Guide for performing Test of Significance and Set Estimation

S350

Arturo Valdivia

(Modified on October 19, 2021. Subject to change.)

1 Steps for performing a Test of Significance (Hypothesis testing)

1.1 Hypotheses

The null hypothesis, H_0, is what we initially assume to be true. It's a statement about a parameter (or many parameters). For example, a statement about μ. The alternative hypothesis, H_1, is what we would conclude (about the parameter) if we were to reject H_0.

When stating these hypotheses follow these guidelines:

- For procedural reasons, we **always** include the equal sign in the statement under the null hypothesis (H_0). For example, $H_0 : \mu = 10$ or $H_0 : \mu \geq 15$. This is needed as the test is performed assuming that H_0 is true.
- The statement under the alternative hypothesis (H_1) correspond to the statement that carries the burden of proof. In other words, this statement can be concluded only if we find evidence against the statement under the null hypothesis.
  - If when setting up our problem, two parties are in conflict, whichever party carries the burden of proof is the one that needs to come up with evidence for its claim. Accordingly, the statement desired by this party is the one presented in H_1. (See Example 9.3)
  - Researches when trying to come up with a relevant conclusion, are typically the ones who carry the burden of proof. Therefore, the desired statement for researchers should be presented under H_1.

1.1.1 Three alternatives

When the hypotheses statements are about μ, we have three possible tests

1. Left-tailed test
$$H_0 : \mu \geq \mu_0$$
$$H_1 : \mu < \mu_0$$

2. Rigth-tailed test
$$H_0 : \mu \leq \mu_0$$
$$H_1 : \mu > \mu_0$$

3. Two-tailed test (non-directional)
$$H_0 : \mu = \mu_0$$
$$H_1 : \mu \neq \mu_0$$

where μ_0 is some number. As illustration, in example 9.3a, a left-tailed test is presented,

$$H_0 : \mu \geq 15$$
$$H_1 : \mu < 15$$

1.2 Test Statistic

The statistic is the method we use to estimate the parameter(s). It is typically constructed using one or many estimators and it's constructed as to follow a known distribution. Many statistics exist, the one we encounter often looks like this:

$$statistic = \frac{estimator - parameter\ under\ H_0}{standard\ error}$$

When constructing a statistic based on the sample mean, $\bar{X}_n$, we need to determine if σ is known or not, as the distribution of the statistic changes.

1.2.1 When σ is known

When the hypothesis is made about μ, and σ is known, the test statistic is given by

$$Z = \frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}}$$

and due to the CLT, $Z \sim N(0, 1)$.

1.2.2 When σ is unknown

When the hypothesis is made about μ, and σ is unknown (as in most real-life problems), the test statistic is given by

$$T = \frac{\bar{X}_n - \mu_0}{S_n/\sqrt{n}}$$

where $S_n = \sqrt{\frac{\sum(X_i - \bar{X}_n)^2}{n-1}}$ is the sample standard deviation, another estimator. As shown in the lab, for the added uncertainty of S_n, T is no longer normal, but $T \sim T_{n-1}$, i.e. T follows a T-distribution with $n - 1$ degrees of freedom.

1.2.3 Observed test statistic

Once you collect a (random) sample of n observations, the sample observed is $\bar{x} = (x_1, \ldots, x_n)$. If σ is known, the observed test statistic is

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

and if *sigma* is unknown, the observed test statistic is

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

where

$$s = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n - 1}}$$

is the sample standard deviation.

1.3 Significance Probability (p-value)

The p-value is the probability of observing a test statistic that is as extreme or more extreme than the one observed in our data, when we assume that H_0 is true. The p-value depends on the hypotheses statements made (see Step 1).

