Hypothesis Testing

Chapter 4, Lab 3: Solutions

OpenIntro Biostatistics

The previous labs in this unit discussed the calculation of point estimates and interval estimates for a population mean. An interval estimate produces a range of plausible values for μ based on the observed data.

Hypothesis testing operates under a different perspective—under the working hypothesis that the population mean is a particular value μ_0, what is the likelihood of observing a sample with mean $\overline{x}$? If there is a very small chance of observing a sample with mean $\overline{x}$ when $\mu = \mu_0$, this represents evidence against the working hypothesis and suggests that the μ does not equal μ_0.

The material in this lab corresponds to Section 4.3.1 - 4.3.2 of *OpenIntro Biostatistics*.

Introduction to hypothesis testing

This lab uses data from the National Health and Nutrition Examination Survey (NHANES), a survey conducted annually by the US Centers for Disease Control (CDC).[1] The complete `NHANES` dataset contains 10,000 observations, which will be the artificial target population.

Are American adults getting enough sleep?

A 2016 study from the CDC reported that more than a third of American adults are not getting enough sleep on a regular basis.[2] The National Sleep Foundation recommends that adults need between 7 to 9 hours of sleep per night to function well. Consistent sleep deprivation is known to increase risk of health problems, negatively affect cognitive processes, and contribute to depressive symptoms.

The dataset `nhanes.samp.adult` in the `oibiostat` package contains sleep information for 135 participants ages 21 years or older that were randomly sampled from the `NHANES` population. The variable `SleepHrsNight` contains the self-reported number of hours a participant usually gets at night on weekdays or workdays.

1. Explore the distribution of `SleepHrsNight` in `nhanes.samp.adult`.

   a) Using numerical and graphical summaries, describe the distribution of nightly sleep hours in `nhanes.samp.adult`.

   The data are roughly symmetric around the mean at 6.9 hours. There is one very low outlier corresponding to one person who reported getting 2 hours of sleep on weekdays. The middle half of participants report getting between 6 and 8 hours of sleep on weekdays.

```
#load the data
library(oibiostat)
data("nhanes.samp.adult")
```

[1] The dataset was first introduced in Chapter 1, Lab 1 (Introduction to Data).

[2] https://www.cdc.gov/media/releases/2016/p0215-enough-sleep.html

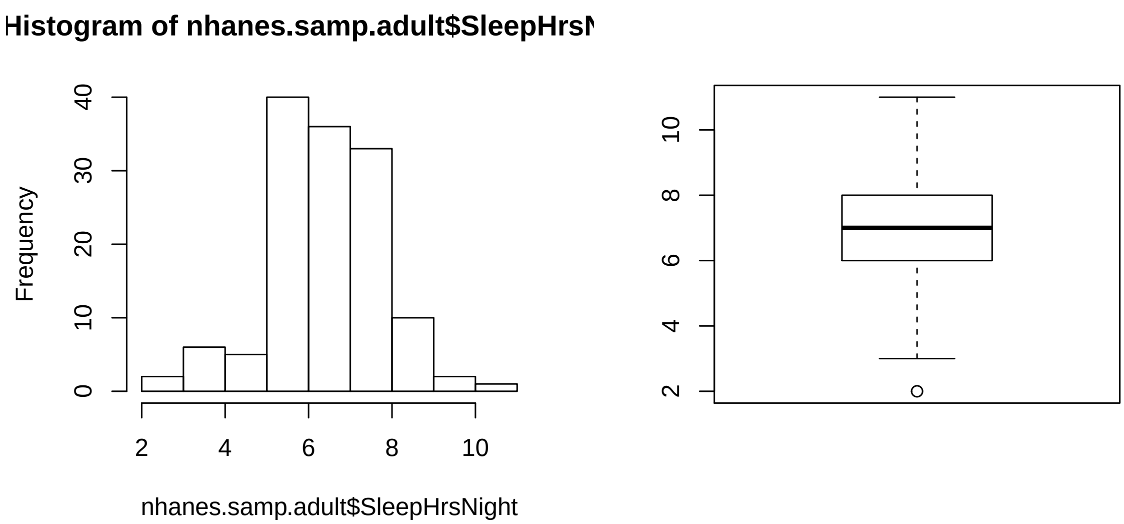
```r
#numerical summaries
summary(nhanes.samp.adult$SleepHrsNight)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   2.000   6.000   7.000   6.896   8.000  11.000
```

```r
sd(nhanes.samp.adult$SleepHrsNight)
```

```
## [1] 1.394414
```

```r
#graphical summaries
par(mfrow = c(1, 2))
hist(nhanes.samp.adult$SleepHrsNight)
boxplot(nhanes.samp.adult$SleepHrsNight)
```



