Mathematics 50
Probability and Statistical Inference
Winter 1999

Statistical Hypothesis Testing

Week 9
March 1-5

We start with confidence interval construction because hypothesis testing and CI are closely related.

1. Exact confidence interval

Let x_i be iid from the normal distribution $N(\mu, \sigma^2)$ with unknown μ and σ^2. Find the exact CI for μ with coverage probability (confidence level) λ.

1.1. Approximate CI

If σ^2 is **known** then

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

so that

$$\frac{\overline{x} - \mu}{\sigma/\sqrt{n}} \sim N(0,1).$$

Thus, to find the exact CI which covers μ with probability λ we proceed as follows:

1. For standard normal variable $z \sim N(0,1)$ we find the critical value z_* such that

   $$\Pr(|z| < z_*) = \lambda.$$

   z_* is the $(1+\lambda)/2$ quantile.

2. CI is found from

   $$\left|\frac{\overline{x} - \mu}{\sigma/\sqrt{n}}\right| < z_*$$

   i.e. confidence limits are:

   $$\left(\overline{x} - \frac{\sigma}{\sqrt{n}} z_*, \overline{x} + \frac{\sigma}{\sqrt{n}} z_*\right).$$

This means that

$$\Pr\left(\overline{x} - \frac{\sigma}{\sqrt{n}} z_* < \mu < \overline{x} + \frac{\sigma}{\sqrt{n}} z_*\right) = \lambda.$$

For example, if $\lambda = .95$ then $z_* = z_{.975} = 1.96$ and CI:

$$\left(\overline{x} - \frac{1.96\sigma}{\sqrt{n}}, \overline{x} + \frac{1.96\sigma}{\sqrt{n}}\right).$$

But in real life we do not know σ^2. Approximation is used: estimate σ^2 as

$$\widehat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

and we use the same CI but with σ substituted by $\widehat{\sigma}$:

$$\left(\overline{x} - \frac{\widehat{\sigma}}{\sqrt{n}}z_*, \overline{x} + \frac{\widehat{\sigma}}{\sqrt{n}}z_*\right)$$

We cannot claim that the confidence level is exactly λ.

This is **two-sided** CI.

We can construct a **one-sided** CI.

1. For standard normal variable $z \sim N(0,1)$ we find critical value $z_* = z_\lambda$ such that

$$\Pr(z < z_*) = \lambda.$$

   z_* is the λ quantile of standard normal distribution.

2. CI is

$$\frac{\overline{x} - \mu}{\widehat{\sigma}/\sqrt{n}} < z_*$$

   This means that

$$\begin{aligned} \lambda &= \Pr\left(\frac{\overline{x} - \mu}{\widehat{\sigma}/\sqrt{n}} < z_*\right) \\ &= \Pr\left(\mu > \overline{x} - \frac{\widehat{\sigma}z_*}{\sqrt{n}}\right) \end{aligned}$$

   i.e. the lower confidence limits are:

$$\left(\overline{x} - \frac{\widehat{\sigma}}{\sqrt{n}}z_*, \infty\right).$$

In order to construct the exact CI when σ^2 is unknown more theory is needed..

2. χ^2-distribution

plays an important role in statistical hypothesis testing and CI construction.

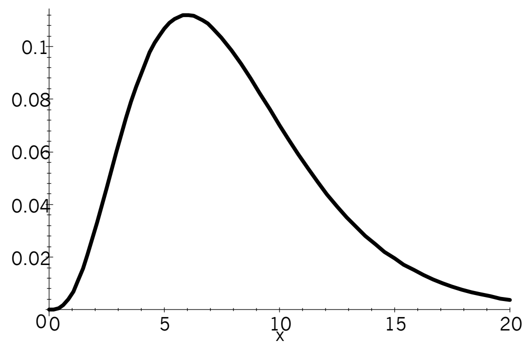
Definition 2.1. *χ^2-distribution with n degrees of freedom is the distribution of the sum of n iid standard normal RVs, $x_i \sim N(0,1)$,*

$$\chi^2 = \sum_{i=1}^{n} x_i^2.$$

The distribution of χ^2 is a special case of the Gamma distribution, the density of χ^2 has the form

$$Ax^{n/2-1}e^{-x/2}, \qquad x > 0$$

where A is a normalizing constant.



χ^2-distribution with 8 d.f.

