STAT 210
Applied Statistics and Data Analysis
Two-Sample Problems

Joaquin Ortega

Fall 2020

Comparing two population means

Comparing two population means

The usual test for comparing the means from two populations is the two-sample t test.

Suppose we have samples

$$x_1, x_2, \ldots, x_{n_1} \quad \text{and} \quad y_1, y_2, \ldots, y_{n_2}$$

and assume they come from the normal distributions

$$N(\mu_1, \sigma_1^2) \quad \text{and} \quad N(\mu_2, \sigma_2^2)$$

and we want to test the null hypothesis

$$H_0 : \mu_1 = \mu_2.$$

Let $\bar{x}$ and $\bar{y}$ be the sample means for each sample, and s_x^2 and s_y^2 be the corresponding variances.

Comparing two population means

Recall that the *standard error* for the sample mean is the standard deviation of the sampling distribution for the mean, given by $\sigma_1/\sqrt{n_1}$ for the first population and $\sigma_2/\sqrt{n_2}$ for the second.

These unknown values are estimated by

$$\frac{s_i}{\sqrt{n}_i}$$

where

$$s_i^2 = \frac{1}{n_1 - 1}\sum_{j=1}^{n_1}(x_j - \bar{x})^2$$

for $i = 1, 2$.

Comparing two population means

The test statistic, in this case, is the difference between the sample means divided by the standard error of the difference of the means:

$$\frac{\bar{x} - \bar{y}}{SEDM}$$

The standard error of the difference of the means is calculated by using the fact that for two independent random variables, the variance of their difference is the sum of the variances.

Comparing two population means

Let X and Y be two random variables and denote by μ_X and μ_Y their means.

$$\begin{aligned}
Var(X - Y) &= E[(X - Y - (\mu_X - \mu_Y))^2] \\
&= E[((X - \mu_X) - (Y - \mu_Y))^2] \\
&= E[(X - \mu_X)^2 + (Y - \mu_Y)^2 - 2(X - \mu_X)(Y - \mu_Y)] \\
&= Var(X) + Var(Y) - 2E[(X - \mu_X)(Y - \mu_Y)] \qquad (1)
\end{aligned}$$

The expected value in the last term is known as the covariance between X and Y:

$$Cov(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$$

If the variables are independent, the covariance is zero and

$$Var(X - Y) = Var(X) + Var(Y)$$

Case 1: Equal variances

If we assume that the two variances are equal, which corresponds to the ‘classical’ test, the standard error is calculated using a ‘pooled’ estimator for the variance based on the standard deviations from the two groups.

The test is implemented in `R` with the function `t.test()` as in the one-sample case. The option `var.equal` must be set to `TRUE`.

Case 1: Equal variances

```r
t.test(fish1,fish2, var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  fish1 and fish2
## t = 0.37877, df = 68, p-value = 0.706
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -2.489833  3.656499
## sample estimates:
## mean of x mean of y
##  49.63333  49.05000
```

Case 1: Equal variances

```r
with(fish.df,t.test(fish ~ group, var.equal = TRUE))
```

```
##
##  Two Sample t-test
##
## data:  fish by group
## t = 0.37877, df = 68, p-value = 0.706
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -2.489833  3.656499
## sample estimates:
## mean in group 1 mean in group 2
##        49.63333        49.05000
```

Case 2: Unequal variances

The classical t test requires equal variances, but there is an extension due to Welch to the case of unequal variances.

In this situation, variances are estimated separately for each sample.

The distribution for the statistic is not the t distribution, but it may be approximated by a t distribution with a number of degrees of freedom that is calculated from the sample standard deviations and the group sizes. This parameter is usually not an integer.

This is the default test in `t.test()`.

Case 2: Unequal variances

```
t.test(fish1,fish2)
```

```
##
##  Welch Two Sample t-test
##
## data:  fish1 and fish2
## t = 0.34458, df = 38.246, p-value = 0.7323
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -2.843034  4.009701
## sample estimates:
## mean of x mean of y
##  49.63333  49.05000
```

The result is usually not very different unless the group sizes and standard deviations are very different.

Comparison of Variances

Comparison of Variances

One may also wish to compare variances from different samples. The test for this is known as Fisher's F test.

The test is based on the quotient between the two variances, but with the condition that the bigger variance must be on top.

The distribution for this quotient is known as Fisher's F distribution. It has two parameters that correspond to the degrees of freedom of the numerator and the denominator.

```
c(var(fish1), var(fish2))
```

```
## [1] 74.17126 15.74103
```

Comparison of Variances

```
(F.ratio <- var(fish1)/var(fish2))
```

```
##              [,1]
## [1,] 4.711972
```

So the group 2 variance is about five times the variance for the first group.

