Tutorial 4 Solutions

Question 1

(10.64) A coin-operated soft drink machine was designed to discharge, on the average, 7 ounces of beverage per cup. In a test of the machine, ten cupfuls of beverage were drawn from the machine and measured. The mean and standard deviation of the 10 measurements were 7.1 ounces and 0.12 ounces, respectively.

(a) Do these data present sufficient evidence to indicate that the mean discharge differs from 7 ounces? Use $\alpha = 0.05$.

The hypotheses we wish to test are:

$$H_0 : \mu = 7 \quad \text{vs} \quad H_1 : \mu \neq 7.$$

Assuming that the data came from a normal distribution,

$$T = \frac{\overline{X} - \mu_0}{S/\sqrt{n}} = \frac{\overline{X} - 7}{S/\sqrt{n}} \sim t_{n-1}.$$

Plugging in the given values, the value of our test statistic is

$$t = \frac{7.1 - 7}{0.12/\sqrt{10}} = 2.635.$$

For a two-tailed test, we reject H_0 if $|t| > t_{n-1,\,\alpha/2}$, where $\alpha/2$ is the area to the right.

$$t_{n-1,\,\alpha/2} = t_{9,\,0.025} = 2.262.$$

Since $2.635 > 2.262$, we reject the null hypothesis in favour of the alternative hypothesis. We conclude that there is sufficient evidence to support the claim that mean discharge differs from 7 ounces.

(b) What is the conclusion if $\alpha = 0.10$?

If $\alpha = 0.10$,

$$t_{n-1,\,\alpha/2} = t_{9,\,0.05} = 1.833.$$

Since $2.635 > 1.833$, we would once again reject the null hypothesis.

(c) Find the p-value of the test statistic.

Since this is a two-tailed test, the p-value is two-times the area to the right of $|t|$ (since the t distribution is a symmetric distribution). In this case, since the value of our test statistic is already positive, we can drop the absolute values.

```r
statistic <- (7.1 - 7) / (0.12 / sqrt(10))

2 * pt(statistic, df=10-1, lower.tail=FALSE)
```

```
## [1] 0.02712501
```

That is to say,

$$p\text{-value} = 2 \times \mathbf{P}\left(T_9 \geq 2.635\right) = 0.027.$$

Since the p-value is less than both 0.05 and 0.10, we would have rejected the null hypothesis once again in both scenarios.

Question 2

(10.78) A manufacturer of hard safety hats for construction workers is concerned about the mean and the variation of the forces its helmets transmit to wearers when subjected to a standard external force. The manufacturer desires the mean force transmitted by helmets to be 800 pounds (or less), well under the legal 1000-pound limit, and desires σ to be less than 40. Tests were run on a random sample of $n = 40$ helmets, and the sample mean and variance were found to be equal to 825 pounds and 2350 pounds^2, respectively.

(a) If $\mu = 800$ and $\sigma = 40$, is it likely that any helmet subjected to the standard external force will transmit a force to a wearer in excess of 1000 pounds? Explain.

Let Y be the force transmitted by a helment. Assuming that Y follows a normal distribution, we can calculate $\mathbf{P}\left(Y > 1000\right)$ directly.

$$\mathbf{P}\left(Y > 1000\right) = \mathbf{P}\left(\frac{Y - 800}{40} > \frac{1000 - 800}{40}\right) = \mathbf{P}\left(Z > 5\right) \approx 0.$$

It is unlikely that a helmet will transmit a force to the wearer in excess of 1000 pounds.

(b) Do the data provide sufficient evidence to indicate that when subjected to the standard external force, the helmets transmit a mean force exceeding 800 pounds?