Theorem 2.2. *Let x_i be iid, $x_i \sim N(\mu, \sigma^2)$. Then*

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \overline{x})^2 \sim \chi^2(n-1).$$

Also, sample mean and variance are independent:

$$\overline{x} \text{ and } \widehat{\sigma}^2 \text{ are independent RVs.}$$

Problem. If $x_i \sim N(\mu, 4), i = 1, ..., 4$ find the probability that the sum of squared distance from the sample mean is more then 10.

Solution. We need to find

$$\begin{aligned}
\Pr\left(\sum_{i=1}^{4}(x_i - \overline{x})^2 > 10\right) &= \Pr\left(\frac{1}{4}\sum_{i=1}^{4}(x_i - \overline{x})^2 > \frac{10}{4}\right) \\
&= \Pr\left(\chi^2(3) > 2.5\right) = 1 - \Pr\left(\chi^2(3) < 2.5\right) \\
&= .475
\end{aligned}$$

3. t-distribution

Definition 3.1. *t-distribution with n degrees of freedom is the distribution of the ratio*

$$t = \frac{x}{\sqrt{\sum_{i=1}^{n} y_i^2/n}}$$

where $x \sim N(0,1)$ and $y_i \sim N(0,1)$ are all independent.

The density of the t-distribution looks alike standard normal density especially for large d.f., n.

4. Exact confidence interval for the true mean of the normal distribution

Let x_i be iid from the normal distribution $N(\mu, \sigma^2)$ with unknown μ and σ^2. Find the exact CI for μ with coverage probability (confidence level) λ. Now σ^2 is **unknown** along with μ.

Calculate

$$\begin{aligned} \overline{x} &= \frac{1}{n}\sum_{i=1}^{n} x_i, \\ \widehat{\sigma}_x^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 \end{aligned}$$

and calculate the statistic

$$t = \frac{\overline{x} - \mu}{\widehat{\sigma}_x}.$$

Then we shall prove that $\sqrt{n}t$ has t - distribution with $n-1$ d.f.

Indeed,

$$\sqrt{n}t = \frac{(\overline{x} - \mu)/\sigma}{\sigma_x/(\sigma\sqrt{n})}$$

But, for the numerator

$$\sqrt{n}\frac{\overline{x} - \mu}{\sigma} \sim N(0,1)$$

because $\overline{x} \sim N(\mu, \sigma^2/n)$. For the denominator,

$$\begin{aligned} \sigma_x &= \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2/\sigma^2} \\ &= \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}((x_i - \mu)/\sigma - (\overline{x} - \mu)/\sigma)^2} \\ &= \sqrt{\frac{\chi^2(n-1)}{n-1}}. \end{aligned}$$

Thus, to find the exact CI which covers μ with probability λ when σ^2 is unknown we proceed as follows:

1. Find the critical value t_* such that

   $$\Pr(|t| < t_*) = \lambda.$$

   where t has $t-$distribution with $n-1$ d.f.

2. CI is

   $$\left|\frac{\overline{x} - \mu}{\widehat{\sigma}/\sqrt{n}}\right| < t_*$$

   i.e.

   $$\left(\overline{x} - \frac{\widehat{\sigma}}{\sqrt{n}}t_*, \overline{x} + \frac{\widehat{\sigma}}{\sqrt{n}}t_*\right).$$

For example, if $n = 6$ and $\lambda = .95$ then $t_* = t_{.975}(5) = 2.571$ and CI:

$$\left(\overline{x} - \frac{2.571\widehat{\sigma}}{\sqrt{n}}, \overline{x} + \frac{2.571\widehat{\sigma}}{\sqrt{n}}\right).$$

4.1. Comparison of exact and approximate CIs

Principally the exact CI have the same form as the approximate one but differ by the coefficient (critical value).

