Lab 8: Two-sample inference

STAT218

This lab focuses on two-sample inference for differences in population means. The main objectives are:

1. Learn to implement two-sample tests in R
2. Practice distinguishing directional and nondirectional tests and providing appropriate specifications to the t.test(...) function in R

The lab uses several datasets for which we will consider two-sample comparisons:

* finch: mean finch beak depths in generations before and after a drought on Daphne Major
* temps: body temperatures and heart rates for men and women

library(tidyverse)  
load('data/finch.RData')  
load('data/temps2.RData')

Examples will utilize the finch data; you’ll practice using the temps data. Here are the summary statistics for the finch data for reference:

finch |>  
 group\_by(year) |>  
 summarize(depth.mean = mean(depth),  
 depth.sd = sd(depth),  
 n = n())

# A tibble: 2 × 4  
 year depth.mean depth.sd n  
 <int> <dbl> <dbl> <int>  
1 1976 9.45 0.962 58  
2 1978 10.2 0.807 65

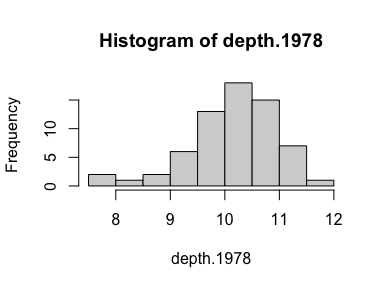
### Checking assumptions for two-sample tests

A two-sample test can be used whenever two one-sample tests are appropriate. So, to check assumptions, we need to inspect the frequency distributions of the variable of interest in *both* samples.

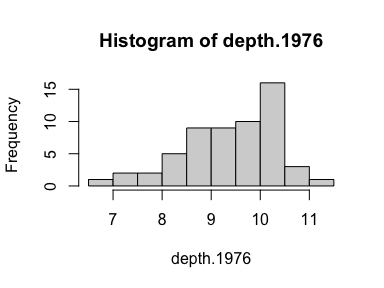
#### Option A: two histograms

One way to do this is to make two separate histograms. To do that, we’ll need to separate the samples. This can be done by ‘filtering’ observations according to whether year is 1978 or 1976.

# separate samples  
finch.1978 <- finch |> filter(year == 1978)  
finch.1976 <- finch |> filter(year == 1976)  
  
# extract depths  
depth.1978 <- finch.1978$depth  
depth.1976 <- finch.1976$depth  
  
# make histograms  
hist(depth.1978)



hist(depth.1976)



Both distributions are a bit left-skewed, but each sample is large enough that this isn’t a problem for performing the test.

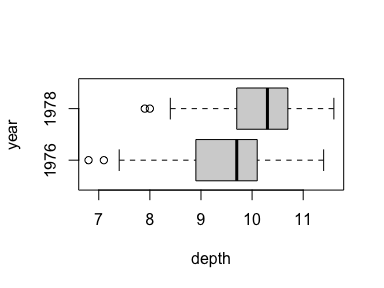
|  |
| --- |
| Your turn 1 |
| Filter the temps data by sex to separate the samples, and make histograms of the heart rates. Comment on whether assumptions seem to be met.  # separate samples  # extract heart rates  # make histograms |

#### Option B: side-by-side boxplots

A slightly more efficient alternative is to make side-by-side boxplots. This doesn’t involve filtering the data, and will produce just a single graphic.

However, some details of the distribution (such as multiple modes) may not be evident from the boxplots, so it’s not a perfect substitute for checking histograms.

# side-by-side boxplots  
boxplot(depth ~ year, data = finch, horizontal = T)



Here we want to see two things:

* approximate symmetry of boxes
* few to no large outliers

While there is a bit of left skewness, the sample sizes are large enough that it’s not a concern.

|  |
| --- |
| Your turn 2 |
| Make side-by-side boxplots for heart rate and reassess test assumptions.  # side-by-side boxplots for heart rate |

### Two-sample -tests

Given that assumptions seem plausible (both samples show little skew and few outliers, and are sufficiently large), we can go ahead with the test. To test whether the drought imposed selection pressure on the finch population, we want to know whether finch beak depth increased after the drought.

