Chapter 18: Comparing two population means

Corinne Riddell

October 26, 2020

### Recap

* So far we’ve learned the z-test and the t-test that apply when the variable of interest is continuous
* These are **one sample** tests, meaning we have one variable of interest, and we take one sample. We’re interested in knowing whether the one sample differs from some null hypothesized value ()

### Next up: comparing two population means

* In Chapter 18, we are interested in comparing two samples from two populations. We want to shed light on the question: Do these two samples have the same underlying mean, or are their means different? Another way of phrasing this is, “is the difference between the means equal to 0 or not?”
* For this question, our null and two-sided alternative hypotheses are:

- You can also write them this way, by rearranging the equation:

### Comparing two population means

In public health, we are often most interested in questions that compare two groups (exposed vs. unexposed, treatment vs. placebo, etc)

* Run a randomized controlled trial where we compare a treated subgroup to a placebo group. Do their mean health outcomes differ?
* Conduct an observational study where we have exposed and unexposed individuals. Do their mean health outcomes differ?

In this chapter, we talk about tests to compare the difference in two continuous means, represented by and

### Comparing two samples, graphically

**Procedure to compare two samples graphically**

* Make two histograms, one for each sample
* Compare their shapes, centers (means or medians) and spreads (standard deviations)
* Or, make two box plots and compare their medians and IQRs

### Conditions for inference comparing two means

* We have two SRSs, representing two distinct populations.
  + The samples are **independent**. That is, the individuals in one sample are unrelated to the individuals in the other sample.
  + We measure the same quantitative variable for both samples.
* Both populations are **Normally distributed**. The means and standard deviations of the populations are unknown. **In practice, it is enough that the distributions have similar shapes and that the data have no strong outliers.**

### Notation

Notation for the population parameters:

|  |  |  |  |
| --- | --- | --- | --- |
| Population | Variable | Population mean | Population SD |
| 1 |  |  |  |
| 2 |  |  |  |

### Notation

Notation for the sample statistics:

|  |  |  |  |
| --- | --- | --- | --- |
| Population | Sample size | Sample mean | Sample SD |
| 1 |  |  |  |
| 2 |  |  |  |

To perform inference about the difference between the means (), we first estimate the difference between the sample means.

### Two-sample test

* With one-sample tests, we have one and we sketch the sampling distribution for . It is centered at with standard error
* With two-samples, we have two sample averages and .
* What do we do?

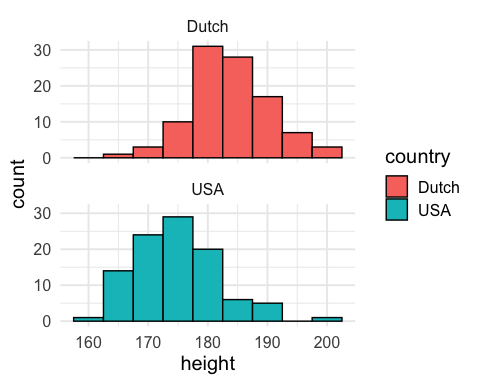
### Example with adolescent heights from two different countries

Are the heights of US and Dutch-born men different?

## # A tibble: 2 x 4  
## country sample\_mean sample\_sd length  
## <chr> <dbl> <dbl> <int>  
## 1 Dutch 184. 6.75 100  
## 2 USA 175. 7.16 100

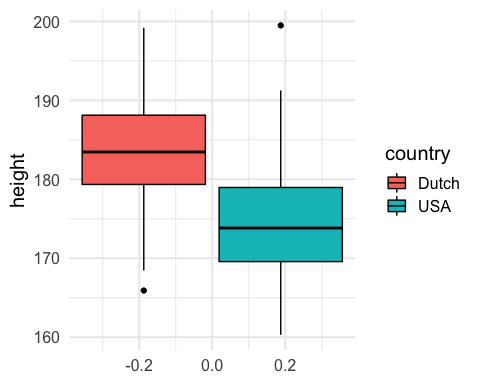
### Compare the histograms

ggplot(height\_data, aes(x = height)) +   
 geom\_histogram(aes(fill = country), binwidth = 5, col = "black") +  
 theme\_minimal(base\_size = 15) +  
 facet\_wrap(~country, nrow = 2)