The hypotheses we wish to test are:

$$H_0 : \mu \leq 800 \quad \text{vs} \quad H_1 : \mu > 800$$

Our test statistic is

$$Z = \frac{\overline{X} - \mu_0}{S/\sqrt{n}} = \frac{\overline{X} - 800}{S/\sqrt{n}},$$

and will have an approximately standard normal distribution as a result of the central limit theorem. Substituting in the provided values (don't forget to take the square root of the provided variance!), the value of our test statistic is

$$z = \frac{825 - 800}{\sqrt{2350}/\sqrt{40}} = 3.262.$$

Rejection region method: As this is an upper-tailed test, we reject H_0 if $z > z_\alpha$.

$$z_\alpha = z_{0.05} = 1.645.$$

Since $3.262 > 1.645$, we reject the null hypothesis in favour of the alternative hypothesis. We conclude that there is sufficient evidence that when subjected to the standard external force, the helmets transmit a mean force exceeding 800 pounds.

p-value method: As this is an upper-tailed test, the p-value is found as the area to the right of the value of our test statistic, under the standard normal distribution.

$$p\text{-value} = \mathbf{P}\left(Z \geq 3.262\right) = 0.00055.$$

Since the p-value is less than $\alpha = 0.05$, we reject the null hypothesis in favour of the alternative hypothesis. We conclude that there is sufficient evidence that when subjected to the standard external force, the helmets transmit a mean force exceeding 800 pounds.

Do notice that the conclusion from both methods is the same!

(c) Do the data provide sufficient evidence to indicate that σ exceeds 40?

Note that if $\sigma = 40$, then $\sigma^2 = 1600$. Therefore, the hypotheses we wish to test are:

$$H_0 : \sigma^2 \leq 1600 \quad \text{vs} \quad H_1 : \sigma^2 > 1600$$

Our test statistic is

$$V = \frac{(n-1)S^2}{\sigma_0^2} \sim \chi^2_{n-1}.$$

Substituting in the provided values, the value of our test statistic is

$$v = \frac{(40-1)2350}{1600} = 57.281.$$

Rejection region method: As this is an upper-tailed test, we reject H_0 if $v > \chi^2_{n-1,\,\alpha}$.

$$\chi^2_{n-1,\,\alpha} = \chi^2_{39,\,0.05} = 54.572$$

Since $57.281 > 54.572$, we reject the null hypothesis in favour of the alternative hypothesis. There is sufficient evidence to conclude that σ^2 exceeds 1600, and as such, that σ exceeds 40.

p-value method: As this is an upper-tailed test, the p-value is found as the area to the right of the value of our test statistic, under the chi^2_{39} distribution.

$$p\text{-value} = \mathbf{P}\left(\chi^2_{39} \geq 57.281\right) = 0.0296.$$

Since the p-value is less than $\alpha = 0.05$, we reject the null hypothesis in favour of the alternative hypothesis. There is sufficient evidence to conclude that σ^2 exceeds 1600, and as such, that σ exceeds 40.

Once again, notice that the conclusion from both methods is the same!

Question 3

The `t4.rds` data set contains the sample means and sample standard deviations of 200 samples of size 60. Suppose we wish to test the following hypotheses 200 times:

$$H_0 : \mu \leq 3.1 \quad \text{vs} \quad H_1 : \mu > 3.1.$$

For each part, assume that $\alpha = 0.05$.

(a) Compute the values of the test statistics, their (approximate) p-values, and record whether the p-values result in a rejection of the null hypothesis.

We first read in the data.

```
t4 <- readRDS("t4.rds")

head(t4)
```

```
##      means      sds
## 1 4.266667 1.929843
## 2 3.816667 2.127238
## 3 4.033333 2.201438
## 4 3.800000 1.857646
## 5 3.900000 1.892895
## 6 3.466667 2.020796
```

As this is a large-sample hypothesis test, our test statistic is:

$$Z = \frac{\overline{X} - \mu_0}{S/\sqrt{n}} = \frac{\overline{X} - 3.1}{S/\sqrt{n}} \overset{\text{approx}}{\sim} N(0, 1)$$

As this is an upper-tailed test, the p-value is computed as the area to the right of the value of the computed statistic, under the standard normal distribution.

Recall that when the p-value is less than α, the null hypothesis is rejected in favour of the alternative hypothesis.

