

IEHC0046 BASIC STATISTICS FOR MEDICAL SCIENCES

Analysis of Continuous Data II: Practical

10 October, 2020

In this practical session, we will learn how to use Stata to compare two means. First, we will compare means from independent samples (i.e. unconnected observations). Next, we will compare means from paired samples.

We'll be working with the ELSA dataset again. Let's also load the `tidyverse()` package as we will need to create some new variables in this session.

Remember to use a script to save your code and to change your working directory so you can load the ELSA dataset easily.

```
load("elsa.Rdata")
```

1. Revising the previous session

1.1. Calculating means and 95% confidence intervals (CIs)

First, to recap from last week, please calculate the 95% confidence intervals for the mean for systolic blood pressure (`sbp`), diastolic blood pressure (`dbp`), BMI (`bmi`) and baseline cholesterol (`chol`).

```
t.test(elsa$sbp)
```

```
##
## One Sample t-test
##
## data:  elsa$sbp
## t = 370.7, df = 2691, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  139.6404 141.1256
## sample estimates:
## mean of x
##  140.383
```

```
t.test(elsa$dbp)
```

```
##
## One Sample t-test
##
## data:  elsa$dbp
## t = 337.22, df = 2691, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  76.87486 77.77409
## sample estimates:
## mean of x
##  77.32448
```

```
t.test(elsa$bmi)

##
## One Sample t-test
##
## data: elsa$bmi
## t = 331.71, df = 2929, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
## 27.36890 27.69438
## sample estimates:
## mean of x
## 27.53164
```

```
t.test(elsa$chol)

##
## One Sample t-test
##
## data: elsa$chol
## t = 269.34, df = 2474, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
## 5.936061 6.023131
## sample estimates:
## mean of x
## 5.979596
```

Note, what we have done here is repetitive. We could have made this simpler using the `map()` function from the package `purrr` which is loaded with `tidyverse`. `map()` takes a collection of objects and repeats a function on each object within the collection. If we pass `map()` a `data.frame`, it will carry out a function for each column (variable). The first input to `map()` is the collection of objects, the second input is the function to be repeated.

```
elsa %>%
  select(sbp, dbp, bmi, chol) %>% # Keep the relevant variables
  map(t.test)

## $sbp
##
## One Sample t-test
##
## data: .x[[i]]
## t = 370.7, df = 2691, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
## 139.6404 141.1256
## sample estimates:
## mean of x
## 140.383
##
## $dbp
##
## One Sample t-test
##
```

```
## data: .x[[i]]
## t = 337.22, df = 2691, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
## 76.87486 77.77409
## sample estimates:
## mean of x
## 77.32448
##
##
## $bmi
##
## One Sample t-test
##
## data: .x[[i]]
## t = 331.71, df = 2929, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
## 27.36890 27.69438
## sample estimates:
## mean of x
## 27.53164
##
##
## $chol
##
## One Sample t-test
##
## data: .x[[i]]
## t = 269.34, df = 2474, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
## 5.936061 6.023131
## sample estimates:
## mean of x
## 5.979596
```

Remember, the pipe operation, `%>%`, takes the object on the left hand side and puts it into the first argument of the function on the right hand side.

1.1. One sample t-test

Q: Compare whether the mean systolic blood pressure in this sample is significantly different from the population mean (estimated to be 130 mmHg).

The variable of interest is systolic blood pressure (`sbp`). Recall you can use the function `look_for()` from the package `janitor` to explore the `elsa` dataset.

```
t.test(elsa$sbp, mu = 130)
```

```
##
## One Sample t-test
##
## data: elsa$sbp
## t = 27.417, df = 2691, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 130
```

```
## 95 percent confidence interval:
## 139.6404 141.1256
## sample estimates:
## mean of x
## 140.383
```

Mean systolic blood pressure in the sample is 140.4 mmHg (95% CI: 139.6, 141.1). This is higher than the reference value of 130 mmHg. The output shows a small p-value (< 0.01) and a large t-statistic. The results suggest that the mean systolic blood pressure in the sample from which ELSA participants are drawn is not 130 mmHg.

2. Comparison of two independent samples: the unpaired t-test

If we are asked to assess the potential differences in a continuous normally distributed outcome (systolic blood pressure at baseline) between two groups of a dichotomous variable (**sex**) then we can use the (unpaired) t-test. This test is only for examining differences between two groups, if you need to do some work in creating the appropriate variable for that, such as recoding an existing variable has more than two categories. The data that we collected in order to make the comparison are from independent samples; the measurements in men are not connected in any way to the measurements in women (they are different people).

