MVE051/MSG810 2017 Lecture 9

Petter Mostad

Chalmers

November 27, 2017

Test for expected value of normal distribution

- Assume $X_1, \dots, X_n$ is a random sample from Normal(μ, σ) with μ, σ unknown. Assume we want to compare $H_0 : \mu = \mu_0$ with $H_1 : \mu \neq \mu_0$ for some fixed μ_0.
- We choose as test statistic

$$T = \frac{\overline{X} - \mu_0}{S/\sqrt{n}}$$

  which has a t distribution with $n-1$ degrees of freedom when H_0 is true.
- The rejection region is all T so that $T < -T_0$ or $T > T_0$ for some T_0. To make the significance become α, we must choose

$$T_0 = t_{\alpha/2}$$

  where $t_{\alpha/2}$ is so that $\Pr\left[T > t_{\alpha/2}\right] = \alpha/2$ when T has a t distribution with $n-1$ degrees of freedom.
- Finally we compute our value for T, compare with T_0, and decide to reject H_0 or not based on this.

One-sided and two-sided tests

- The test in the example above is a *two-sided* test. An alternative is a *one-sided* test, where for example $H_0 : \mu \leq \mu_0$ and $H_1 : \mu > \mu_0$.
- The test statistic is the same, but the rejection region becomes all T so that $T > T_0$, where

$$T_0 = t_\alpha$$

  where t_α is so that $\Pr[T < t_\alpha] = \alpha$ when T has a t distribution with $n - 1$ degrees of freedom.
- We use $\mu = \mu_0$ to compute the significance.
- Correspondingly one can construct a one-sided test with $H_0 : \mu \geq \mu_0$ and $H_1 : \mu < \mu_0$.

Hypothesis testing

1. Two models are established: The *Null hypothesis* H_0 and the *alternative* hypothesis H_1. (H_1 is often what one "wants to show statistically").
2. A *test statistic* T (i.e., a function of a random sample) is established, so that
   - The distribution of the test statistic T can be computed when H_0 is true.
   - The test statistic has one type of values (often small) when H_0 is true and generally another type of values (often large) when H_1 is true.
3. A *rejection region* F is established (generally one or more intervals) and one decides to *reject* H_0 if T is in F while H_0 is *not rejected* if T is not in F.
4. T is computed from observed data, compared with F, and rejected or not based on this.

Properties of hypothesis tests

- Type I and type II errors.
- We assume we can find the distribution of the test statistic T when we assume H_0 is true: Thus we can compute the probability of Type I errors before data is observed. This probability is often denoted α, and called the *significance* of the test. We often choose the rejection region so that $\alpha = 0.05$.
- Similarly, we write β for the probability for Type II errors. This probability cannot always be computed as easily, without further specifying H_1. The *strength* of the test is $1 - \beta$.
- One tries to choose the test statistic maximizing the test strength while the significance is fixed (often at $\alpha = 0.05$).

Interpreting a hypothesis test

- Correct interpretation: If you use the hypothesis test as a decision rule, then in a (long) series of such decisions, you will reject a proportion α of the correct H_0 hypotheses.
- Note: A hypothesis test does not give you the probability that H_0 or H_1 is true.
- Note: If you reject H_0, it does not mean that H_1 is true.
- Note: If you do not reject H_0, it does not mean it is proven that H_0 is true!
- Note: Wether you reject or not will not only depend on the data, H_0, and H_1, but also on the choice of test statistic.

Significance testing

- Significance testing represents a further development of the ideas of hypothesis testing: Instead of first deciding a significance level α, we compute the value of the test statistic on the data and then the *smallest significance level* α *which would make it possible to reject* H_0 *with this test statistic*.
- This smallest significance level is called the *p value* of the test.
- Doing the previous hypothesis test with a significance level of 0.05 corresponds to first computing the p-value and then rejecting H_0 if the p value is 0.05 or less.

Example: Significance testing for the expectation of a normal distribution

- Assume $X_1, \dots, X_n$ is a random sample from Normal(μ, σ) with μ, σ unknown. Assume you would like to compare $H_0 : \mu = \mu_0$ with $H_1 : \mu \neq \mu_0$ for some known μ_0.
- We choose the same test statistic as before:

$$T = \frac{\overline{X} - \mu_0}{S/\sqrt{n}}$$

  which has a t distribution with $n-1$ degrees of freedom when H_0 is true.
- We compute the value of T for our data, find $t_{\alpha/2}$ so that $T = -t_{\alpha/2}$ (if $T < 0$) or $T = t_{\alpha/2}$ (if $T > 0$), and use the table for the t-distibution with $n-1$ degrees of freedom to compute α, which becomes the p value.
- In a corresponding way we can compute the p value for one-sided tests.