We will operate on the above data frame as we have done in the past tutorials by repeatedly piping (again, requires **R** 4.1+) and `transform()`ing.

```
alpha <- 0.05

t4 <- t4 |>
  transform(statistic = (means - 3.1) / (sds / sqrt(60))) |>
  transform(pval = pnorm(statistic, lower.tail=FALSE)) |>
  transform(reject = pval < alpha)

head(t4)
```

```
##      means      sds statistic         pval reject
## 1 4.266667 1.929843  4.682744 1.415297e-06   TRUE
## 2 3.816667 2.127238  2.609617 4.532183e-03   TRUE
## 3 4.033333 2.201438  3.284022 5.116853e-04   TRUE
## 4 3.800000 1.857646  2.918843 1.756664e-03   TRUE
## 5 3.900000 1.892895  3.273702 5.307427e-04   TRUE
```

```
## 6 3.466667 2.020796  1.405480 7.993932e-02  FALSE
```

(b) Compute lower confidence bounds for the true mean and record whether the lower confidence bounds contain the value of 3.1.

Recall that a large-sample lower confidence bound for the true mean, μ, is given by:

$$\overline{X} - z_\alpha \cdot \frac{S}{\sqrt{n}},$$

where α is the area to the right.

We will continue `transform()`ing our data set as before.

```
zval <- qnorm(alpha, lower.tail=FALSE)

t4 <- t4 |>
  transform(lower = means - zval * sds / sqrt(60)) |>
  transform(contained = 3.1 > lower)

head(t4)
```

```
##      means      sds statistic         pval reject    lower contained
## 1 4.266667 1.929843  4.682744 1.415297e-06   TRUE 3.856865     FALSE
## 2 3.816667 2.127238  2.609617 4.532183e-03   TRUE 3.364948     FALSE
## 3 4.033333 2.201438  3.284022 5.116853e-04   TRUE 3.565859     FALSE
## 4 3.800000 1.857646  2.918843 1.756664e-03   TRUE 3.405530     FALSE
## 5 3.900000 1.892895  3.273702 5.307427e-04   TRUE 3.498044     FALSE
## 6 3.466667 2.020796  1.405480 7.993932e-02  FALSE 3.037551      TRUE
```

(c) What is the proportion of hypothesis tests where you rejected the null hypothesis? What is the proportion of lower confidence bounds that contained the value of 3.1? What do you notice?

The proportion of hypothesis tests where I rejected the null hypothesis was:

```
mean(t4$reject)
```

```
## [1] 0.945
```

The proportion of lower confidence bounds that contained the value of 3.1 was:

```
mean(t4$contained)
```

```
## [1] 0.055
```

We should notice that the two proportions above sum to one. Exploring the `reject` and `contained` columns of our data set, we should also notice that whenever `reject` takes a value of `TRUE`, `contained` takes a value of `FALSE` (and vice-versa). This is actually not a coincidence! For a fixed level of α, there is a relationship between:

- A two-tailed hypothesis test and a two-sided confidence interval
- An upper-tailed test and a lower confidence bound
- A lower-tailed test and an upper confidence bound

In the context of our problem, our alternative hypothesis was of the form

$$\mu > 3.1,$$

while our lower confidence bound was of the form

$$\mu > \overline{X} - z_\alpha \cdot \frac{S}{\sqrt{n}}.$$

By the structure of a lower confidence bound, to find the null value of 3.1 not contained in the lower confidence bound means that the lower confidence bound contains only values greater than 3.1. By the interpretation of the lower confidence bound, all values contained within are *plausible* for the unknown parameter μ. As such, this suggests that only values greater than 3.1 are plausible values for μ, which supports the claim in the alternative hypothesis!

In general, if the null value is not contained in a one-sided confidence bound or a two-sided confidence interval, then the null hypothesis is rejected in favour of the corresponding alternative hypothesis.

(d) If the true value of μ was 4, in how many hypothesis tests did you commit an error? What type of error is this?

If the true value of μ was 4, then the null hypothesis was actually false. Thus, the error would arise in the tests where we failed to reject the null hypothesis despite it being false. This would be a type II error.

```r
# Number of tests where a type II error occurred
sum(t4$reject == FALSE)
```

```
## [1] 11
```

```r
# Proportion of tests where a type II error occurred
mean(t4$reject == FALSE)
```

```
## [1] 0.055
```

Do note that this proportion is the same as the proportion of lower confidence bounds that contained the null value (calculated in the previous part), which resulted in a failure to reject the null hypothesis.