Partial answer key

7.6

a.

The box plot looks symmetric with one outlier on the top, and we don’t see any severe departure from normality.



> speed <- read.table("C:/sdata/vehicle.txt", header=F, sep=" ")

> boxplot(x = speed, col = "lightblue", main = "Box plot", ylab = "Speed(mph)")

b.

The 95% CI on σ is (8.17, 10.81).

> alpha<-0.05

> n <- 100

> #Interval

> lower<-sqrt((n - 1)\*var(speed)/qchisq(p = 1 - alpha/2, df = n - 1))

> upper<-sqrt((n - 1)\*var(speed)/qchisq(p = alpha/2, df = n - 1))

> data.frame(lower, upper)

 V1 V1.1

V1 8.17154 10.81162

c.

The level of significance is α = 0.05.

Confidence interval method:

1.

Ho: σ ≤ 10
Ha: σ > 10

2.

The 95% one-sided CI on σ is (8.34, ∞).

> sqrt((n - 1)\*var(speed)/qchisq(p = 1 - alpha, df = n - 1))

 V1

V1 8.34207

3.

Do not reject Ho because 10 is in the interval for σ.

4.

There is not sufficient evidence to indicate the standard deviation in vehicle speeds exceeds 10 mph.

Test statistic method:

1.

Ho: σ ≤ 10
Ha: σ > 10

2.



> sigma.sq<-10^2

> chisquare<-(n - 1)\*var(speed)/sigma.sq

> chisquare

 V1

V1 85.75258

3.



> qchisq(p = 1 - alpha, df = n - 1)

[1] 123.2252



4.

Do not reject Ho because 85.8 < 123.2.

5.

There is not sufficient evidence to indicate the standard deviation in vehicle speeds exceeds 10 mph.

P-value method:

1.

Ho: σ ≤ 10
Ha: σ > 10

2.

P(X > χ2) = 0.826

> 1 - pchisq(q = chisquare, df = n - 1)

 V1

V1 0.8261634

3.

α = 0.05.

4.

Do not reject Ho because 0.826>0.05.

5.

There is not sufficient evidence to indicate the standard deviation in vehicle speeds exceeds 10 mph.

d. We can apply this inference to all cars on that section of interstate at comparable times of day.

7.7

a.

The distribution is a little skewed to the right, but we don’t see any severe departure from normality.



> reading <- read.table("C:/sdata/reading.txt", header=F, sep=" ")

> boxplot(x = reading, col = "lightblue", main = "Box plot", ylab = "Reading scores")

> stripchart(x = reading, method = "jitter", vertical = TRUE, pch = 1, col = "red", add =

 TRUE)

b.

The 99% CI on σ2 is (89.18, 162.41).

> lower<-(n - 1)\*var(reading)/qchisq(p = 1 - alpha/2, df = n - 1)

> upper<-(n - 1)\*var(reading)/qchisq(p = alpha/2, df = n - 1)

> data.frame(lower, upper)

 V1 V1.1

V1 89.18238 162.4118

c.

The level of significance is chosen to be α = 0.01.

Confidence interval method:

1.

Ho: σ2 ≤ 90
Ha: σ2 > 90

2.

A 99% one-sided C.I on σ2 is (91.6, ∞)

> (n - 1)\*var(reading)/qchisq(p = 1 - alpha, df = n - 1)

 V1

V1 91.5681

3.

Reject Ho because 90 is not in the interval for σ2.

4.

There is enough evidence to indicate the variance of reading scores is greater than 90.

Test statistic method:

1.

Ho: σ2 ≤ 90
Ha: σ2 > 90

2.



> sigma.sq<-90

> chisquare<-(n - 1)\*var(reading)/sigma.sq

> chisquare

 V1

V1 195.4196

3.



> qchisq(p = 1 - alpha, df = n - 1)

[1] 192.0730



4.

Reject Ho because 192.1 < 195.4196

5.

There is enough evidence to indicate the variance of reading scores is greater than 9.

P-value method:

1.

Ho: σ2 ≤ 90
Ha: σ2 > 90

2.

P(X > χ2) = 0.006

> 1 - pchisq(q = chisquare, df = n - 1)

 V1

V1 0.006

3.

α = 0.01.

4.

Reject Ho because 0.006 < 0.01.

5.

There is enough evidence to indicate the variance of reading scores is greater than 90.

7.15

a.

95% CI on σold is (0.212, 0.344).

> machine <- read.csv("C:/sdata/exp07-15.txt", header=T, sep=" ")

> alpha<-0.05

> n1 <- n2 <- 40

> var.old <- var(machine$X.OLD.)

> var.new <- var(machine$X.NEW.)

> #old

> lower<-sqrt((n1 - 1)\*var.old/qchisq(p = 1 - alpha/2, df = n1 - 1))

> upper<-sqrt((n1 - 1)\*var.old/qchisq(p = alpha/2, df = n1 - 1))

> data.frame(lower, upper)

 lower upper

1 0.2192302 0.3436435

95% CI on σnew is (0.137, 0.197).

