T-tests: one-sample, two-sample, and paired tests

XDASI Fall 2021

Contents

Review

- Random samples
- Standard normal distribution
  - What is a *z*-score?
  - What are the quantiles?
  - How do you capture height vs. area?
- Sampling distribution of the sample mean
- Standard error
- Confidence intervals
- What is a *z*-score?
- *t*-statistics vs. *z*-statistics : when is each appropriate?
- P values : one- vs. two-sided
- What does qt(0.975) represent?

This worksheet illustrates basic concepts and practical application of significance tests for different experimental scenarios.

Sample data

To illustrate the various types of *t*-test, we will use a simple case study where a drug was administered to 10 random patients (or test subjects) and its effect was compared to that of a placebo pill. Three different experimental designs are illustrated:

- Treatment sample vs. control population (one-sample test)
- Two random samples: treatment vs. control (two-sample test)
- One sample measured before and after treatment (paired test)

In all three scenarios, measurements were collected to answer the question:

Is there a significant difference between the control subjects and the those who were given the drug?

To generate sample data for this exercise, I simulated some normally distributed data for a large control population and wrote this to a file (`placebo_pop.csv`). I then drew two random samples from normal distributions with different means.

```
# population data
# simulate 1000 samples from a normal with the same parameters
Placebo_pop = rnorm(1000, mean = 50.482, sd = 2.693)
head(Placebo_pop)
## [1] 49.24308 47.52654 55.64826 47.13595 48.98226 48.94788

# write to file
# Placebo_pop = data.frame(value = Placebo_pop)
# write.table(Placebo_pop, file = "placebo_pop.csv", row.names = FALSE, col.names = FALSE)

# load from file
# Placebo_pop = read.table("placebo_pop2.csv", header=FALSE)
# Placebo_pop = Placebo_pop[,1]  # turns data frame into a simple vector
# head(Placebo_pop)

# sample data
Placebo = c(54,51,58,44,55,52,42,47,58,46)
Drug = c(54,73,53,70,73,68,52,65,65,60)
```

Sample vs. population parameters

Compare the mean and SD of the Placebo population and the Placebo sample. Do these look pretty different? Why?

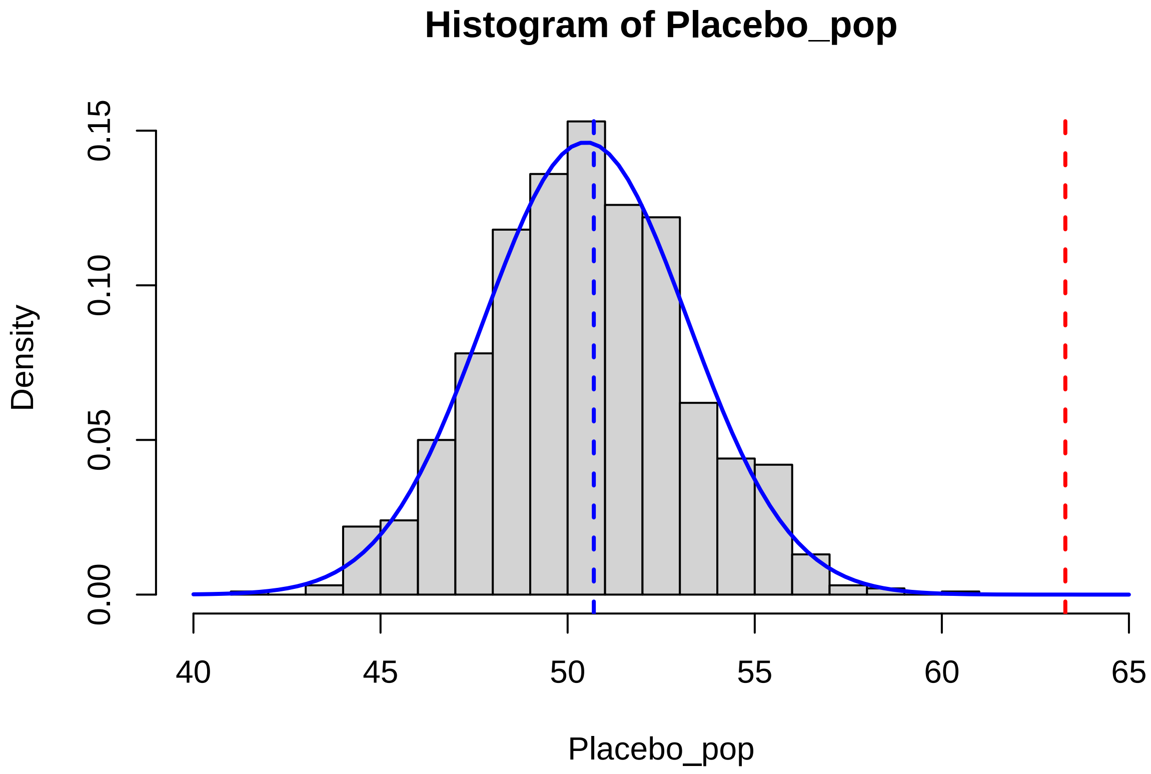
```
## check the parameters from the true population
paste("Pop. mean = ",round(mean(Placebo_pop),3),
      " ; Pop. SD = ",round(sd(Placebo_pop),4))
## [1] "Pop. mean =  50.484  ; Pop. SD =  2.7283"

## check the parameters from the estimated population
paste("Sample mean = ",round(mean(Placebo),3),
      " ; Sample SD = ",round(sd(Placebo),4))
## [1] "Sample mean =  50.7  ; Sample SD =  5.7164"
paste("Drug sample = ",round(mean(Drug),3),
      "; Drug sample = ",round(sd(Drug),4))
## [1] "Drug sample =  63.3 ; Drug sample =  8.111"
```

