

# Inference Concept Check

Chapter 4, Lab 4

*OpenIntro Biostatistics*

## Topics

- Relationship between hypothesis tests and confidence intervals
- One-sided hypothesis tests and confidence intervals
- Choosing between one-sided and two-sided tests
- Definition of  $\alpha$  as the acceptable error probability

The previous labs in this unit discussed the mechanics of calculating confidence intervals and conducting hypothesis tests. This lab focuses on examining some conceptual details of inference.

The material in this lab corresponds to Sections 4.3.3 - 4.3.5 of *OpenIntro Biostatistics*.

## Hypothesis Testing and Confidence Intervals

Suppose we would like to assess whether the mean weight in the yrbss “population” is different from 60 kilograms, using the information from a random sample. The YRBSS data was introduced in the first lab in this unit (Chapter 4, Lab 1).

Run the following code chunk to take a random sample of size 100 from yrbss.complete, the version of yrbss that has no missing weight values.

```
#load the dataset
library(oibiostat)
data("yrbss")

#remove rows with missing values
yrbss.complete = yrbss[complete.cases(yrbss$weight), ]

#set parameters
sample.size = 100

#set seed for pseudo-random sampling
set.seed(5011)

#obtain random sample of row numbers
sample.rows = sample(1:nrow(yrbss.complete), sample.size)

#create yrbss.sample
yrbss.sample = yrbss.complete[sample.rows, ]
```

1. Calculate a 95% confidence interval for  $\mu_{weight}$  using the data in yrbss.sample. Does it seem like the population mean weight will be different from 60 kilograms?

2. Test  $H_0 : \mu_{weight} = 60$  kg versus  $H_A : \mu_{weight} \neq 60$  kg at the  $\alpha = 0.05$  significance level.
  - a) What point on the  $t$ -distribution (with  $df = 99$ ) has area of 0.025 to the left? What point on the  $t$ -distribution (with  $df = 99$ ) has area of 0.025 to the right?  
 These points mark off the *rejection region* as defined by  $\alpha$ ; the  $t$ -statistic calculated from  $\bar{x}$  must lie within the tail areas bounded by these points in order to constitute sufficient evidence against  $H_0$ .
  - b) Calculate the  $t$ -statistic from `yrbss.sample`. Confirm that it lies within the rejection region and that the associated  $p$ -value is less than  $\alpha$ .
3. The relationship between a hypothesis test and the corresponding confidence interval is defined by  $\alpha$ . Suppose that a two-sided test is conducted at significance level  $\alpha$ ; the confidence level of the matching interval is  $(1 - \alpha)\%$ . For example, a 95% confidence interval can be compared to a two-sided hypothesis test with  $\alpha = 0.05$ .
  - a) What is the margin of error,  $m$ , from the confidence interval calculated in Question 1?
  - b) What values of  $\bar{x}$  would correspond to the  $t$ -points that mark off the rejection region? In other words, what weight values for the sample mean would be considered ‘extreme’ enough to warrant rejecting  $H_0$ ?
  - c) How far apart are the two values calculated in part b) from each other? How does the distance relate to  $m$ , as calculated in part a)?

### One-sided hypothesis tests and confidence intervals

One-sided confidence intervals for a population mean provide either a lower bound or upper bound, but not both. One-sided confidence intervals have the form:

$$(\bar{x} - m, \infty) \text{ or } (-\infty, \bar{x} + m), \text{ where } m = \frac{s}{\sqrt{n}} t^*.$$

In general, for a confidence interval of  $(1 - \alpha)100\%$ ,  $t^*$  is the point on a  $t$  distribution with  $n - 1$  degrees of freedom that has area  $1 - \alpha$  to the left. For a 95% confidence interval,  $\alpha = 0.05$ ;  $t^*$  is the point on a  $t$  distribution with area  $1 - 0.05 = 0.95$  to the left.

A one-sided hypothesis test with  $\alpha = 0.05$  and  $H_A : \mu > \mu_0$  corresponds to a one-sided 95% confidence interval that has a lower bound, but no upper bound:  $(\bar{x} - m, \infty)$ . If instead,  $H_A : \mu < \mu_0$ , the test corresponds to a one-sided 95% interval with no lower bound:  $(-\infty, \bar{x} + m)$ .

4. Suppose that in a random sample of 150 adults from the US population, the average amount of nightly sleep is 6.80 hours, with standard deviation 1.60 hours.
  - a) Is there evidence that on average, American adults sleep less than 7 hours per night? Use  $\alpha = 0.10$ .
  - b) Calculate a one-sided lower 90% confidence interval for the average amount of sleep per night among US adults; i.e., of the form  $(-\infty, \bar{x} + m)$ . Compare the information obtained from a confidence interval versus a hypothesis test and assess whether the results of the test are practically significant, as opposed to statistically significant.