### Compare the boxplots

ggplot(height\_data, aes(y = height)) +   
 geom\_boxplot(aes(fill = country), col = "black") +  
 theme\_minimal(base\_size = 15)



### Example with adolescent heights from two different populations

Are the heights of US and Dutch-born men different?

and so the difference is

What would happen if we took another set of two samples of USA and Dutch-born men? We would expect these sample means to change. We could draw an approximate **sampling distribution for the difference** between these two means.

### The sampling distribution of

The distribution of the difference between two independent random variables has a mean equal to the difference of their respective means and a variance equal to the sum of their respective variances. That is:

* The mean of the sampling distribution for is .
* The variance of the sampling distribution is:
* The standard deviation of the sampling distribution is:
* Our *estimate* of the standard deviation of the sampling distribution is:

### Recall the t-test

We need to generalize this by replacing each piece in the z-test by the calculations on the previous slide:

The **two-sample** t-test is therefore:

The two-sample t statistics has approximately a distribution. The approximation is accurate when both sample sizes are greater than or equal to 5.

### Degrees of freedom for the two-sample t-test…

is bananas.

Where:

* is the sample variance for sample #1
* is the sample variance for sample #2
* is the sample size for sample #1
* is the sample size for sample #2

### Confidence intervals for the two-sample t-test

,

where is the critical value with area C between and under the density curve with the appropriate degrees of freedom.

### Hypothesis testing when you have two samples

Here we consider the two-sided alternative hypothesis, but you could be interested in the one-sided alternative in either direction:

Obtain the **two-sample t-test** statistic

Under the two-sided alternative:

where the **p-value** is the probability, assuming is true, of getting a test statistic that we saw or a more extreme value. We compute the p-value as the corresponding area under the distribution with the appropriate degrees of freedom.

### Example, continued

Let R do the work for you:

t.test(height\_data\_wide %>% pull(usa),   
 height\_data\_wide %>% pull(dutch),   
 alternative = "two.sided")

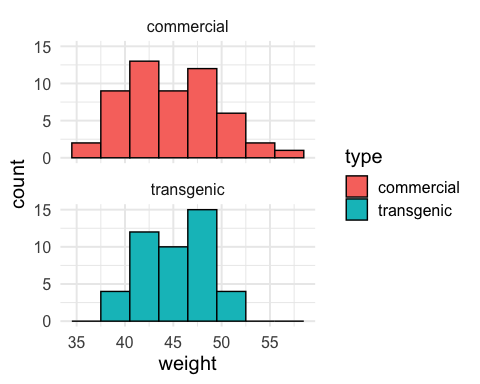
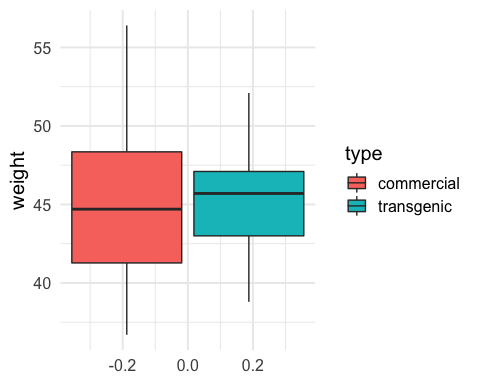
##   
## Welch Two Sample t-test  
##   
## data: height\_data\_wide %>% pull(usa) and height\_data\_wide %>% pull(dutch)  
## t = -9.1103, df = 197.35, p-value < 2.2e-16  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -10.905418 -7.024282  
## sample estimates:  
## mean of x mean of y   
## 174.7042 183.6690

Note that t.test gives you both the t-test results (including the t-statistic (called “t” in the output), df, and p-value), as well as the 95% CI. We got both because we performed a two-sided test.