Placebo population distribution

Let's take a look at the "population" and see where the two samples means fall relative to the control distribution.

```
## visualize the distribution and sample means
hist(Placebo_pop, breaks=20, freq=FALSE, xlim=c(40,65))
xfit = seq(40,65,length=100)
yfit = dnorm(xfit,mean=mean(Placebo_pop),sd=sd(Placebo_pop))
lines(xfit,yfit,col="blue",lwd = 2)
abline(v = mean(Placebo), col="blue", lwd=2, lty=2)
abline(v = mean(Drug), col="red", lwd=2, lty=2)  # off scale
```



t-distrubution

Confidence Intervals

The standard error is very helpful because it gives us an idea of how close our data are to the actual mean. We can use the SE to help define **Confidence Intervals (CI)** for the the actual population means from which the samples were drawn.

Here, we assume that we have random samples and that the measurements are normally distributed in the population. The estimated distribution of the sample means will then follow a t-distribution, which is similar to a standard normal distribution but with heavier tails. As the sample size increases, the tails of the t-distribution become smaller and it resembles closely the Z-distribution.

Let's calculate the 95% CI for the mean of each population that the samples are drawn from. Here are the steps we need:

1. Use the `qt()` *quantile function* to identify the critical t-score, t_{crit}, for the the 95% CI with the appropriate df.
2. Estimate the mean and SEM for the Placebo and Drug samples.
3. Use these values to compute the 95% CI's for the two samples.

```
## t_critical: t-score for 97.5% area (range is from 2.5% to 97.5%)
## also need degrees of freedom for this distribution (length - 1)
##   df = same for Placebo and Drug (9)

# Placebo and Drug are the same length so this works for both df
t_0.975 = qt(p = 0.975, df = length(Placebo)-1)  # t-score for 97.5th percentile
paste0("t_crit = ",round(t_0.975,4))
## [1] "t_crit = 2.2622"

# note this is bigger than the corresponding z-score would be
# n_0.975 = qnorm(p = 0.975)
# n_0.975


# placebo SEM
Pmean = mean(Placebo)
Pse=sd(Placebo)/sqrt(length(Placebo))
Pmean
## [1] 50.7
Pse
## [1] 1.8077

# drug SEM
Dmean = mean(Drug)
Dse=sd(Drug)/sqrt(length(Drug))
Dmean
## [1] 63.3
Dse
## [1] 2.564934

# placebo and drug 95% CI
c(Pmean - Pse * t_0.975, Pmean + Pse * t_0.975)
## [1] 46.6107 54.7893
c(Dmean - Dse * t_0.975, Dmean + Dse * t_0.975)
## [1] 57.49772 69.10228
```

Draw a strip chart showing the mean, SEM, and 95%CI.

```
# probably very significant b/c CI's do not overlap
```

Using the rules of thumb given in your textbook, can you conclude by eye that the samples come from different populations? Why or why not?

```
# not for sure, but it's very unlikely since CI's don't overlap
```

One-sample t-test

A one-sample t-test compares a sample against expected parameters for a larger population.

The `t.test()` function in R takes a vector of input data (a sample) and automatically finds the degrees of freedom for the corresponding t-distribution based on the length of the vector.