Observe, first, which sample appears first in the dataset: 1976. R will treat this as the first sample; to keep track of directions, we’ll want to formulate the hypotheses as a comparison between 1976 (first sample) and 1978 (second sample).

We want to test whether the mean of the first sample is less than the mean of the second:

Let’s carry out the test at the 5% significance level. The inputs to t.test(...) that implement this test are:

* a formula <VARIABLE> ~ <SAMPLE> as the first argument: formula = depth ~ year
* a data frame containing the variable names mentioned in the formula: data = finch
* a null value for the difference: mu = 0
* an alternative: alternative = 'less'
* a confidence level to complement the significance level of the test: conf.level = 0.95

# perform t test (notice which group comes first)  
t.test(formula = depth ~ year, data = finch, mu = 0, alternative = 'less', conf.level = 0.95)

Welch Two Sample t-test  
  
data: depth by year  
t = -4.5727, df = 111.79, p-value = 6.255e-06  
alternative hypothesis: true difference in means between group 1976 and group 1978 is less than 0  
95 percent confidence interval:  
 -Inf -0.4698812  
sample estimates:  
mean in group 1976 mean in group 1978   
 9.453448 10.190769

Take a moment to inspect the output and identify each number appearing. We’d report the test result as follows:

The data provide very strong evidence that mean beak depth increased in the generation of finches following the drought (*T* = -4.5727 on 111.79 degrees of freedom, *p* < 0.0001). With 95% confidence, the mean beak depth is estimated to have increased by at least 0.4699 mm, with a point estiamte of 0.7373 mm (SE 0.1612).

The point estimate and standard error can be retrieved by storing the output of t.test(...).

# store t test result  
tt.rslt <- t.test(formula = depth ~ year, data = finch, mu = 0, alternative = 'less', conf.level = 0.95)  
  
# estimates  
tt.rslt$estimate

mean in group 1976 mean in group 1978   
 9.453448 10.190769

# estimate for difference in means  
tt.rslt$estimate |> diff()

mean in group 1978   
 0.737321

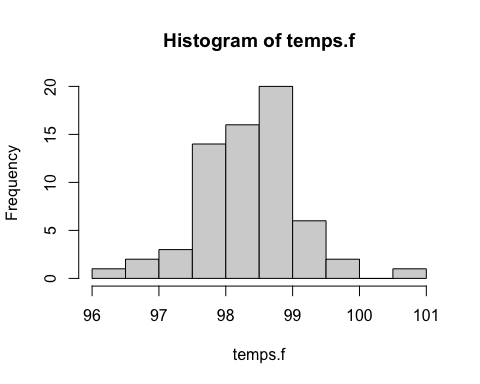
# standard error for estimate of difference in means  
tt.rslt$stderr

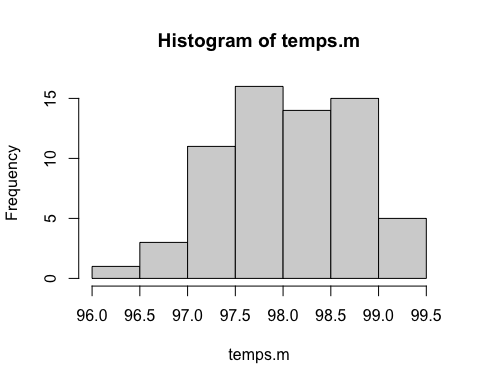
[1] 0.1612445

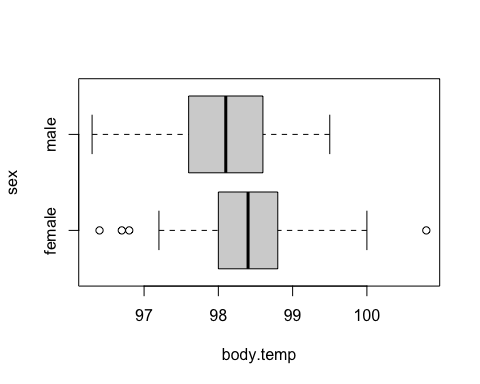
|  |
| --- |
| Your turn 3 |
| Test whether mean heart rate differs between men and women at the 1% significance level. Report the test result, confidence interval, and point estimate and standard error for the difference in means.  # perform t test  # store t test result  # estimate for difference in means  # standard error |

