One sample and paired ttest in R

Dr. Osman

February 11, 2015

One sample hypothesis testing about the mean

One sample hypothesis testing about the mean

1. Set up the null hypothesis
   population mean = particular value
   $H_0 : \mu = \mu_0$

One sample hypothesis testing about the mean

1. Set up the null hypothesis
   population mean = particular value
   $H_0 : \mu = \mu_0$
2. Set up the alternative hypothesis
   $H_1 : \mu \neq \mu_0$

Example

Sample mean birth weight in 141 babies was 3.01 with s.e. 0.04.Suppose prior to the study it was thought population mean was 3.25 kg.What are the null and the alternative hypothesis?

Example

Sample mean birth weight in 141 babies was 3.01 with s.e. 0.04.Suppose prior to the study it was thought population mean was 3.25 kg.What are the null and the alternative hypothesis?

$H_0 : \mu = 3.25$ VS $H_1 : \mu \neq 3.25$

- Set the level of significance

  α usually given as 0.05.

- Compute the test statistic

$$t = \frac{\bar{x} - \mu_0}{se(\bar{x})}$$

where

$$se = \frac{s}{\sqrt{n}}$$

degrees of freedom=number of observations minus one.$(n - 1)$

Example

sample mean $= 3.01$
standard error (sample mean) $= 0.04$
population mean $= 3.25$

$$t = \frac{3.01 - 3.25}{0.04)} = -6.00$$

$df = n - 1 = 141 - 1 = 140$

Calculate appropriate p-value

To get P-value,

Use statistical tables

$P<0.001$

Calculate appropriate p-value in R

```
> pt(-6, df=141)
[1] 7.90706e-09
```

$P<0.001$

Interpretation

- Compare the p-value to the α level.
  Reject H_0 when P-value < 0.05
  Reject H_0 when P-value ≥ 0.05

P-value

- probability of observing our data assuming that the null hypothesis is true
- probability of observing our data given that the population mean birth weight is 3.25 kg

P<0.001

Conclusion

- unlikely that we would get our sample mean if the population mean birth weight was really 3.25 kg
- have found strong evidence to suggest that the population mean $\neq 3.25kg$

Paired comparisons

- When it is not feasible to assume that two groups of data are independent
- Used to compare means of the same population/subjects under different conditions
- Takes the correlation into account
- The differences between paired observations are assumed to be normally distributed
- More powerful since it reduces inter-subject variability

Examples of paired comparisons

- Pre- and post-test scores for a student receiving tutoring
- Fuel efficiency readings of two fuel types observed on the same automobile
- Sunburn scores for two sunblock lotions, one applied to the individual's right arm, one to the left arm
- Political attitude scores of husbands and wives

Worked Example

A stimulus is being examined to determine its effect on systolic blood pressure. Twelve men participate in the study. Their systolic blood pressure is measured both before and after the stimulus is applied.The values for BP before the stimulus are 20, 20, 21, 22, 23, 22, 27, 25, 27, 31,30 and 28.

The values for the BP after the stimulus are given as 19, 22, 24, 24, 25, 25, 26, 26, 28, 28,29 and 32.

Is there a change in BP before and after the stimulus was applied?

```
> bp_a <-c(20, 20, 21, 22, 23, 22, 27, 25, 27, 30, 31,30)
> bp_b <- c(19, 22, 24, 24, 25, 25, 26, 26, 28, 32,28,29)
>  t.test(bp_a,bp_b, paired=TRUE)
```

```
Paired t-test

data:  bp_a and bp_b
t = -1.5202, df = 11, p-value = 0.1567
alternative hypothesis: true difference in means is not equal
95 percent confidence interval:
 -2.0398769  0.3732102
sample estimates:
mean of the differences
             -0.8333333
```

Conclusion

The t test is not significant ($t=-1.09$, $p=0.1567$), indicating that the stimuli did not significantly affect systolic blood pressure.

Comparing group means: Two sample T-test

- Used when comparing means of two independent groups.
- E.g. Comparing if the average miles per gallon between two models of a car brand.

Assumptions:

- Assumes normal distribution, within each group, of the variable being compared
- The sampling distribution of the difference is also normally distributed
- Normally assumes equal variance between groups of the variable, but this assumption can be relaxed.

Calculating the standard error of the difference in means

1. Look at the distributions of the two groups.. Are the standard deviations similar? If so, we can calculate the pooled standard deviation.
2. Calculate the pooled standard deviation.

$$s = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{(n_1 - 1) + (n_2 - 1)}}$$

3. Estimate the standard error.

$$se = s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

Compute the test statistic as

$$t = \frac{\bar{x}_1 - \bar{x}_1 - \Delta}{se(\bar{x})}$$