Dr Isabella Gollini

Probability and Statistics

Lab – The testing of hypotheses and t-tests

Test procedure at the $100\alpha\%$ significance level. Conventionally, we use $\alpha = 0.05, 0.01$ or 0.001, i.e., tests at the 5%,1% or 0.1% significance level.

1. We begin by setting up a *statistical model*, which provides a theoretical framework for analysing the data.

2. We shall usually consider two competing hypotheses, a *null hypothesis* H_0 and an *alternative hypothesis* H_1.

3. We construct a test statistic whose distribution under H_0 is known, at least approximately.

4. Given α, if t is the observed value of the test statistic on a particular occasion, we *reject* H_0 *at the* $100\alpha\%$ *significance level* if and only if $|t| \geq k_\alpha$ if the test is one-sided, or $|t| \geq k_{\alpha/2}$ if the test is two-sided.

5. An alternative way of describing this procedure is in terms of p-values. We *reject* H_0 *at the* $100\alpha\%$ *significance level* if and only if p-value $\leq \alpha$.

One sample t-tests

1. **Statistical Model**

$$x_1, x_2, \ldots, x_n \overset{iid}{\sim} N\left(\mu, \sigma^2\right)$$

   with μ and σ^2 unknown.

2. **H_0 and H_1**

Null hypothesis H_0	Alternative hypothesis H_1				
	two-sided		one-sided greater		one-sided less
$\mu = \mu_0$	$\mu \neq \mu_0$	or	$\mu > \mu_0$	or	$\mu < \mu_0$

3. **Test Statistic**

   To test the null hypothesis H_0 we use the test statistic

$$t = \frac{(\bar{x} - \mu_0)}{\frac{s}{\sqrt{n}}} \quad \sim t_{n-1} \ .$$

4. **Rejection region Significance-level-α test**

two-sided		one-sided greater		one-sided less
$\lvert t\rvert \geq t_{n-1,\frac{\alpha}{2}}$	or	$t \geq t_{n-1,\alpha}$	or	$t \leq -t_{n-1,\alpha}$

5. **p-value**

two-sided		one-sided greater		one-sided less
$2(1 - F(\lvert t\rvert))$	or	$1 - F(t)$	or	$F(t)$

   where $F(t)$ is the c.d.f. of a t-distribution with $n-1$ degrees of freedom.

`R` function

Two-sided H_1:

```
t.test(x, mu = mu0)
```

One-sided greater H_1:

```
t.test(x, mu = mu0, alternative = "greater")
```

One-sided less H_1:

```
t.test(x, mu = mu0, alternative = "less")
```

Two-sample t-tests – Independent Samples

1. **Statistical Model**

$$x_1, x_2, \ldots, x_n \overset{iid}{\sim} N\left(\mu_1, \sigma^2\right) \qquad \text{and} \qquad y_1, y_2, \ldots, y_m \overset{iid}{\sim} N\left(\mu_2, \sigma^2\right)$$

   with μ_1, μ_2 and σ^2 unknown. The two samples are independent.

2. **H_0 and H_1**

Null hypothesis H_0	Alternative hypothesis H_1				
	two-sided		one-sided greater		one-sided less
$\mu_1 = \mu_2$	$\mu_1 \neq \mu_2$	or	$\mu_1 > \mu_2$	or	$\mu_1 < \mu_2$

3. **Test Statistic** Under H_0,

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{s^2 \left(\frac{1}{n} + \frac{1}{m}\right)}} \quad \sim t_{n+m-2} \ .$$

   where s^2 is the *pooled estimate of variance*, based on the data from both samples:

$$s^2 = \frac{(n-1)s_1^2 + (m-1)s_2^2}{n+m-2} \ ,$$

   where s_1^2 and s_2^2 are the sample variances for the samples from Population 1 and Population 2, respectively.

4. **Rejection region Significance-level-α test**

two-sided		one-sided greater		one-sided less
$\lvert t\rvert \geq t_{n+m-2,\frac{\alpha}{2}}$	or	$t \geq t_{n+m-2,\alpha}$	or	$t \leq -t_{n+m-2,\alpha}$

5. **p-value**

two-sided		one-sided greater		one-sided less
$2(1 - F(\lvert t\rvert))$	or	$1 - F(t)$	or	$F(t)$

   where $F(t)$ is the c.d.f. of a t-distribution with $n+m-2$ degrees of freedom.

`R` function

Two-sided H_1:

```
t.test(x, y, var.equal = TRUE)
```

One-sided greater H_1:

```
t.test(x, y, var.equal = TRUE, alternative = "greater")
```

One-sided less H_1:

```
t.test(x, y, var.equal = TRUE, alternative = "less")
```

Two-sample t-tests – Paired comparisons

1. **Statistical Model**

   $x_1, x_2, \ldots, x_n$ and $y_1, y_2, \ldots, y_n$ are paired (not independent) samples from Normal distributions with means μ_X and μ_Y (both unknown) respectively.

