Hypothesis Testing

Chapter 4, Lab 3

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

[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

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.

b) Specify a significance level, α.

c) Calculate the test statistic.

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

d) Calculate the p-value.

e) Draw a conclusion.

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?
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.
   b) Specify a significance level, α.
   c) Calculate the test statistic. (Note: Be careful of missing data values.)
   d) Calculate the p-value.
   e) Confirm your calculations in parts c) and d) using `t.test()`.
   f) Draw a conclusion.

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

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?