D.f. is crucial:

Critical values for $\lambda = .95$

d.f., $n-1$	Normal	t-distribution
	large sample appr.	exact, small sample
3	1.96	3.182
5	1.96	2.571
7	1.96	2.365
10	1.96	2.228
20	1.96	2.086
30	1.96	2.042
100	1.96	1.984
500	1.96	1.96

4.2. Exact CI for regression coefficients

If in simple linear regression, the distribution of the error term is normal and x_i are fixed then

$$t = \frac{b - \beta}{\widehat{\sigma}_b} \tag{4.1}$$

has t-distribution with $n - 2$ d.f., where b is the OLS estimate and

$$\widehat{\sigma}_b = \frac{s}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}}$$

is the SD of b, and $s = \widehat{\sigma}$,

$$s^2 = \frac{SS}{n-2}.$$

Therefore, the exact CI with the confidence level λ is

$$(b - t_*\widehat{\sigma}_b, b + t_*\widehat{\sigma}_b)$$

which covers the true β with probability λ and

$$\Pr(|t| < t_*) = \lambda$$

i.e. t_* is the $(1 + \lambda)/2$ quantile of the t-distribution with $n - 2$ d.f.

Statistic (4.1) is called **t-statistic.**

We say that the slope is **statistically significant** if CI does not cover 0 (the true value β cannot be 0 with high probability).

4.3. Regression prediction

One of the application of regression may be to predict the value of y given the value of x. After the intercept and slope are estimated we can predict y given x as follows

$$\widehat{y} = a + bx.$$

In particular, if $x = x_i$ we obtain the prediction of y at the observation point

$$\widehat{y}_i = a + bx_i, \quad i = 1, ..., n.$$

The difference, $e_i = y_i - \widehat{y}_i$ is called the OLS residual. y-hat is also a RV. If the error term in regression has normal distribution then y-hat has the normal distribution as well. Since a and b are unbiased estimators of α and β y-hat is also an unbiased estimator of y, i.e. $E(\widehat{y}) = \alpha + \beta x$. It is possible to prove that the variance of y-hat is given by

$$var(\widehat{y}) = \sigma^2 \left(1 + \frac{(x - \overline{x})^2}{\sum (x_i - \overline{x})^2}\right).$$

5. Hypothesis testing. Major concepts, Rice 9.

Problem set up: There are n iid observations $x_1, .., x_n$ drawn from a general population with a density $f(x;\theta)$. The statistical hypothesis is formulated in terms of the true parameter θ as

$$H_0 : \theta = \theta_0.$$

where θ_0 is a predefined value, e.g. $\theta_0 = 0$.

For example, if the general population is annual American family income and x_i is a sample of Vermont family incomes we may hypothesize that the population average of Vermont family income, μ is μ_0 =\$25,000 (national average, as announced by the federal government). We can test this hypothesis using data x_i.

H_0 is called the *null* hypothesis. The *alternative* hypothesis is

$$H_A : \theta \neq \theta_0$$

This is *simple* hypothesis. An example of *composite* hypothesis is $H_0 : \theta < \theta_0$ with alternative $H_A : \theta \geq \theta_0$. The alternative hypothesis $H_A : \theta \neq \theta_0$ is *two-sided*, the alternative hypothesis $H_A : \theta > \theta_0$ is *one-sided.*

We will deal with simple null hypotheses.

Continuing our example, the alternative might be that $\mu \neq \$25,000$ (two-sided) or $\mu < \$25,000$ (one-sided).

Statistical hypothesis is verified (tested) using a *test statistic*, t.

Test statistic is the engine of hypothesis testing.

For example, it is reasonable to reject H_0 if $|t| > c$ where c is a certain number and t is an estimator for θ, the number c is called *critical value*. Then, the region $|t| > c$ is called *rejection region* and $|t| \leq c$ is called *acceptance region*.

Example. A new drug to lose weight is suggested. How to verify whether the new drug is more effective than the old one. We collect data on the weight loss:

Difference in weight loss after one month treatment, L.B.
(11 volunteers)

Old drug-New drug
-10
+1
-3
-4
-14
-7
+2
-6
-23
-10
+4

How to test the hypothesis that the new drug is better than the old one? We make an assumption that the difference follows a normal distribution, i.e. $x_i \sim N(\mu, \sigma^2)$ are iid. We test the hypothesis

$$H_0 : \mu = 0.$$

with the alternative $H_A : \theta \neq 0$. A reasonable test statistic is $\overline{x}$. We say that drugs have different effect if $|\overline{x}|$ is big. More precisely, H_0 is rejected if $|\overline{x}| > c$ where c is the critical value. What is c? This is a big question. To find the critical value we have to do more investigation.