In particular, when making inferences about μ, we have:

1.3.1 When σ is known:

1. Left-tailed test: $p\text{-value} = P(Z \leq z)$

In R: `pnorm(z)`

2. Rigth-tailed test: $p\text{-value} = P(Z \geq z)$

In R: `1 - pnorm(z)`

3. Two-tailed test (non-directional): $p\text{-value} = P(Z \geq |z|)$

In R, `2*(1 - pnorm(abs(z)))`

1.3.2 When σ is unknown

1. Left-tailed test: $p\text{-value} = P(T \leq t)$

In R: `pt(t, df = n-1)`

2. Rigth-tailed test: $p\text{-value} = P(T \geq t)$

In R: `1 - pt(t, df = n-1)`

3. Two-tailed test (non-directional): $p\text{-value} = P(T \geq |t|)$

In R, `2*(1 - pt(abs(t), df = n-1))`

1.4 Conclusion and Interpretation of results

We want to reject H_0 if the data (sample) obtained produces an estimate that is highly unlikely to be obtained by chance if H_0 is true. The smaller the p-value, the more evidence to reject H_0. While how small p-value needs to be is somewhat arbitrary, it should be guided by how important is not to make a mistake by rejecting H_0 when it is actually true (Type I Error) or by failing to reject H_0 when it is actually false (Type II Error).

To guide the decision process, some reasonable values (although still arbitrary) can be:

- If p-value > 0.1, do not reject H_0.
- If p-value < 0.001, reject H_0.
- If $0.001 \leq p\text{-value} \leq 0.1$, decide based on your own perception of this uncertainty (different people may make different decisions).
  - An alternative is to come up with a significance level, α, such that ff p-value $\leq \alpha$, we reject H_0, and if p-value $> \alpha$, we fail to reject H_0.
  - The choice of α should be provided before collecting and/or observing the sample.

2 Set Estimation (Confidence Intervals)

2.1 Two-tailed confidence intervals

Set estimation is an inferential method alternative and/or complementary to test of significance. Instead of using the information of our sample to reject a claim (or not), we use it to obtain a set of plausible values for the parameter of interest (such as μ). In that sense, a set estimate, or a confidence interval, is is an **interval estimate** of the parameter of interest. (contrast this with a **point estimate**). For many problem we'll encounter, a confidence interval looks like this:

$$estimator \pm margin\ of\ error$$

where

$$margin\ of\ error = critical\ value \times standard\ error$$

where the critical value is constructed based on the level of confidence we would like to attach to our interval. The level of confidence is a number close to 1 (100%) and it represents how confident we would like to be on our method such that the interval constructed would in fact contain the parameter of interest. In this context, to obtain an interval with a $(1-\alpha) \times 100\%$ level of confidence for μ this formula becomes:

2.1.1 When σ is known (two-tailed confidence interval)

$$\bar{x} \pm q \times \frac{\sigma}{\sqrt{n}}$$

where q is the $(1 - \frac{\alpha}{2})$-quantile. In R, `q = qnorm(1-alpha/2)`.

2.1.2 When σ is unknown (two-tailed confidence interval)

$$\bar{x} \pm q \times \frac{s}{\sqrt{n}}$$

where q is the $(1 - \frac{\alpha}{2})$-quantile. In R, `q = qt(1-alpha/2, n-1)`.

2.2 Sample Size

If the length of a confidence interval is given by L, then it is easy to see that

$$L = 2 \times q\frac{\sigma}{\sqrt{n}}$$

solving for n we get

$$n = \left(\frac{2q\sigma}{L}\right)^2$$

It's easier to simply use the case where σ is knwon, even if you use an estimate for this purpose, and use `qnorm()` to obtain the appropriate value of q. (See the textbook for more details).

2.3 One-tailed confidence intervals (Not needed for Exam 2)

2.3.1 Left-tailed

- When σ is known

$$\left(-\infty\ , \quad \bar{x} + q \times \frac{\sigma}{\sqrt{n}}\right)$$

where q is the $(1-\alpha)$-quantile. In R, `q = qnorm(1-alpha)`.

- When σ is unknown

$$\left(-\infty\ , \quad \bar{x} + q \times \frac{s}{\sqrt{n}}\right)$$

where q is the $(1-\alpha)$-quantile. In R, `q = qt(1-alpha, n-1)`.

2.3.2 Rigth-tailed

- When σ is known,

$$\left(\bar{x} - q \times \frac{\sigma}{\sqrt{n}}\ , \quad \infty\right)$$

where q is the $(1-\alpha)$-quantile. In R, `q = qnorm(1-alpha)`.

- When σ is unknown,

$$\left(\bar{x} - q \times \frac{\sigma}{\sqrt{n}}\ , \quad \infty\right)$$

where q is the $(1-\alpha)$-quantile. In R, `q = qt(1-alpha, n-1)`.