Now calculate the probability of getting a value at least as big as this one and multiply this by 2.

```
2*(1-pf(F.ratio, length(fish1)-1, length(fish2)-1))
```

```
##                  [,1]
## [1,] 1.033866e-05
```

Comparison of Variances

```r
var.test(fish1,fish2)
```

```
##
##  F test to compare two variances
##
## data:  fish1 and fish2
## F = 4.712, num df = 29, denom df = 39, p-value = 1.034e-05
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  2.401777 9.577864
## sample estimates:
## ratio of variances
##           4.711972
```

Wilcoxon's test

Wilcoxon's test

This non-parametric test is useful when the assumption of normality is not justified.

The test statistic W is calculated as follows:

- Mix the two samples and order them.
- Calculate the ranks for the elements of each group and sum the ranks for one of the groups

The problem now is reduced to sampling n_1 values from the numbers 1 to $n_1 + n_2$.

Wilcoxon's test

For the fish example,

```r
wilcox.test(fish1,fish2)
```

```
## Warning in wilcox.test.default(fish1, fish2): cannot compute exact p
## ties

##
##  Wilcoxon rank sum test with continuity correction
##
## data:  fish1 and fish2
## W = 561.5, p-value = 0.6512
## alternative hypothesis: true location shift is not equal to 0
```

Tests on paired samples.

Tests on paired samples.

In the second example we considered in the previous video, we had results for a reading test before and after the subjects go through a reading comprehension improvement method.

Let us now focus on one of the groups, group 2, say. We would like to know if the treatment (method) had some effect on the outcome (reading comprehension).

This is an example of paired data: we have scores before and after treatment for the same subjects.

In this case, the two samples are clearly not independent: the score after treatment will be related to the score before treatment, because it corresponds to the same subject.

Tests on paired samples.

Analyzing this data as if the samples were independent could lead to severe errors.

Remember that we showed that

$$Var(X - Y) = Var(X) + Var(Y) - 2Cov(X, Y)$$

If the covariance between X and Y is positive, by the equation above, the variance of the difference will be smaller than the sum of the individual variances.

This is an advantage because it makes easier it to detect significant differences between the means.

In `R`, the same `t.test` function with the option `paired` set to `TRUE` will do the test for paired data.

Tests on paired samples.

```r
t.test(group2.pre, group2.post)
```

```
##
##  Welch Two Sample t-test
##
## data:  group2.pre and group2.post
## t = -2.3618, df = 196.87, p-value = 0.01916
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -2.954324 -0.265676
## sample estimates:
## mean of x mean of y
##     77.43     79.04
```

Tests on paired samples.

```r
t.test(group2.pre, group2.post,paired = TRUE)
```

```
##
##  Paired t-test
##
## data:  group2.pre and group2.post
## t = -2.4167, df = 99, p-value = 0.01749
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -2.9318775 -0.2881225
## sample estimates:
## mean of the differences
##                   -1.61
```

```r
cor(group2.pre,group2.post)
```

```
## [1] 0.04503046
```

A more extreme example

This example comes from Box, Hunter & Hunter, *Statistics for Experimenters*.

In the experiment, two types of material for shoe soles are compared, A and B.

These were randomly assigned to the left and right shoes of 10 boys so that each boy had one shoe with each material.

```
library(MASS)
shoes
```

```
## $A
##  [1] 13.2  8.2 10.9 14.3 10.7  6.6  9.5 10.8  8.8 13.3
##
## $B
##  [1] 14.0  8.8 11.2 14.2 11.8  6.4  9.8 11.3  9.3 13.6
```

A more extreme example

For illustration purposes, let's do a standard t test without pairing. We will consider the two alternatives available, equal and unequal variances.

```
with(shoes,t.test(A, B, var.equal = TRUE))
```

```
##
##  Two Sample t-test
##
## data:  A and B
## t = -0.36891, df = 18, p-value = 0.7165
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -2.744924  1.924924
## sample estimates:
## mean of x mean of y
##     10.63     11.04
```

A more extreme example

```r
with(shoes,t.test(A, B))
```

```
##
##  Welch Two Sample t-test
##
## data:  A and B
## t = -0.36891, df = 17.987, p-value = 0.7165
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -2.745046  1.925046
## sample estimates:
## mean of x mean of y
##     10.63     11.04
```

A more extreme example

And now we do a paired test.

```
with(shoes,t.test(A, B, paired = TRUE))
```

```
##
##  Paired t-test
##
## data:  A and B
## t = -3.3489, df = 9, p-value = 0.008539
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.6869539 -0.1330461
## sample estimates:
## mean of the differences
##                   -0.41
```

```
with(shoes,cor(A,B))
```

```
## [1] 0.9882255
```

The Wilcoxon test for matched pairs.

There is also a non-parametric test for paired samples, which is the one-sample Wilcoxon signed-rank test applied to the differences.

```r
with(shoes, wilcox.test(A, B, paired = TRUE))
```

```
## Warning in wilcox.test.default(A, B, paired = TRUE): cannot compute
## value with ties

##
##  Wilcoxon signed rank test with continuity correction
##
## data:  A and B
## V = 3, p-value = 0.01431
## alternative hypothesis: true location shift is not equal to 0
```