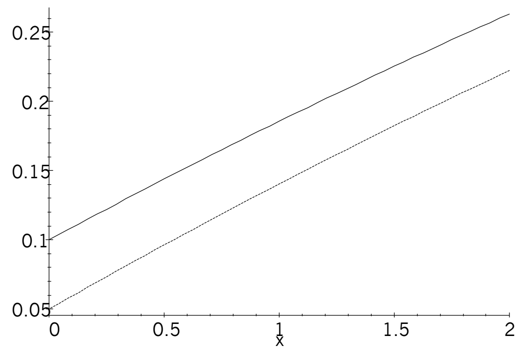
5.1. Two types of errors in hypothesis testing

Type 1 error: H_0 is true but we reject it. The probability of this error is denoted by α. It is called *significance level*.

Type 2 error: H_0 is accepted but in fact H_A is true, i.e. H_0 is wrong. This probability is denoted by β. The probability to reject H_0 when it is not true is $1 - \beta$, it is called *power*.

Typically, power,$1 - \beta$ is a function of alternative θ. Moreover, often power is an increasing function of θ.

Ideally, we should have $\alpha = 0$ and $\beta = 0$ (power=1). But is impossible to achieve this because power approaches α when alternative θ approaches θ_0.



Uncomparable statistical tests

We cannot minimize the two errors simultaneously.

5.2. How to find the critical value

Since it is impossible to minimize two types of errors simultaneously statisticians decided to fix type I error at certain possibly low level, α (say $\alpha = .05$). Thus, $\alpha = .05$ means that the probability to reject the null hypothesis when it is actually true is α. This defines the critical value of the test statistic in the following way.

We assume that the distribution of the test statistic t is known under the null, i.e. $\theta = \theta_0$. If the rejection rule is $|t| > c$, then we find c from the equation

$$\Pr(|t| > c|\theta = \theta_0) = \alpha,$$

by the definition. For testing the null hypothesis in our Example,

$$\frac{\overline{x} - \mu}{\widehat{\sigma}} \sim t(n-1)$$

so that

$$\frac{\overline{x}}{\widehat{\sigma}} \sim t(n-1)$$

and we choose c from

$$\Pr(|t\text{-dsirtibution with n-1 d.f.}|>\text{c}) = .05.$$

Finally, if

$$\frac{|\overline{x}|}{\widehat{\sigma}} > c$$

we reject H_0.

6. Confidence intervals and hypothesis testing

Let we want to test the null hypothesis

$$H_0 : \theta = 0$$

with the alternative

$$H_A : \theta \neq 0$$

under significance level α. We construct a $100(1-\alpha)\%$ CI for θ as

$$(\widehat{\theta}_1, \widehat{\theta}_2).$$

Then we say that H_0 is accepted if and only if the interval $(\widehat{\theta}_1, \widehat{\theta}_2)$ contains 0. Otherwise, we reject H_0. This rule provides significance level α for H_0 testing (see Rice for the proof). In particular, the best CI leads to the most powerful statistical test.

For example, if the 95% CI for the regression slope does not contains 0 then we say that the relationship between x and y is statistically significant with probability .95.

7. Hypothesis testing in regression. P-value

How to test that x and y are, in fact, correlated, i.e. y depends on x in linear regression

$$y_i = \alpha + \beta x_i + \varepsilon_i$$

As we can see y depends on x if and only if $\beta = 0$. Thus the question on the dependence can be reformulated in the language of hypothesis testing as

$$H_0 : \beta = 0.$$

If this hypothesis cannot be rejected based on the data we say that the relationship (and the slope) is *statistically significant.*

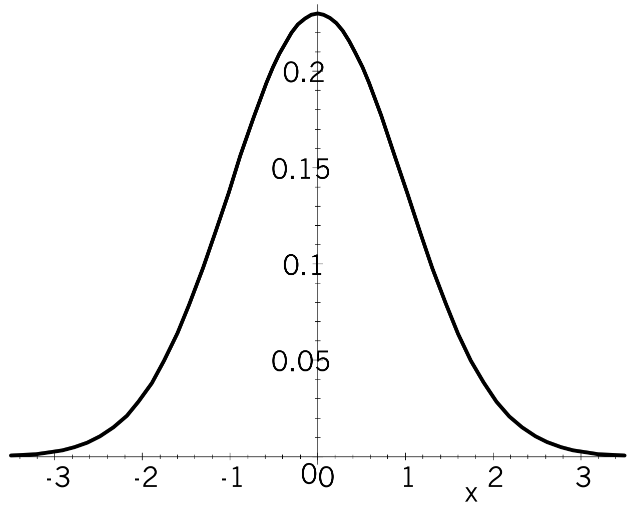
To test H_0 we use the fact that if $\beta = 0$ then

$$\frac{b}{s_b} \sim t(n-2),$$

so that the test statistic is

$$t = \frac{b}{s_b}$$

Larger absolute value of t indicates that the null hypothesis is not true, i.e. $\beta \neq 0$.