> lower<-sqrt((n - 1)\*var.new/qchisq(p = 1 - alpha/2, df = n2 - 1))

> upper<-sqrt((n - 1)\*var.new/qchisq(p = alpha/2, df = n2 - 1))

> data.frame(lower, upper)

 lower upper

1 0.1592825 0.2496755

b.

Use a level of significance is α = 0.05.

Confidence interval method:

1.

Ho:  ≤ 1
Ha:  > 1

2.

A 95% one-sided C.I for  is (3.23, ∞)

> var.old/var.new \* qf(p = 1 - alpha, df1 = n1 - 1, df2 = n2 - 1)

[1] 3.228885

3.

Reject Ho because 1 is not in the interval.

4.

There is sufficient evidence that the new type of machine has less variability of fills than the old machine.

Test statistic method:

1.

Ho:  ≤ 1
Ha:  > 1

2.

$$f=\frac{s\_{old}^{2}}{s\_{new}^{2}}=1.89$$

> F.stat<-var.old/var.new

> F.stat

[1] 1.894368

3.

$$F\_{0.95,41,41}=1.70$$

> qf(p = 1 - alpha, df1 = n1 - 1, df2 = n2 - 1)

[1] 1.704465



4.

Reject Ho because 1.70 < 1.89

5.

There is sufficient evidence that the new type of machine has less variability of fills than the old machine.

P-value method:

1.

Ho:  ≤ 1
Ha:  > 1

2.

P(X > 1.89) = 0.0034

> 1 - pf(q = F.stat, df1 = n1 - 1, df2 = n2 - 1)

[1] 0.02466079

3.

α = 0.05.

4.

Reject Ho because 0.025 < 0.05.

5.

There is sufficient evidence that the new type of machine has less variability of fills than the old machine.

c.

The box plots indicate the samples for both machine types are normally distributed.



#Suppose x1 and y1 contain the data for old and new, respectively

ex7.15<-data.frame(fill=c(x1,y1), location=rep(c("old", "new"),

 each=40))

boxplot(formula = fill ~ location, data = ex7.15, col = "lightblue",

 main = "Box and dot plot", ylab = "amount of fill(ounces)", xlab =

 "machine type", pars = list(outpch=NA))

stripchart(x = fill ~ location, data = ex7.15, method = "jitter", vertical

 = TRUE, pch = 1, col="red", add=T)

7.20

a.

There is no evidence against normality from the plots.



> # Suppose x and y contain the data for brand I and II, respectively.

> ex7.20 <- data.frame(miles=c(x,y), brand=rep(c("I", "II"), each = 10))

> head(ex7.20)

 miles brand

1 38.9 I

2 39.7 I

3 42.3 I

4 39.5 I

5 39.6 I

6 35.6 I

> boxplot(formula = miles ~ brand, data = ex7.20, col = "lightblue",

 main = "Longevity of Two Brands of Tires", ylab = "Miles to Tire Wearout (1000

 miles)", xlab = "Brand of Tires")

> stripchart(x = miles ~ brand, data = ex7.20, method = "jitter", vertical = TRUE, pch =

 1, col = "red", add = T)

b.

The 95% CI for σ1 (Brand I) is (1.344, 3.568). The 95% CI for σ2 (Brand II) is (3.838, 10.185).

> alpha<-0.05

> n1 <- n2 <- 10

> #Interval

> lower<-sqrt((n1 - 1)\*var(x)/qchisq(p = 1 - alpha/2, df = n1 - 1))

> upper<-sqrt((n1 - 1)\*var(x)/qchisq(p = alpha/2, df = n1 - 1))

> data.frame(lower, upper)

 lower upper

1 1.344146 3.567551

> #Interval

> lower<-sqrt((n2 - 1)\*var(y)/qchisq(p = 1 - alpha/2, df = n2 - 1))

> upper<-sqrt((n2 - 1)\*var(y)/qchisq(p = alpha/2, df = n2 - 1))

> data.frame(lower, upper)

 lower upper

1 3.837527 10.18533

The 95% CI for μ1 (Brand I) is (37.392, 40.188). The 95% CI for μ2 (Brand II) is (36.679, 44.661).

> t.test(x = x, conf.level = 0.95)

 One Sample t-test

data: x

t = 62.7708, df = 9, p-value = 3.334e-13

alternative hypothesis: true mean is not equal to 0

95 percent confidence interval:

 37.39207 40.18793

sample estimates:

mean of x

 38.79

> t.test(x = y, conf.level = 0.95)

 One Sample t-test

data: y

t = 23.0519, df = 9, p-value = 2.586e-09

alternative hypothesis: true mean is not equal to 0

95 percent confidence interval:

 36.67893 44.66107

sample estimates:

mean of x

40.67

c.

For comparing the population variances of the two brands:

The level of significance is α = 0.05.

Confidence interval method:

1.

Ho:  = 1
Ha:  ≠ 1

2.