Run a one-sample t-test for each sample against Placebo_pop. Do the samples look like they both came from the same parent population?

```
# Placebo sample
t.test(Placebo, mu=mean(Placebo_pop))
##
##  One Sample t-test
##
## data:  Placebo
## t = 0.11935, df = 9, p-value = 0.9076
## alternative hypothesis: true mean is not equal to 50.48425
## 95 percent confidence interval:
##  46.6107 54.7893
## sample estimates:
## mean of x
##      50.7

# Drug sample
t.test(Drug, mu=mean(Placebo_pop))
##
##  One Sample t-test
##
## data:  Drug
## t = 4.9965, df = 9, p-value = 0.0007425
## alternative hypothesis: true mean is not equal to 50.48425
## 95 percent confidence interval:
##  57.49772 69.10228
## sample estimates:
## mean of x
##      63.3
```

You can inspect the `t.test` object using `str()`; it's a list containing a bunch of information about the results of the t-test.

Notice that the output of `t.test()` includes an estimate for the 95% CI. We can check our manual calculations by extracting just the CIs from the function output. The precise expression to get the CI is `t.test()$confint[1:2]`.

Do this for both the Drug and the Placebo samples. Are these the same as what you calculated above?

```
# confidence intervals and p-values from t-tests
#str(Pt.test)

t.test(Placebo, mu=mean(Placebo_pop))$conf.int[1:2] # lower and upper CI limits
## [1] 46.6107 54.7893
```

```
#t.test(Placebo, mu=mean(Placebo_pop))$conf.int      # also indicates 0.95 is the range

t.test(Drug, mu=mean(Placebo_pop))$conf.int[1:2]
## [1] 57.49772 69.10228

t.test(Placebo, mu=mean(Placebo_pop))$p.value
## [1] 0.9076172
t.test(Drug, mu=mean(Placebo_pop))$p.value
## [1] 0.0007424835
```

Two-sample t-test

The **null hypothesis** is that the samples come from the same population, so their means should be the same.

$$H_o : \mu_{Drug} = \mu_{Placebo}$$
$$H_A : \mu_{Drug} \neq \mu_{Placebo}$$

Difference between sample means

Another way to frame H_o is to say that we expect the **difference between the two sample means to be 0**. We know this because if we were to plot a distribution of random variables with the same mean, we would expect to get a mean difference of 0. This is the null hypothesis for a two-sample t-test.

$$H_o : \mu_{Drug} - \mu_{Placebo} = 0$$
$$H_A : \mu_{Drug} - \mu_{Placebo} \neq 0$$

We already know that the sampling distribution of the sample mean is normally distributed, and that the sum or difference of two normally distributed variables is also normally distributed.

Therefore, the difference in sample means must also be normally distributed. For small sample sizes, we can use t-statistics to test whether the means are the same, and we can estimate (with 95% confidence) the true difference in the means of the two populations.

95% CI for the difference in sample means

To calculate the 95% CI for $\bar{\mu}_D - \bar{\mu}_P$ using the t-distribution, we will need several things:

- the observed mean difference
- the SE of the difference
- the degrees of freedom
- the critical t-value

In this example, the standard error is the same whether we use the **pooled variance** formula (for equal variances) or the simpler formula from **Welch's approximate** t-test (for unequal variances), which uses individual sample variances:

$$SE_{\bar{Y}_1 - \bar{Y}_2} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

For ***independent samples***, the degrees of freedom are just 2 less than the sum of the two sample sizes: $df = n_1 + n_2 - 2$.

```
# mean difference
mean_diff = mean(Drug) - mean(Placebo)  # 12.6
mean_diff
## [1] 12.6

# SE of the difference
se_diff = sqrt ( var(Placebo)/length(Placebo) +
                   var(Drug) /length(Drug) )
se_diff
## [1] 3.13794

# critical t-value
t_crit = abs(qt(p=0.025, df=18))  # same as qt(p=0.975, df=18)
t_crit
## [1] 2.100922

# 95% CI
c(mean_diff - t_crit * se_diff, mean_diff +  t_crit * se_diff)
## [1]  6.007433 19.192567
```

Notice that the confidence interval does not span 0, which means it is extremely unlikely that the true difference between the Drug and Placebo populations is 0.

Test statistic and p-value

Since we expect that we have random samples drawn from a normal distribution, and our samples are pretty small, it is valid to use t-statistics to test our null hypothesis.

The t-score for the difference in the means is computed in the same way as the z-score for a normal distribution: it is simply the difference *standardized* by the standard error. We already computed the mean and SE above, so this will be very easy!

First, calculate by hand the test statistic (the t-score for the observed difference) and use it to find the probability that our null hypothesis is true (the p-value for our test statistic).