The same principle previously discussed for hypothesis testing for comparing a mean with a hypothesised value applies to the comparison of two sample means. In this case, the null hypothesis states that there is no difference between the two means in the populations from which the sample means are drawn. If the null hypothesis is correct, then the difference between the two means that we have observed in the sample would be due to sampling variation. We then estimate the probability of observing this difference in the sample if the null hypothesis (of no difference in the population) were true. We do this by taking the difference between the two means in our sample and assessing its relative position in the respective sampling distribution of mean differences. In other words, we assess whether the difference between the sample means lies as far away from the hypothesised difference between the population means so that to be included in the margins of the sampling distribution that correspond to the extreme 5% of means (2.5% lower and 2.5% upper) or not.

We can compare independent means using the syntax `t.test(x ~ group, data)`. For the present question:

```
t.test(sbp ~ sex, elsa)
```

```
##
## Welch Two Sample t-test
##
## data: sbp by sex
## t = 1.9418, df = 2667.9, p-value = 0.05227
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.01426405 2.92279692
## sample estimates:
## mean in group male mean in group female
## 141.1976 139.7434
```

Q: Are there differences in systolic blood pressure at baseline between men and women?

The output provides the mean systolic blood pressure in each group (males and females). It also provides a 95% CI for the difference in the means (-0.014, 2.922) and a p-value for whether the difference in means is statistically significantly different from zero ($p = 0.05227$). We can infer that the p value is greater than 0.05 if the 95% CIs overlap zero, which they do in this case - but only just. The results state that if we assume the population mean in males and females is zero, we would get a difference in means at least as large as 5.227% of samples.

Q: Are there differences by BMI groups?

For this, we need to recode the grouped BMI (bmi4) into two groups: (i) people with BMI under 25, (ii) people with BMI over 25. Let's use the function `fct_collapse()` from the package `forcats` (part of the `tidyverse`) to do this.

```
levels(elsa$bmi4)

## [1] "Under 20" "Over 20-25" "Over 25-30" "Over 30"

elsa <- elsa %>%
  mutate(bmi_bin = fct_collapse(bmi4,
                                "Less than 25" = levels(bmi4)[1:2],
                                "Over 25" = levels(bmi4)[3:4]))
table(elsa$bmi_bin, elsa$bmi4, useNA = "ifany")
```

```
##
##           Under 20 Over 20-25 Over 25-30 Over 30 <NA>
## Less than 25      63      814          0          0      0
## Over 25           0          0      1336      717      0
## <NA>              0          0          0          0     199
```

I used the `table()` function to confirm that we recoded the BMI variable correctly. This is good practice.

Now, we can compare mean systolic blood pressure by BMI groups.

```
t.test(sbp ~ bmi_bin, elsa)

##
## Welch Two Sample t-test
##
## data:  sbp by bmi_bin
## t = -8.9472, df = 1412, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -9.122468 -5.841646
## sample estimates:
## mean in group Less than 25      mean in group Over 25
##           134.9582              142.4403
```

The output provides the mean systolic blood pressure in each group (low and high BMI). It also provides a 95% CI for the difference in the means (-9.12, -5.84) and a p-value for whether the difference in means is statistically significantly different from zero ($p < 0.01$). We can infer that the p value is less than 0.05 because the 95% CIs do not overlap zero. The null hypothesis of no difference can be rejected. The results suggest that systolic blood pressure is higher among high BMI individuals.

3. Comparison of two dependent samples: the paired t-test

For the next few exercises, we are going to use a different dataset: `contin2.Rdata`. This is data extracted from a trial assessing the effectiveness of surgical periodontal treatment (treatment of the severe inflammation of the gums) using different oral and general health outcomes.

```
load("contin2.Rdata")
```

Up to now, the difference between the two means referred to independent observations (separate groups of subjects). However, there are cases where the data are paired. Data are considered to be **paired** in the following circumstances:

1. When the **same individuals are studied more than once**, usually in different circumstances (**pre- and post-treatment** measurements). This happens, for example, in a clinical trial evaluating the effect of a treatment (e.g. periodontal treatment) on an outcome (e.g. quality of life); the quality of life of

each patient is measured first before and then after treatment. Then, for each patient we have a pair of measurements of the outcome and we should consider all these pairs when we estimate the effect of treatment on the outcome.

2. When we have two **different groups of subjects who have been individually matched** (e.g. in a matched pair case-control study or in a clinical trial with matched controls). In this case, individuals are matched during the sample selection, so that they have key characteristics in common (e.g. age, sex, socioeconomic status etc.). Then, the data does not consist of two independent groups, but rather of case-control pairs and this pairing should be considered in the analysis.

When the data are paired, this should guide the analysis appropriately. In such cases, we need to focus on the **difference in the outcome measurement between each pair**. The analysis is effectively reduced to a one-sample problem: we calculate the difference between each pair ($x = x_1 - x_2$) and treat these differences as a single sample of differences.