   The analysis will be based on consideration of the differences $d_i = x_i - y_i$:

   $$d_i \overset{iid}{\sim} N\left(\mu_D, \sigma_D^2\right)$$

   where

   $$\mu_D = \mu_X - \mu_Y \ ,$$

   and σ_D^2 are unknown.

2. H_0 **and** H_1

   The hypotheses that we test may be expressed in terms of μ_X and μ_Y or, equivalently, in terms of $\mu_D = \mu_X - \mu_Y$.

	Null hypothesis H_0	Alternative hypothesis H_1		
		two-sided	one-sided greater	one-sided less
In terms of μ_X, μ_Y:	$\mu_X = \mu_Y$	$\mu_X \neq \mu_Y$	$\mu_X > \mu_Y$	$\mu_X < \mu_Y$
In terms of μ_D:	$\mu_D = 0$	$\mu_D \neq 0$	$\mu_D > 0$	$\mu_D < 0$

3. **Test Statistic** Under H_0,

   $$t = \frac{\bar{x} - \bar{y}}{\frac{s_D}{\sqrt{n}}} = \frac{\bar{d}}{\frac{s_D}{\sqrt{n}}} \quad \sim t_{n-1} \ ,$$

   where $\bar{d}$ and s_D^2 are the sample mean and sample variance for $d_1, d_2, \ldots, d_n$.

4. **Rejection region Significance-level-α test**

two-sided		one-sided greater		one-sided less
$\lvert t\rvert \geq t_{n-1,\frac{\alpha}{2}}$	or	$t \geq t_{n-1,\alpha}$	or	$t \leq -t_{n-1,\alpha}$

5. p**-value**

two-sided		one-sided greater		one-sided less
$2(1 - F(\lvert t\rvert))$	or	$1 - F(t)$	or	$F(t)$

   where $F(t)$ is the c.d.f. of a t-distribution with $n-1$ degrees of freedom.

`R` function

Two-sided H_1:

```r
t.test(x, y, paired = TRUE)
```

or

```r
t.test(x - y)
```

One-sided greater H_1:

```r
t.test(x, y, paired = TRUE, alternative = "greater")
```

or

```r
t.test(x - y, alternative = "greater")
```

One-sided less H_1:

```r
t.test(x, y, paired = TRUE, alternative = "less")
```

or

```r
t.test(x - y, alternative = "less")
```

Hypothesis testing for a proportion

1. **Statistical Model**

   X is the number of success in n trials. $\hat{p} = \frac{r}{n}$ is the proportion of members in the sample with a characteristic. p is the proportion of individuals in the population who have the specified characteristic.

2. **H_0 and H_1**

Null hypothesis H_0	Alternative hypothesis H_1				
	two-sided		one-sided greater		one-sided less
$p = p_0$	$p \neq p_0$	or	$p > p_0$	or	$p < p_0$

3. **Test Statistic**

   To test the null hypothesis H_0 we use the test statistic

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \, , \sim N(0, 1)$$

4. **Rejection region Significance-level-α test**

two-sided		one-sided greater		one-sided less
$\lvert z \rvert \geq z_{\frac{\alpha}{2}}$	or	$z > z_\alpha$	or	$z < -z_\alpha$

5. **p-value**

two-sided		one-sided greater		one-sided less
$2(1 - \Phi(\lvert z \rvert))$	or	$1 - \Phi(z)$	or	$\Phi(z)$

   where $\Phi(z)$ is the c.d.f. of a standard Normal distribution.

R function

Two-sided H_1:

```
prop.test(x, n, p = p0, correct = FALSE)
```

One-sided greater H_1:

```
prop.test(x, n, p = p0, correct = FALSE, alternative = "greater")
```

One-sided less H_1:

```
prop.test(x, n, p = p0, correct = FALSE, alternative = "less")
```

Hypothesis testing for comparing two proportions

1. **Statistical Model**

   X_1 is the number of success in n_1 trials in population 1. $\hat{p}_1 = \frac{X_1}{n_1}$ is the proportion of members in a sample from population 1 with a characteristic. X_2 is the number of success in n_2 trials in population 2. $\hat{p}_2 = \frac{X_2}{n_2}$ is the proportion of members in a sample from population 2 with a characteristic. The two samples are independent.