A 95% CI on  is (0.03, 0.49).

> var.test(x = x, y = y, conf.level = 0.95)

 F test to compare two variances

data: x and y

F = 0.1227, num df = 9, denom df = 9, p-value = 0.004505

alternative hypothesis: true ratio of variances is not equal to 1

95 percent confidence interval:

 0.03047313 0.49392768

sample estimates:

ratio of variances

 0.1226846

3.

Reject Ho because 1 is not in the interval.

4.

There is sufficient evidence to indicate the population variances of the two tire brands are different.

Make sure you understand why this result makes sense relative to the box and dot plots.

Test statistic method:

1.

Ho:  = 1
Ha:  ≠ 1

2.

$$f=\frac{s\_{I}^{2}}{s\_{II}^{2}}=0.123$$

3.



> qf(p = alpha, df1 = n1 - 1, df2 = n2 - 1)

[1] 0.3145749

> qf(p = 1 - alpha, df1 = n1 - 1, df2 = n2 - 1)

[1] 3.178893



4.

Reject Ho because 0.123 < 0.315.

5.

There is sufficient evidence to indicate the population variances of the two tire brands are different.

P-value method:

1.

Ho:  = 1
Ha:  ≠ 1

2.

The p-value is 0.004.

3.

α = 0.05.

4.

Reject Ho because 0.004 < 0.05.

5.

There is sufficient evidence to indicate the population variances of the two tire brands are different.

For comparing the two population means, the hypotheses are:

Ho: μI - μII = 0
Ha: μI - μII ≠ 0

We can use the following code to perform the hypothesis tests. The conclusion is that there is not enough evidence to say the mean miles till tire wearout for the two brands are different. Be sure to use all three methods on your own, like what you did in chapter 6.

> t.test(x = x, y = y, var.equal = FALSE, conf.level = 0.95)

 Welch Two Sample t-test

data: x and y

t = -1.0057, df = 11.176, p-value = 0.3358

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

 -5.986589 2.226589

sample estimates:

mean of x mean of y

 38.79 40.67

7.22

a.

The level of significance is α = 0.05.

Confidence interval method:

1.

Ho:  ≥ 1
Ha:  < 1

2.

A 95% one-sided C.I for  is (0, 1.158)

> ex7.22 <- read.table("C:/sdata/ex7-22.txt", header=T, sep=",")

> x<-as.numeric(ex7.22[1,])

> y<-as.numeric(ex7.22[2,])

> ex7.22 <- data.frame(dollar=c(x,y), portfolio=rep(c(1, 2), each = 10))

> head(ex7.22)

 dollar portfolio

1 130 1

2 135 1

3 135 1

4 131 1

5 129 1

6 135 1

> var.test(x = x, y = y, conf.level = 0.95, alternative = "less")

 F test to compare two variances

data: x and y

F = 0.3642, num df = 9, denom df = 9, p-value = 0.07425

alternative hypothesis: true ratio of variances is less than 1

95 percent confidence interval:

 0.000000 1.157770

sample estimates:

ratio of variances

 0.3642053

3.

Do not reject Ho because 1 is not in the interval.

4.

There is not sufficient evidence to conclude that portfolio 2 has a higher risk than portfolio 1.

Test statistic method:

1.

Ho:  ≥ 1
Ha:  < 1

2.

$$f=\frac{s\_{1}^{2}}{s\_{2}^{2}}=0.364$$

3.



> alpha<-0.05

> n1 <- n2 <- 10

> qf(p = alpha, df1 = n1 - 1, df2 = n2 - 1)

[1] 0.3145749



4.

Do not reject Ho because 0.364 > 0.315.

5.

There is not sufficient evidence to conclude that portfolio 2 has a higher risk than portfolio 1.

P-value method:

1.

Ho:  ≥ 1
Ha:  < 1

2.

The p value is 0.074.

3.

α = 0.05.

4.

Do not reject Ho because 0.074 > 0.05.

5.

There is not sufficient evidence to conclude that portfolio 2 has a higher risk than portfolio 1.

b.

The 95% CI for  is (0.09, 1.47).

> var.test(x = x, y = y, conf.level = 0.95)

 F test to compare two variances

data: x and y

F = 0.3642, num df = 9, denom df = 9, p-value = 0.1485

alternative hypothesis: true ratio of variances is not equal to 1

95 percent confidence interval:

 0.09046343 1.46628824

sample estimates:

ratio of variances

 0.3642053

The 95% CI for  is then (0.3, 1.21).

c.

Normality seems to be reasonable for both portfolios.



> boxplot(formula = dollar ~ portfolio, data = ex7.22, col = "lightblue",

 main = "Comparison of Risk of Two Portfolios", ylab = "Yearly Returns (thousands

 of dollars)", xlab = "Portfolio type")

> stripchart(x = dollar ~ portfolio, data = ex7.22, method = "jitter", vertical = TRUE,

 pch = 1, col = "red", add = T)