3.1. Differences before and after treatment for the whole sample

If data are paired (both outcomes are measured on the same person), then you use a paired t-test to examine differences. The code for this is below.

```
t.test(contin2$oidpsco2, contin2$oidpsco1, paired = TRUE)

##
## Paired t-test
##
## data: contin2$oidpsco2 and contin2$oidpsco1
## t = -3.1052, df = 44, p-value = 0.003322
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -5.851590 -1.245379
## sample estimates:
## mean of the differences
## -3.548485
```

The first two arguments are the two (paired) vectors and the `paired` argument is set to `TRUE`. The difference is the first vector minus the second (in this case, the change between measurement 1 and 2).

Q: Are there differences before and after treatment for the whole sample in terms of quality of life (oidpsco1 and oidpsco2)?

The sample had lower quality of life score, comparing measurement 2 (oidpsco2) to measurement 1 (oidpsco1). (A lower score on the measure indicates better quality of life.) The confidence interval does not overlap zero and the p-value is very small (< 0.01). This suggests that quality of life improved through time.

3.2. Differences before and after treatment between treatment (tx) groups.

Again, we use the paired t-test here but we need to carry out the test separately for different subgroups in the sample. We can use sub-setting as so:

```
levels(contin2$tx)

## [1] "control (scale+polish)" "test (intensive care + extractions)"

t.test(contin2$oidpsco2[contin2$tx == levels(contin2$tx)[1]],
       contin2$oidpsco1[contin2$tx == levels(contin2$tx)[1]],
       paired = TRUE)

##
## Paired t-test
##
```

```
## data: contin2$oidpsco2[contin2$tx == levels(contin2$tx)[1]] and contin2$oidpsco1[contin2$tx == levels(contin2$tx)[2]]
## t = -2.3922, df = 17, p-value = 0.02857
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -8.3309442 -0.5225912
## sample estimates:
## mean of the differences
## -4.426768

t.test(contin2$oidpsco2[contin2$tx == levels(contin2$tx)[2]],
       contin2$oidpsco1[contin2$tx == levels(contin2$tx)[2]],
       paired = TRUE)
```

```
##
## Paired t-test
##
## data: contin2$oidpsco2[contin2$tx == levels(contin2$tx)[2]] and contin2$oidpsco1[contin2$tx == levels(contin2$tx)[1]]
## t = -2.0171, df = 26, p-value = 0.05412
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -5.98241201 0.05648616
## sample estimates:
## mean of the differences
## -2.962963
```

Q: Are the results similar for the treatment groups? Are they similar with the ones shown for the whole sample?

The results for the quality of life indicate improvement in quality of life in the control group (difference of 4.4; 95% CI: 0.5, 8.3; p=0.029), while the test group reported some improvement in quality of life, though the respective result was marginally not significant (difference of 3.0; 95% CI: -0.1, 6.0; p=0.054).

3.3. Testing the difference in quality of life changes between treatment groups.

This is about assessing the difference between two groups in relation to difference before and after treatment. The data are NOT paired, since you only have one observation per participant, i.e. the difference between before- and after-treatment. Firstly we need to calculate this variable for every participant (difference in quality of life before and after treatment), then we will use that to assess differences in that variable between the subgroups (treatment and control).

Let's create a new variable using the `mutate()` function from the `tidyverse` and use the `t.test()` formula syntax for Section 2 to test for differences between groups.

```
contin2 <- contin2 %>%
  mutate(oidpsco_diff = oidpsco2 - oidpsco1)
t.test(oidpsco_diff ~ tx, contin2)

##
## Welch Two Sample t-test
##
## data: oidpsco_diff by tx
## t = -0.61956, df = 35.864, p-value = 0.5395
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -6.256093 3.328483
## sample estimates:
## mean in group control (scale+polish)
```

```
## -4.426768
## mean in group test (intensive care + extractions)
## -2.962963
```

Q: Is there a difference between the treatment groups in relation to the before- and after-treatment differences in probing pocket depth?

The improvement in the quality of life score in the control group was greater (since lower scores indicates better quality of life, the higher the score the worse the quality of life) than the respective reduction in the treatment group: -2.96 (-5.98, 0.06) vs. -4.42 (-8.33, -0.52). The difference in these differences was -1.46 (-6.2, 3.3). The confidence intervals for this difference are wide. We can't reject the null hypothesis that there is actually no difference in quality of life between the treatment and the control group.

Formative Exercise

Practice more with the paired t-test. Please assess whether there are differences before and after treatment in probing pocket depth (ppd_0 – ppd_2m).