2. **H_0 and H_1**

Null hypothesis H_0	Alternative hypothesis H_1				
	two-sided		one-sided greater		one-sided less
$p_1 = p_2$	$p_1 \neq p_2$	or	$p_1 > p_2$	or	$p_1 < p_2$

3. **Test Statistic**

   To test the null hypothesis H_0 we use the test statistic

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\hat{q}\left(1/n_1 + 1/n_2\right)}} \,, \sim N(0,1)$$

   where $\hat{p} = \frac{X_1+X_2}{n_1+n_2}$, and $\hat{q} = 1 - \hat{p}$

4. **Rejection region Significance-level-α test**

two-sided		one-sided greater		one-sided less
$\lvert z \rvert \geq z_{\frac{\alpha}{2}}$	or	$z > z_\alpha$	or	$z < -z_\alpha$

5. **p-value**

two-sided		one-sided greater		one-sided less
$2(1 - \Phi(\lvert z \rvert))$	or	$1 - \Phi(z)$	or	$\Phi(z)$

   where $\Phi(z)$ is the c.d.f. of a standard Normal distribution.

R function

Two-sided H_1:

```r
prop.test(c(x1, x2), c(n1, n2), correct = FALSE)
```

One-sided greater H_1:

```r
prop.test(c(x1, x2), c(n1, n2), correct = FALSE, alternative = "greater")
```

One-sided less H_1:

```r
prop.test(c(x1, x2), c(n1, n2), correct = FALSE, alternative = "less")
```

Example 1 - From 2013 exam

In an experiment to investigate whether there was significant evidence of an underlying difference between the mean measurements given by two types of caliper, Caliper 1 and Caliper 2, the diameter of a ball bearing was measured by 12 inspectors, each using the two types of caliper. The results are given in the table below.

Diameter of ball bearing in cm.

Inspector	Caliper 1	Caliper 2
1	0.265	0.264
2	0.265	0.265
3	0.266	0.264
4	0.267	0.266
5	0.267	0.267
6	0.265	0.268
7	0.267	0.264
8	0.267	0.265
9	0.265	0.265
10	0.268	0.267
11	0.268	0.268
12	0.265	0.269

1. State precisely the statistical model that is being used, defining carefully any notation that you use. Specify the null and alternative hypotheses in terms of the model parameters.

2. Write down a general formula for the test statistic that is used in the parametric test and state its distribution under the null hypothesis.

3. Draw conclusions in the present case.

Solution

Load the data

```
Caliper1 <- c(0.265, 0.265, 0.266, 0.267, 0.267, 0.265,
              0.267, 0.267, 0.265, 0.268, 0.268, 0.265)

Caliper2 <- c(0.264, 0.265, 0.264, 0.266, 0.267, 0.268,
              0.264, 0.265, 0.265, 0.267, 0.268, 0.269)

cbind(Caliper1, Caliper2)
```

1. The two samples are paired (not independent).

   We assume that the data comes from Normal distributions with means μ_X and μ_Y (unknown) in order to perform a two sample t-test for paired samples.

   The null hypothesis is that there is no difference between the mean of the two measurements: $H_O : \mu_X = \mu_Y$, while the alternative hypothesis is that there is not significant difference $H_1 : \mu_X \neq \mu_Y$.

2. The test statistic is,

$$t = \frac{\bar{x} - \bar{y}}{\frac{s_D}{\sqrt{n}}} = \frac{\bar{d}}{\frac{s_D}{\sqrt{n}}} \ ,$$

where $\bar{d}$ and s_D^2 are the sample mean and sample variance for $d_1, d_2, \ldots, d_n$. under H_0 it follows a t distribution with $n-1$ degrees of freedom.

In this case:

```
d <- Caliper1 - Caliper2
d

n <- length(d)
n

dbar <- mean(d)
dbar

sD <- sd(d)
sD
```

The test statistics is

```
t <- dbar / (sD / sqrt(n))
t
```

and the p-value:

```
pval <- 2 * (1 - pt(abs(t), df = n - 1))
pval
```

We can perform the test by using the function `t.test`:

```
test <- t.test(Caliper1, Caliper2,
               paired = TRUE)
test
```

let's check that this is the same as:

```
test2 <- t.test(d)
test2
```

3. Fill in the conclusions:

.

Example 2

A pharmaceutical company wishes to determine whether its new allergy product (A) is any better at reducing the level of a certain histamine in the blood stream than its current product (B). Two independent random samples of individuals were drawn from groups of people using product A and product B, respectively, and their histamine levels (in mg per cubic litre) were recorded. The data are given below.

Product A:	16.61	15.38	15.70	17.58	16.66	17.13
Product B:	18.66	19.52	16.98	18.19	17.20	

1. State carefully the statistical model that underlies an appropriate analysis, specifying in particular the unknown parameters.

2. Stating in terms of the model parameters the hypotheses that you are testing, write down a test statistic to investigate whether the mean level of the histamine for Product A is less than for Product B. Find the corresponding p-value and draw conclusions.