.freedom,
    df2 = denominator.degrees.of.freedom,
    fill.color = "salmon1"
)

segments(
    lower.critical.value, 0,
    lower.critical.value, 1.5,
    lty = "solid",
    lwd = 2,
    col = "black"
)

text(
    lower.critical.value,
    1.6,
    paste( "L =", round( lower.critical.value, 2 ) )
)


shade.under.f.density.curve(
    initial.x = upper.critical.value,
    final.x = 50,
    df1 = numerator.degrees.of.freedom,
    df2 = denominator.degrees.of.freedom,
    fill.color = "salmon1"
)

segments(
    upper.critical.value, 0,
    upper.critical.value, 1.5,
    lty = "solid",
    lwd = 2,
    col = "black"
)

text(
    upper.critical.value,
    1.6,
    paste( "U =", round( upper.critical.value, 2 ) )
)


segments(
    -4, 0, 4, 0,
```

```
    lty = "solid",
    lwd = 2,
    col = "gray50"
)

curve(
    df(
        x,
        df1 = numerator.degrees.of.freedom,
        df2 = denominator.degrees.of.freedom
    ),
    lty = "solid",
    lwd = 3,
    col = "black",
    add = TRUE
)
```



End of problem 7

Problem 8: Conducting the Two-Sample Test on Variances

Part (a)

Calculate the sample variance for Group A. Store this value in a variable, and report it using a `cat()` statement.

Solution

```
group.a.sample.variance <-
    var( group.a.data )

cat( "Group A sample variance:",
     round( group.a.sample.variance, 5 ) )
```

```
## Group A sample variance: 55.89073
```

Part (b)

Calculate the sample variance for Group B. Store this value in a variable, and report it using a `cat()` statement.

Solution

```
group.b.sample.variance <-
    var( group.b.data )

cat( "Group B sample variance:",
     round( group.b.sample.variance, 5 ) )
```

```
## Group B sample variance: 105.0201
```

Part (c)

Calculate the F statistic for this observed data under the null hypothesis $H_0 : \sigma_A^2 = \sigma_B^2$. Store this value in a variable, and report it using a `cat()` statement.

Solution

```
f.statistic <-
    group.a.sample.variance /
    group.b.sample.variance

cat( "F statistic:",
     round( f.statistic, 5 ) )
```

```
## F statistic: 0.53219
```

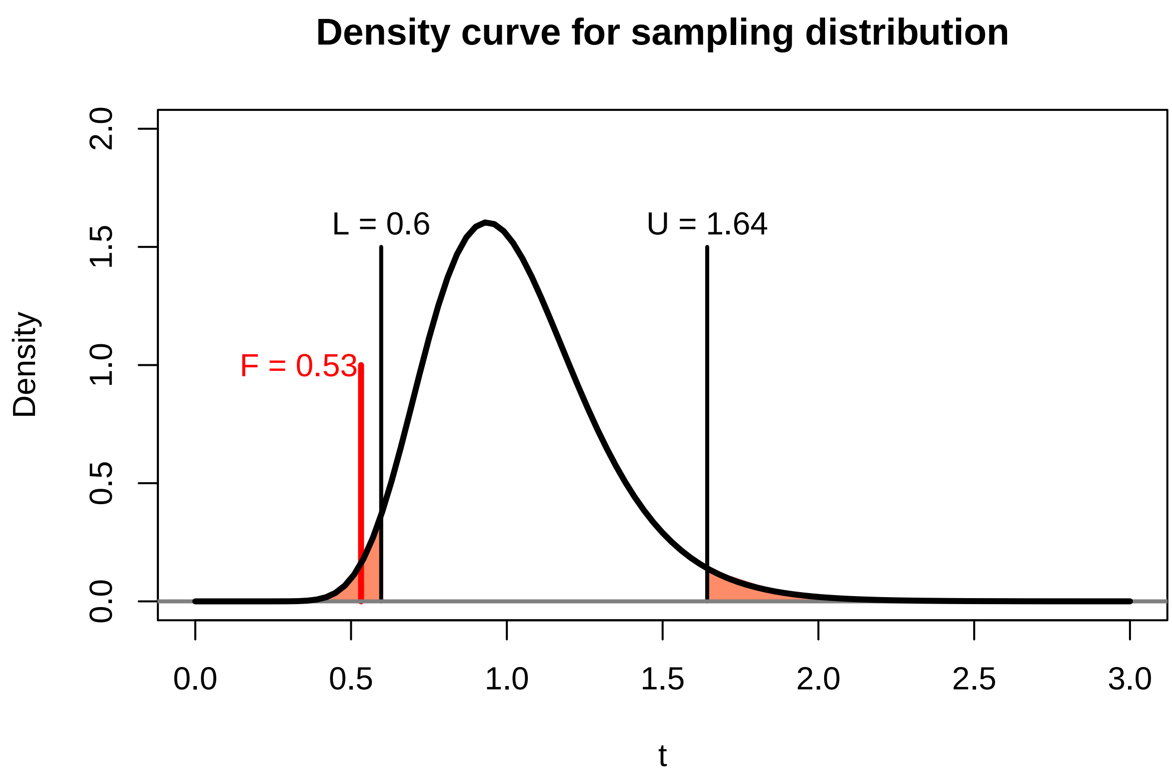
Part (d)

Copy the code from part (e) of problem 7 for the graph of the density curve of the test statistic under the null hypothesis. Then add a vertical line to indicate the observed F statistic, and annotate it with text.

Solution