### Example 2: Transgenic chickens

Infection of chickens with the avian flu is a threat to both poultry production and human health. A research team created transgenic chickens resistant to avian flu infection. Could the modification affect the chicken in other ways? The researchers compared the hatching weights (in grams) of 45 transgenic chickens and 54 independently selected commercial chickens of the same breed.

transgenic <- c(38.8, 39.0, 39.7, 40.0, 40.8, 40.9, 41.0, 41.0, 41.0, 42.5, 42.6, 43.0,  
 43.0, 43.4, 43.5, 43.5, 43.8, 44.4, 44.7, 44.7, 44.7, 45.3, 45.7, 45.8,   
 46.4, 46.5, 46.6, 46.7, 46.7, 46.8, 46.9, 47.1, 47.1, 47.1, 47.3, 47.6,  
 47.7, 48.1, 48.3, 49.3, 49.3, 49.8, 50.3, 50.9, 52.1)  
  
commercial <- c(36.7, 37.1, 38.9, 39.5, 39.5, 39.8, 40.0, 40.2, 40.3, 40.5, 40.5, 40.7,  
 41.1, 41.2, 41.5, 41.5, 41.6, 41.6, 41.7, 42.4, 43.1, 43.3, 43.3, 43.4,  
 43.7, 44.1, 44.2, 45.2, 45.3, 45.4, 46.0, 46.1, 46.4, 46.6, 46.6, 46.9,   
 47.3, 47.5, 48.1, 48.2, 48.4, 48.6, 49.0, 49.1, 49.3, 49.6, 50.1, 50.2,   
 50.4, 50.6, 52.2, 53.0, 55.5, 56.4)  
  
chicken\_data <- data.frame(weight = c(transgenic, commercial),   
 type = c(rep("transgenic", 45), rep("commercial", 54)))



### Estimate the size of the difference between the two means

means <- chicken\_data %>%   
 group\_by(type) %>%   
 summarise(mean\_weight = mean(weight))

## `summarise()` ungrouping output (override with `.groups` argument)

means

## # A tibble: 2 x 2  
## type mean\_weight  
## <chr> <dbl>  
## 1 commercial 45.0  
## 2 transgenic 45.1

diff\_means <- means[1, 2] - means[2, 2]  
diff\_means

## mean\_weight  
## 1 -0.1533333

The estimated mean difference is -0.153 grams.

### Estimate the standard error

chicken\_stats <- chicken\_data %>%   
 group\_by(type) %>%   
 summarise(mean\_weight = mean(weight),  
 sd\_weight = sd(weight),   
 n = length(weight))

## `summarise()` ungrouping output (override with `.groups` argument)

Use the output to calculate the SE:

### Calculate the t-statistic

What is the chance of observing the t-statistic -0.193 on the t-distribution with the appropriate degrees of freedom?

To answer this, we would need to calculate the degrees of freedom using the complicated formula from a few slides previous. We won’t do this. Instead, we will ask R to do the test for us (and verify that our calculated t-statistic matches R’s test)

### t.test in R

Pay attention to the arguments specified by t.test. Below, the first argument is the weight data for the commercial chickens and the second argument is the weight data for the transgenic chickens.

commercial\_weight <- chicken\_data %>% filter(type == "commercial") %>% pull(weight)  
transgenic\_weight <- chicken\_data %>% filter(type == "transgenic") %>% pull(weight)  
  
t.test(commercial\_weight, transgenic\_weight, alternative = "two.sided")

##   
## Welch Two Sample t-test  
##   
## data: commercial\_weight and transgenic\_weight  
## t = -0.19293, df = 95.344, p-value = 0.8474  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -1.731044 1.424377  
## sample estimates:  
## mean of x mean of y   
## 44.98889 45.14222

What happens if you flip the order?

* If you flip the order, the test statistic flips signs (from negative to positive or vice versa) as does the values for the CI. The p-value and df stays the same.

### Robustness of the two-sample t-test

* These procedures are more robust than the one-sample t-test, especially if the data are skewed.
* When the sizes of the two samples are equal and the two populations being compared have similar shapes, the two-sample t-test will work well for sample sizes as small as .
* When the two populations have different shapes, larger samples are needed (e.g., one skewed left and the other skewed right).