b) Based on the distribution of nightly sleep hours in the sample, does it seem that the population mean nightly sleep hours may be outside the range defined as adequate (7 - 9 hours)?

Yes, it seems like the population mean nightly sleep hours will be outside the range defined as adequate; the sample mean at 6.9 hours is just below the lower end of this range.

2. Calculate a 95% confidence interval for nightly sleep hours using `nhanes.samp.adult` and interpret the interval.

We are 95% confident that the population average nightly sleep hours is between (6.66, 7.13) hours.

```r
#calculate a confidence interval
t.test(nhanes.samp.adult$SleepHrsNight, na.rm = TRUE, conf.level = 0.95)$conf.int
```

```
## [1] 6.658933 7.133659
## attr(,"conf.level")
## [1] 0.95
```

3. Conduct a hypothesis test to assess whether on average, American adults are getting enough sleep. Let μ_0 be 8 hours, the midpoint of the range defined as adequate.

   a) Formulate null and alternative hypotheses. The symbol μ denotes the population mean, while μ_0 refers to the numeric value specified by the null hypothesis.

   The null hypothesis is that the mean nightly sleep hours in the population equals 8 hours ($H_0 : \mu = 8$). Based on prior knowledge that Americans tend to be sleep-deprived, it is reasonable to assume that if H_0 is not true, then the population mean sleep hours would be less than 8 hours ($H_A : \mu < 8$).

b) Specify a significance level, α.

   The significance level α defines a "rare" event. Typically, $\alpha = 0.05$. An α level of 0.05 means that an event occurring with probability lower than 5% will be considered sufficient evidence against the null hypothesis. For this test, let $\alpha = 0.05$.

c) Calculate the test statistic.

$$t = \frac{\overline{x} - \mu_0}{s/\sqrt{n}}$$

   The test statistic is

$$t = \frac{\overline{x} - \mu_0}{s/\sqrt{n}} = \frac{6.89 - 8}{1.39/\sqrt{135}} = -9.20$$

```
#use r as a calculator
x.bar = mean(nhanes.samp.adult$SleepHrsNight)
mu.0 = 8
s = sd(nhanes.samp.adult$SleepHrsNight)
n = length(nhanes.samp.adult$SleepHrsNight)

t = (x.bar - mu.0)/(s/sqrt(n))
t
```

```
## [1] -9.19661
```

d) Calculate the p-value.

   The p-value is extremely small; $P(T \leq -9.20) = 3.16 \times 10^{-16}$.

```
pt(t, df = n - 1, lower.tail = TRUE)
```

```
## [1] 3.164008e-16
```

```
#alternatively, use t.test to calculate the t-stat and p-value
t.test(nhanes.samp.adult$SleepHrsNight, mu = 8, alternative = "less")
```

```
##
##  One Sample t-test
##
## data:  nhanes.samp.adult$SleepHrsNight
## t = -9.1966, df = 134, p-value = 3.164e-16
## alternative hypothesis: true mean is less than 8
```

```
## 95 percent confidence interval:
##      -Inf 7.095073
## sample estimates:
## mean of x
##  6.896296
```

e) Draw a conclusion.

The p-value is less than α; the results are significant at $\alpha = 0.05$. There is sufficient evidence to reject the null hypothesis and accept the alternative that mean nightly sleep hours in the population is less than 8 hours. The chance of drawing a sample with mean nightly sleep hours 6.90 hours if the distribution were actually centered at 8 hours is almost zero! These data support the conclusion that American adults are not getting enough sleep.