```
plot(
    x = NULL,
    xlim = c(0, 3),
    ylim = c(0, 2),
    main = "Density curve for sampling distribution",
    xlab = "t",
    ylab = "Density"
)

shade.under.f.density.curve(
    initial.x = 0,
    final.x = lower.critical.value,
    df1 = numerator.degrees.of.freedom,
    df2 = denominator.degrees.of.freedom,
    fill.color = "salmon1"
)


segments(
    f.statistic, 0,
    f.statistic, 1,
    lty = "solid",
    lwd = 3,
    col = "red"
)




text(
    x = f.statistic - 0.2,
    y = 1,
    labels =
        paste( "F =", round(f.statistic, 2) ),
    col = "red"
)





segments(
    lower.critical.value, 0,
    lower.critical.value, 1.5,
    lty = "solid",
    lwd = 2,
    col = "black"
)
```

```
text(
    lower.critical.value,
    1.6,
    paste( "L =", round( lower.critical.value, 2 ) )
)


shade.under.f.density.curve(
    initial.x = upper.critical.value,
    final.x = 50,
    df1 = numerator.degrees.of.freedom,
    df2 = denominator.degrees.of.freedom,
    fill.color = "salmon1"
)

segments(
    upper.critical.value, 0,
    upper.critical.value, 1.5,
    lty = "solid",
    lwd = 2,
    col = "black"
)

text(
    upper.critical.value,
    1.6,
    paste( "U =", round( upper.critical.value, 2 ) )
)


segments(
    -4, 0, 4, 0,
    lty = "solid",
    lwd = 2,
    col = "gray50"
)

curve(
    df(
        x,
        df1 = numerator.degrees.of.freedom,
        df2 = denominator.degrees.of.freedom
    ),
    lty = "solid",
    lwd = 3,
    col = "black",
    add = TRUE
)
```



Part (f)

Using the critical values you calculated in problem 3, perform a two-sided test of the null hypothesis h_0 : $\sigma_1^2 = \sigma_2^2$ using a significance level of $\alpha = 0.05$. Does this data constitute strong evidence against the null hypothesis? Explain your answer with one or two sentences.

Solution

The observed value of the F statistic was $F = 0.53$, and this is less than the lower critical value $L = 0.60$. Thus, we reject the null hypothesis, and conclude that this data constitutes strong evidence against the null hypothesis.

Part (g)

Calculate the lower endpoint of the 95% confidence interval for the ratio of the true population variances. Report your result using a `cat()` statement, rounding to 5 decimal places. Then calculate the upper endpoint of the 95% confidence interval for the ratio of the true population variances. Report your result using a `cat()` statement, rounding to 5 decimal places.

Solution

```
lower.confidence.interval.endpoint <-
    (group.a.sample.variance / group.b.sample.variance) /
    qf(
        1 - significance.level / 2,
```

```
        df1 = numerator.degrees.of.freedom,
        df2 = denominator.degrees.of.freedom
    )

cat( "Lower confidence interval endpoint:",
     round( lower.confidence.interval.endpoint, 5 ) )
```

```
## Lower confidence interval endpoint: 0.32389
```

```
upper.confidence.interval.endpoint <-
    (group.a.sample.variance / group.b.sample.variance) /
    qf(
        significance.level / 2,
        df1 = numerator.degrees.of.freedom,
        df2 = denominator.degrees.of.freedom
    )

cat( "Upper confidence interval endpoint:",
     round( upper.confidence.interval.endpoint, 5 ) )
```

```
## Upper confidence interval endpoint: 0.89186
```

Part (h)

Calculate the two-sided p-value for this observed data. Report your result using a `cat()` statement, rounding to 5 decimal places.

Solution

```
p.value <-
    2 *
    pf(
        f.statistic,
        df1 = numerator.degrees.of.freedom,
        df2 = denominator.degrees.of.freedom
    )

p.value <-
    cat( "p-value:",
         round( p.value, 5 ) )
```

```
## p-value: 0.0171
```

Part (i)

Now use the built-in R function `var.test()` to conduct the two-sample test on variances. How do the results of this analysis compare with your previous work?

Solution

```r
var.test(
    group.a.data,
    group.b.data
)
```

```
##
##  F test to compare two variances
##
## data:  group.a.data and group.b.data
## F = 0.53219, num df = 53, denom df = 72, p-value = 0.0171
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.3238868 0.8918624
## sample estimates:
## ratio of variances
##          0.5321907
```