The density of the t-distribution

Therefore, to test the hypothesis $H_0 : \beta = 0$ we proceed as follows:

1. Determine the critical value t_* based on the t-distribution with $n-2$ d.f. and significance level α, e.g. $\alpha = .05$. For example, as follows from Rice Table 4 if $n = 10$ and $\alpha = .05$ the value t_* is such that $\Pr(|t_8| > t_*) = .05$ where 8=df. It corresponds to the value t_p such that $\Pr(t_8 < t_p) = (1 + .05)/2 = .975$, i.e. $t_p = t_{.975} = 2.306$.

2. Reject H_0 if $|t| > t_*$. In our example we reject that $\beta = 0$ if the t-statistic, $|b/s_b| > 2.306$.

In fact, once we calculated b/s_b we can calculate

$$\Pr(|t| > |b|/s_b \text{ under assumption } \beta = 0)$$

since t has known distribution. This probability is called $P-value$. We can test the null hypothesis using P-value as follows:

Let we use $\alpha = .05$ significance level.

If P-value is less then .05 then we reject $\beta = 0$ with 5% error.

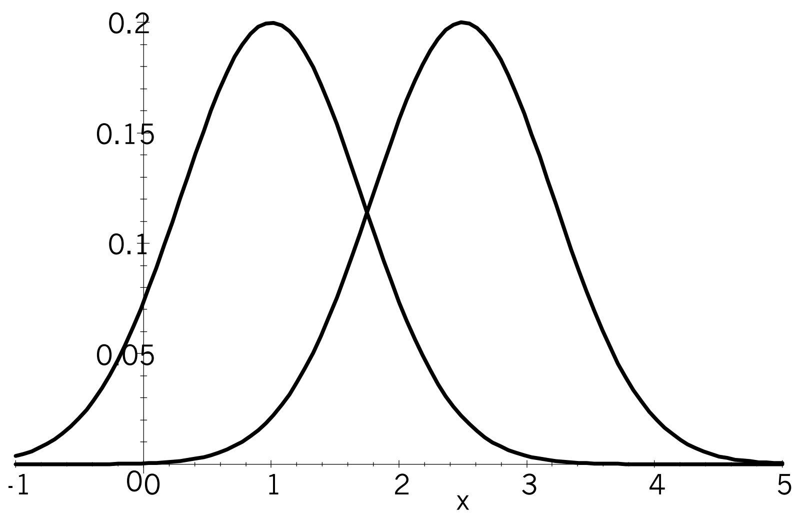
Thus, P-value is simpler to use because you do not need to know the critical value of t-statistic (you do not use any table at all).

P-value provides the empirical (data based) measure of the error to reject $H_0 : \beta = 0$ when it is true (type I error).

8. Comparing two samples

There are two possible situations when two groups (samples) are compared. In the first case observations from the first and the second group do not match and therefore may have different sample size. In the second case observations match and therefore the sample size in comparing groups must be the same.

These cases correspond to the following unpaired and paired t-test.



Two nornal densities with different population means

8.1. Unpaired t-test (Rice 11.2)

Let's imagine that in our weight-problem example now we have two groups of people. The first group used the old drug and consists of $n = 11$ volunteers. The second group used the new drug and consists of $m = 7$ people.

Weight loss after one month treatment, L.B.
(18 volunteers)

Old drug, X	New drug, Y
-10	-5
+1	-16
-3	+2
-4	-12
-14	-19
-7	-7
+2	+2
-6	
-23	
-10	
+4	

Assumptions:

1. All observations are independent, within group and between groups.
2. The first group represents the general population of all people could take the old drug, this distribution is normal $x \sim N(\mu_x, \sigma_x^2)$.
3. The second group represents the general population of all people could take the new drug with normal distribution $y \sim N(\mu_y, \sigma_y^2)$.
4. The variances in the two groups are the same: $\sigma_x^2 = \sigma_y^2 = \sigma^2$.