Interpretation of the p value

- A p-value will tell you something about how "extreme" your data is in the direction of indicating that H_1 is true, compared to random variations in data expected under H_0.
- It is the probability of observing the observed T or "something more extreme" in the direction of H_1, if H_0 is true.
- Some wrong interpretations:
  - The p value does NOT give you the probability that H_0 is true.
  - The p value cannot be directly related to the probability that H_1 is true.
- Remember that the p value may depend on the choice of test statistic, and not only on the data and the hypotheses.

Confidence interval for a proportion when n is large

- If $X_1, \dots, X_n$ is a random sample where each X_i is 1 with probability p and 0 otherwise, if $z_{\alpha/2}$ is so that $[-z_{\alpha/2}, z_{\alpha/2}]$ contains $100(1-\alpha)\%$ of the probability in a standard normal distribution, and if we define $\hat{p} = \overline{X}$ and

$$\begin{aligned} L_1 &= \hat{p} - z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n} \\ L_2 &= \hat{p} + z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n} \end{aligned}$$

  then $[L_1, L_2]$ is an *approximate* $100(1-\alpha)\%$ confidence interval for p when n is large.
- Proof: When n is large we have approximately $\overline{X} \sim \text{Normal}(p, \sqrt{p(1-p)/n})$. When n is big we can even approximate $p(1-p)$ with $\hat{p}(1-\hat{p})$.

Estimation of sample size

- Given a confidence interval where L_1 and L_2 depend on n, we can ask: How large does n need to be for the confidence interval to be shorter than some given number?
- Example 1: We want to estimate a proportion p and we have a guess $\hat{p}$. How large sample do we need for the length of the confidence interval to be maxumum $2d$?

$$n \geq z_{\alpha/2}^2 \frac{\hat{p}(1-\hat{p})}{d^2}$$

- Example 2: We want to estimate a proportion p without using a guess. How large sample do we need for the length of the confidence interval to be maximum $2d$?

$$n \geq z_{\alpha/2}^2 \frac{1}{4d^2}$$

This is based on that we always have $\hat{p}(1-\hat{p}) \leq 1/4$ (as $0 \leq \hat{p} \leq 1$).

Hypothesis test for proportion, with n large

- (Example for comparison; not on exam)
- Assume $X_1, \dots, X_n$ is a random sample with X_i equal to 1 with probability p and otherwise 0. Assume we want to compare $H_0 : p = p_0$ with $H_1 : p \neq p_0$ for some known p_0.
- We write $\hat{p} = \overline{X}$ and choose as test statistic

$$\frac{\hat{p} - p_0}{\sqrt{p_0(1 - p_o)/n}}$$

  which approximately has a standard normal distribution when H_0 is true and n is large.
- The rejection region consists of all values outside $[-z_{\alpha/2}, z_{\alpha/2}]$, where $z_{\alpha/2}$ is so that $\Pr\left[Z > z_{\alpha/2}\right] = \alpha/2$ when Z has a standard normal distribution.
- Correspondingly, one can do one-sided tests where we use z_α instead of $z_{\alpha/2}$, and significance tests, where we compute a p-value.

Confidence interval for difference between proportions when n is large

- If $X_1, \dots, X_{n_1}$ och $Y_1, \dots, Y_{n_2}$ are random samples where each X_i is 1 with probability p_1 and 0 otherwise, and correspondingly for Y_i and p_2, if $z_{\alpha/2}$ is so that $[-z_{\alpha/2}, z_{\alpha/2}]$ contains $100(1-\alpha)\%$ of the probability in a standard normal distribution, and if $\hat{p}_1 = \overline{X}$, $\hat{p}_2 = \overline{Y}$ and

$$\begin{aligned} L_1 &= \hat{p}_1 - \hat{p}_2 - z_{\alpha/2}\sqrt{\hat{p}_1(1-\hat{p}_1)/n_1 + \hat{p}_2(1-\hat{p}_2)/n_2} \\ L_2 &= \hat{p}_1 - \hat{p}_2 + z_{\alpha/2}\sqrt{\hat{p}_1(1-\hat{p}_1)/n_1 + \hat{p}_2(1-\hat{p}_2)/n_2} \end{aligned}$$

  then $[L_1, L_2]$ is an *approximative* $100(1-\alpha)\%$ confidence interval for $p_1 - p_2$ when n is large.
- Proof: When n is large we have, approximately, $\overline{X} - \overline{Y} \sim \text{Normal}(p_1 - p_2, \sqrt{p_1(1-p_1)/n_1 + p_2(1-p_2)/n_2})$. When n is large we can even approximate $p_1(1-p_1)$ with $\hat{p}_1(1-\hat{p}_1)$ and $p_2(1-p_2)$ with $\hat{p}_2(1-\hat{p}_2)$.