- c) Suppose we were instead interesting in testing whether American adults sleep more than 7 hours per night on average. Calculate the  $p$ -value for the test and the corresponding 90% confidence interval.

### Choosing between one-sided and two-sided tests

5. A standard test for diabetes involves measuring blood sugar levels after an overnight fast. Someone without diabetes typically has a fasting blood sugar level of around 5.31 mmol/L. High fasting blood sugar levels are indicative of diabetes or prediabetes.

Neighborhood and community-level factors are known to influence diabetes risk. People living in poorer neighborhoods tend to be at higher risk of diabetes; these neighborhoods often lack grocery stores, recreational facilities, and green space.

- a) For the following scenarios, choose whether to conduct a one-sided or two-sided test. Formulate null and alternative hypotheses.
- Suppose that you are interested in learning whether mean fasting blood sugar levels in Neighborhood A, a middle-class neighborhood, is different from 5.31 mmol/L.
  - Neighborhood B is primarily inhabited by low-income families. If the mean fasting blood sugar level in Neighborhood B is higher than normal, you will start recruiting participants for an antidiabetic drug trial from the neighborhood.
- b) The dataset `sugar.levels.A` contains simulated fasting blood sugar levels for 100 individuals from Neighborhood A; `sugar.levels.B` contain the simulated data from Neighborhood B. Both datasets are in the `oibiostat` package.
- Use `t.test()` to test the hypotheses from part a). Use  $\alpha = 0.05$ . Summarize your conclusions.
  - The values in the two datasets are actually identical. Why are the  $p$ -values from the two tests different from each other?

### The definition of $\alpha$ as the acceptable error probability

The significance level  $\alpha$  can be thought of as the value that quantifies how rare or unlikely an event must be in order to represent sufficient evidence against the null hypothesis. For example, an  $\alpha$  level of 0.05 means that an event occurring with probability lower than 5% will be considered sufficient evidence against  $H_0$ .

It is also possible to think about  $\alpha$  in the context of decision errors. Hypothesis tests can potentially result in incorrect decisions. Rejecting the null hypothesis when the null is true is referred to as a **Type I** error. A Type I error occurs with probability  $\alpha$ , since  $\alpha$  determines the cutoff point for rejecting the null hypothesis. If  $\alpha = 0.05$ , there is a 5% chance of incorrectly rejecting  $H_0$ .

To explore this idea, let's return to the `yrbss` repeated sampling simulation.

Recall that `yrbss.complete` is our artificial "population", where the mean weight,  $\mu_{weight}$ , is 67.91 kg. Let 67.91 kg be  $\mu_0$  and test  $H_0 : \mu = 67.91$  kg versus  $H_A : \mu \neq 67.91$  kg. A simulation can be run in which a hypothesis test is performed on each sample drawn from the population; since in this

setting, it is known that  $\mu$  truly is 67.91 kg, each instance of rejecting  $H_0$  represents an instance of making a Type I error.

6. Run the code chunk shown in the template to take 1,000 random samples of size 100 from `yrbss.complete`. The code calculates the  $t$ -statistics from each sample and returns TRUE if the  $t$ -statistic from a sample is within the rejection region.

```
#set parameters
sample.size = 100
replicates = 1000
alpha = 0.05

#set seed
set.seed(2017)

#create empty lists
t.stat = vector("numeric", replicates)

#calculate t-statistics from each sample
for (k in 1:replicates) {
  sample.rows = sample(nrow(yrbss.complete), sample.size)

  #define constants
  mu.0 = mean(yrbss.complete$weight)
  sample.mean = mean(yrbss.complete$weight[sample.rows])
  sample.sd = sd(yrbss.complete$weight[sample.rows])

  #calculate t statistics
  t.stat[k] = (sample.mean - mu.0) / (sample.sd / sqrt(sample.size))
}

#define upper and lower bounds of rejection region
reject.ub = qt(1 - (alpha)/2, df = sample.size - 1)
reject.lb = qt(alpha/2, df = sample.size - 1)

#is the t-stat in the rejection region?
in.rejection.region = (t.stat >= reject.ub) | (t.stat <= reject.lb)

table(in.rejection.region)
```

- a) With  $\alpha = 0.05$ , what percentage of samples result in the (incorrect) conclusion that the population mean is not 67.91 kg?
- b) What happens when  $\alpha$  is changed? Test  $\alpha = 0.10$  and  $\alpha = 0.01$ . Compare your results to those in part a).
- c) Why not eliminate the chance of Type I error completely by letting  $\alpha = 0$ ? What implications would this have for hypothesis testing?