Hypothesis Testing: Cholesterol Level

High cholesterol is a major controllable risk factor for coronary heart disease, heart attack, and stroke. According to general guidelines, a total cholesterol level below 5.2 mmol/L is desirable, in the 5.2 - 6.2 mmol/L range is borderline high, and above 6.2 mmol/L is high.

4. Describe the distribution of total cholesterol in `nhanes.samp.adult`. Does it seem that most individuals have a cholesterol level considered desirable?

The distribution shows some right skew, with a few individuals having very high total cholesterol levels. The median is at 5.08 mmol/L; thus, at least 50% of individuals in the sample have a cholesterol level considered desirable (i.e., below 5.2 mmol/L).

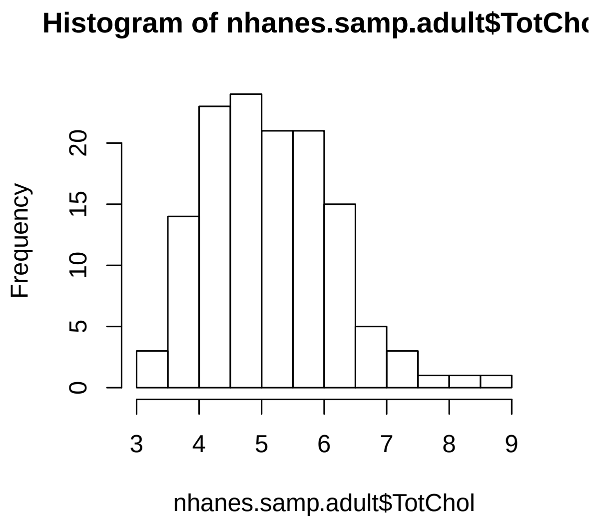
```r
#numerical summaries
summary(nhanes.samp.adult$TotChol)
```

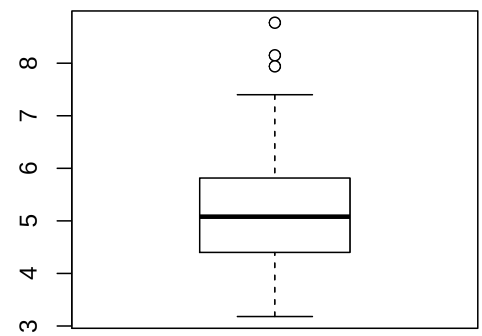
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.    NA's 
##   3.180   4.400   5.080   5.163   5.803   8.770       3
```

```r
sd(nhanes.samp.adult$TotChol, na.rm = TRUE)
```

```
## [1] 1.045627
```

```r
#graphical summaries
par(mfrow = c(1, 2))
hist(nhanes.samp.adult$TotChol)
boxplot(nhanes.samp.adult$TotChol)
```





5. Conduct a hypothesis test to assess whether mean total cholesterol in the `NHANES` "population" is equal to 5.2 mmol/L, using the data in `nhanes.samp.adult`.

   a) Choose whether to conduct a one-sided or two-sided test. Formulate null and alternative hypotheses.

Conduct a two-sided test to assess whether there is a difference in either direction; $H_0: \mu = 5.2$ mmol/L versus $H_A: \mu \neq 5.2$ mmol/L.

   b) Specify a significance level, α.

Let $\alpha = 0.05$.

   c) Calculate the test statistic. (Note: Be careful of missing data values.)

The test statistic is -0.411.

```
#use r as a calculator
x.bar = mean(nhanes.samp.adult$TotChol, na.rm = TRUE)
mu.0 = 5.2
s = sd(nhanes.samp.adult$TotChol, na.rm = TRUE)
n = length(nhanes.samp.adult$TotChol) - sum(is.na(nhanes.samp.adult$TotChol))

t = (x.bar - mu.0)/(s/sqrt(n))
t
```

```
## [1] -0.4112094
```

   d) Calculate the p-value.

For a two-sided hypothesis test, the p-value is the total area from both tails of the t distribution that are beyond the absolute value of the observed t-statistic. Here, $2 \times P(T \leq -0.411) = 0.682$.