### Practice problems

1. Using the temps2 dataset, test whether mean body temperature is lower for men.
   1. Check the assumptions for the test by making *both* a pair of histograms *and* side-by-side boxplots.
   2. Perform the test at the 1% significance level.
   3. Report the test result, confidence interval, and point estimate and standard error for the difference in means.







Welch Two Sample t-test  
  
data: body.temp by sex  
t = 2.2854, df = 127.51, p-value = 0.01197  
alternative hypothesis: true difference in means between group female and group male is greater than 0  
99 percent confidence interval:  
 -0.008923783 Inf  
sample estimates:  
mean in group female mean in group male   
 98.39385 98.10462

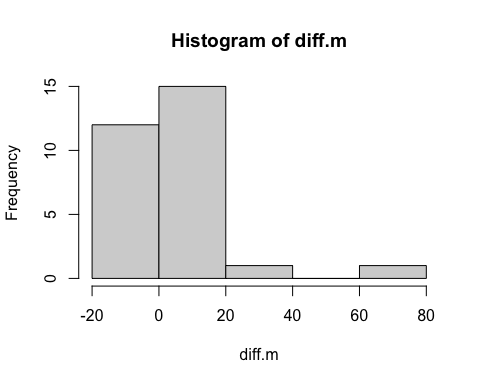
mean in group female mean in group male   
 98.39385 98.10462

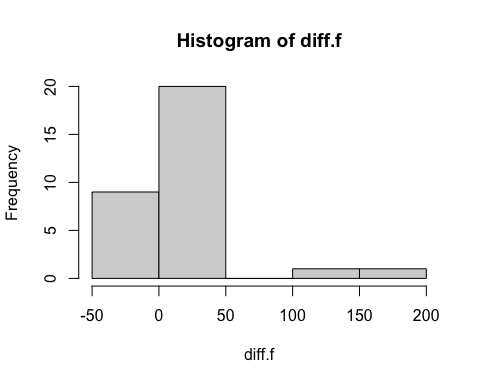
mean in group male   
 -0.2892308

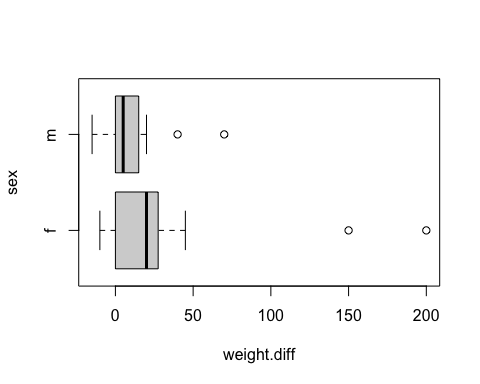
[1] 0.126554

1. Using the brfss2 data, test whether actual body weight exceeds desired body weight by more for women than for men.
   1. Check the assumptions for the test by making *both* a pair of histograms *and* side-by-side boxplots.
   2. Perform the test at the 1% significance level.
   3. Report the test result, confidence interval, and point estimate and standard error for the difference in means.

# A tibble: 6 × 4  
 sex weight wtdesire weight.diff  
 <chr> <dbl> <dbl> <dbl>  
1 m 265 225 40  
2 m 150 150 0  
3 m 137 150 -13  
4 f 159 125 34  
5 f 145 125 20  
6 f 125 120 5







Welch Two Sample t-test  
  
data: weight.diff by sex  
t = 2.048, df = 38.875, p-value = 0.02368  
alternative hypothesis: true difference in means between group f and group m is greater than 0  
99 percent confidence interval:  
 -3.108832 Inf  
sample estimates:  
mean in group f mean in group m   
 26.354839 9.517241

mean in group m   
 -16.8376

[1] 8.22135