Next, perform a two-sample test using the `t.test()`. Try both the standard version and Welch's approximate t-test.

```
# compute the test statistic for the mean difference
t_score = mean_diff / se_diff
t_score
## [1] 4.015373

# compute a two-sided p-value
2*pt(q = t_score,
     df = (length(Placebo)+length(Drug)-2),
     lower.tail = F)
## [1] 0.0008115625

# perform two versions of a 2-sample test using R
t.test(Drug, Placebo, var.equal = T)  # equal variances
##
```

```
##  Two Sample t-test
##
## data:  Drug and Placebo
## t = 4.0154, df = 18, p-value = 0.0008116
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   6.007433 19.192567
## sample estimates:
## mean of x mean of y
##      63.3      50.7
t.test(Drug, Placebo, var.equal = F)  # Welch (unequal variances)
##
##  Welch Two Sample t-test
##
## data:  Drug and Placebo
## t = 4.0154, df = 16.171, p-value = 0.0009803
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   5.95359 19.24641
## sample estimates:
## mean of x mean of y
##      63.3      50.7
```

How do the results of the manual calculation and the two t-tests compare?

What is similar and what is different about these results in comparison with the one-sample tests above?

Paired t-test

What if the data are taken from the ***same individuals***? For example, we could test the same patients before and after treatment. In this kind of experimental design, we say that the data are ***paired***. If there is no difference between the two measurements for each individual – for example, a new drug for blood pressure has no measurable benefit – then we would expect that our before and after values would be about the same on average.

The **paired t-test** is performed in the same way as the one-sample t-test, except we use the ***mean difference between paired measurements*** from the two samples, $\bar{X}_D$, to compute a test statistic. This is a nice design because it controls for inter-individual variation, however it is not appropriate when the two samples are truly independent.

Null hypothesis

Our ***null hypothesis*** is usually that the mean paired difference, D_o, is zero (we could set it to something else if our null hypothesis is that the difference between them is something else . . .).

For paired data, we assume that the two sets of measurements are arranged in the same order as the corresponding individuals (because we have good record-keeping practices!) The test statistic is:

$$t^* = \frac{\bar{X}_D - D_0}{\frac{s_D}{\sqrt{n}}} = \frac{\bar{X}_D - D_0}{SE_D} = \frac{\sqrt{n}}{s_D}(\bar{X}_D - D_0)$$

where $\bar{X_D}$ is the mean of the pairwise differences, s_D is the standard deviation of the pairwise differences, and D_0 is what we are testing (which in this case is 0).

Test statistic and p-value

To compute this t-statistic, we can simply subtract one vector from the other to obtain pair-wise differences for each individual, take the mean, and divide by the standard error. We then find the p value in the usual way. First, find the t-statistic and the p-value by hand.

```
pair_diff = Drug-Placebo

se_diff = sd(pair_diff)/sqrt(length(pair_diff))
t_stat = mean(pair_diff)/se_diff
t_stat
## [1] 4.108696
2*pt(t_stat, df=9, lower.tail = F)
## [1] 0.002642154
```

Relationship between one-sample and paired-sample tests

Study the above equation to convince yourself that the form of the t-statistic for a ***paired test*** is the same as that for a ***one-sample*** t-test, substituting $\bar{X}_D$ for $\bar{X}$, D_o for μ_o, and s_D for s.

It is important to note that the result from the **one-sample** t-test using the **paired differences** between the two samples gives the same result as a **two-sample** t-test with the **paired** option!

Use the `t.test()` function to perform a one-sample test that compares the paired difference vector against the expected mean difference, and then perform a paired t-test to verify that the two methods are equivalent.

```
# one-sample test: sample diff vs. Exp(mu) = 0
t.test(pair_diff, mu=0)
##
##  One Sample t-test
##
## data:  pair_diff
## t = 4.1087, df = 9, p-value = 0.002642
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##   5.662718 19.537282
## sample estimates:
## mean of x
##      12.6

# paired t-test: t.test(treatment, control, ...)
t.test(Drug, Placebo, paired = T, var.equal = T)
##
##  Paired t-test
##
## data:  Drug and Placebo
## t = 4.1087, df = 9, p-value = 0.002642
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   5.662718 19.537282
## sample estimates:
## mean of the differences
##                    12.6
```

Note: when performing any t-test with two samples, you must specify the **test set first**, and the **control set second** – otherwise you get a t-score that is on the opposite side of the distribution. This will not change your p-value, however your **confidence interval** will have the wrong sign!!!