```
2*pt(t, df = n - 1, lower.tail = TRUE)
```

```
## [1] 0.681591
```

e) Confirm your calculations in parts c) and d) using `t.test()`.

```
t.test(x = nhanes.samp.adult$TotChol, mu = 5.2, alternative = "two.sided")
```

```
##
##  One Sample t-test
##
## data:  nhanes.samp.adult$TotChol
## t = -0.41121, df = 131, p-value = 0.6816
## alternative hypothesis: true mean is not equal to 5.2
## 95 percent confidence interval:
##  4.982536 5.342616
## sample estimates:
## mean of x
##  5.162576
```

f) Draw a conclusion.

The p-value is greater than α; there is insufficient evidence to reject the null hypothesis in favor of the alternative. Based on these data, the population mean total cholesterol level may not be different from 5.2 mmol/L. Note that it would be inappropriate to conclude that the population mean total cholesterol level *is* 5.2 mmol/L; failure to reject the null does not imply that the null is true.

Hypothesis Testing: Body Temperature

Mean body temperature is commonly accepted to be 98.6 degrees Fahrenheit. The origin of this benchmark value is credited to Carl Wunderlich, a scientist working in clinical thermometry in the late 19th century. A study was conducted in 1992 to evaluate whether population mean body temperature among healthy adults is really 98.6 F. Data were collected from healthy volunteers who had agreed to participate in a separate set of vaccine trials; these data are in the dataset `thermometry` in the `oibiostat` package.

6. Conduct a formal hypothesis test to evaluate if mean body temperature is really 98.6 F. Be sure to clearly specify the hypotheses and report the conclusions.

It is reasonable to conduct a two-sided test since there is no prior knowledge as to the nature of the difference. Test $H_0 : \mu = 98.6$ F versus $H_A : \mu \neq 98.6$ F. Let $\alpha = 0.05$. The t-statistic is -5.45, which results in a p-value of 2.41×10^{-7}. Since $p < \alpha$, there is sufficient evidence to reject the null hypothesis and conclude that population mean body temperature is not equal to 98.6 F. Since the sample mean (98.2 F) is lower than 98.6 F, the data suggests that population mean body temperature is lower than 98.6 F.

```
#load the data
data("thermometry")

#calculate the test statistic and p-value
t.test(thermometry$body.temp, mu = 98.6, alternative = "two.sided")
```

```
##
##  One Sample t-test
##
## data:  thermometry$body.temp
## t = -5.4548, df = 129, p-value = 2.411e-07
## alternative hypothesis: true mean is not equal to 98.6
## 95 percent confidence interval:
##  98.12200 98.37646
## sample estimates:
## mean of x
##  98.24923
```

7. Calculate a 99% confidence interval for mean body temperature in healthy adults. Does the interval contain 98.6 F?

   The 99% confidence interval (98.08, 98.42) F does not contain 98.6 F.

```r
#calculate a confidence interval
t.test(thermometry$body.temp, mu = 98.6, conf.level = 0.99)$conf.int
```

```
## [1] 98.08111 98.41735
## attr(,"conf.level")
## [1] 0.99
```

8. Briefly summarize the conclusions of this analysis in language accessible to a general audience. What is a possible reason for the observed discrepancy between the 1992 data and Wunderlich's data?

   Based on this analysis, there is sufficient evidence to conclude that mean body temperature is actually lower than 98.6 F. In a group of 130 healthy adults, mean body temperature was about 98.2 F; it would be very unlikely to observe a mean of 98.2 F in a randomly drawn group of individuals if in the population, mean body temperature is actually 98.6 F.

   One possible reason for the discrepancy between the 1992 data and Wunderlich's data is that today's thermometers are more reliable. In the paper reporting the results of this study, the authors discussed that thermometers of Wunderlich's time required 15-20 minutes to equilibrate, and that Wunderlich also preferred measuring body temperature via the underarm. In modern medicine, the preferred sites for measuring body temperature are the mouth or rectum.