Question: is $\mu_x = \mu_y$. i.e. can the difference between $\overline{x}$ and $\overline{y}$ be explained only by sampling error, or

$$H_0 : \mu_x = \mu_y?$$

The algorithm is as follows (significance level=α):

1. Calculate **pooled sample variance**
$$s_p^2 = \frac{\sum_{i=1}^n (x_i - \overline{x})^2 + \sum_{i=1}^m (y_i - \overline{y})^2}{n + m - 2}.$$
2. Calculate test statistic
$$T = \frac{\overline{x} - \overline{y}}{s_p \sqrt{1/n + 1/m}} \sim t(n + m - 2) \quad \text{under } H_0.$$
3. Find the critical value t_* such that $\Pr(|t| > t_*) = \alpha$ where t has t-distribution with $n + m - 2$ d.f..
4. If $|T| > t_*$ then we reject H_0. Otherwise we accept it.

Example (continued). For the old drug we have

$$\overline{x} = (-10 + 1 - 3 - 4 - 14 - 7 + 2 - 6 - 23 - 10 + 4)/11 = -6.364$$

and

$$\sum_{i=1}^n (x_i - \overline{x})^2 = 610.55$$

For the new drug we have

$$\overline{y} = (-5 - 16 + 2 - 12 - 19 - 7 + 2)/7 = -7.8571$$

and

$$\sum_{i=1}^{n}(y_i - \overline{y})^2 = 410.86$$

Calculate the pooled sample variance

$$s_p^2 = \frac{610.55 + 410.86}{11 + 7 - 2} = 63.838$$

and

$$s_p = \sqrt{63.838} = 7.99.$$

Calculate T-statistic,

$$T = \frac{-6.364 + 7.8571}{7.99\sqrt{1/11 + 1/7}} = .3865$$

Find t_* using Table 4 with df=(11+7-2)=16 to obtain $t_* = 2.120$.

Conclusion: Since $.3865 < 2.120$ we cannot statistically prove that there is a difference between the two drugs. Probably, more volunteers is required.

8.2. Paired t-test (Rice 11.3)

The drawback of the above example is that volunteers have different 'starting' conditions: different age, sex, initial weight, etc. We can design experiment in such a way that two people taking different drug match, i.e. they may have the same initial weight, are of the same age and sex. Such a design is called **pair-matched** design.

Weight loss after one month treatment, L.B.
Pair-matched design by sex, age and initial weight
(7×2 volunteers)

Old drug, X	New drug, Y
-3	-5
-12	-16
1	-2
-4	-8
-4	-9
-1	2
-9	-11

Notice, the number of people in the two groups is the same.
Assumptions:

1. Pairs are independent but observations within pair may be dependent.

2. The first group represents the general population of all people could take the old drug, this distribution is normal $x \sim N(\mu_x, \sigma_x^2)$.

3. The second group represents the general population of all people could take the new drug with normal distribution $y \sim N(\mu_y, \sigma_y^2)$.

Question: is $\mu_x = \mu_y$. i.e. is the difference between $\overline{x}$ and $\overline{y}$ can be explained only by sampling error, or

$$H_0 : \mu_x = \mu_y?$$

We take the difference, $d_i = x_i - y_i$. Then,

$$d_i \sim N(\theta, \sigma^2)$$

are iid, where

$$\theta = \mu_x - \mu_y$$

with the null hypothesis

$$H_0 : \theta = 0.$$

The algorithm is as follows (significance level=α and the critical value of t-statistic is t_*):

1. Calculate sample mean and variance

$$\begin{aligned} \overline{d} &= \frac{1}{n}\sum_{i=1}^{n} d_i \\ \widehat{\sigma}^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(d_i - \overline{d})^2 \end{aligned}$$

2. Calculate test statistic

$$\frac{\overline{d}}{\widehat{\sigma}} \sim t(n-1).$$

3. Find the critical value t_* such that $\Pr(|t| > t_*) = \alpha$ where t has t-distribution with $n-1$ d.f.

4. If $|\overline{d}|/\widehat{\sigma} > t_*$ we reject H_0, otherwise we accept it.

What kind of t-test is appropriate in the Team Project?