Mathematical statistics
Week 5/b: Concept of hypothesis testing

Péter Elek and Ádám Reiff

October 17, 2013

Examples.

- H_0 : A particular drug is not effective in reducing the blood pressure.
- H_1 : The drug is effective in reducing the blood pressure.

- H_0 : Enterprise zones do not have an effect on buisiness investment.
- H_1 : Enterprise zones increase business investment.

- H_0 : The defendant is innocent.
- H_1 : The defendant is guilty.

Process of hypothesis testing.

- H_0 is called *null hypothesis*.
- H_1 is called *alternative hypothesis*.
- Suppose we have a a sample $X_1, X_2, \ldots, X_n$. Our decision is based on a sample statistic $T(X_1, X_2, \ldots, X_n)$, which is called the *test statistic*.
- Our decision:
  - We *reject* H_0 in favor of H_1 if T is in the *rejection region* (e.g.$T > c$, where c is the *critical value*).
  - Otherwise we *do not reject* H_0.

Errors and power of a test.

- Probability of *Type I error*: $\Pr(\text{reject } H_0 | H_0 \text{ is true})$
  - i.e. we reject H_0 when it is in fact true
  - we denote the probability of this by α
- Probability of *Type II error*: $\Pr(\text{do not reject } H_0 | H_0 \text{ is false})$
  - i.e. we do not reject H_0 when it is in fact false
  - we denote the probability of this by β
- *Power of a test*: $\pi = \Pr(\text{reject } H_0 | H_0 \text{ is false}) = 1 - \beta$
  - i.e. we reject H_0 when it is false

Tradeoff between α and β.

- There is a tradeoff between α and β.
  - If we reduce α, we have to allow more β.
  - A proper solution would be to choose α and β based on the relative costs of the two types of error. However, this is not the route followed in classical statistics.
- Philosophy in classical statistics: we only reject H_0 if there is "overwhelming" evidence against it (i.e. we fix and keep α low).
  - (We try to find those tests, which have the smallest β for a given α.)
  - Typical values for α: 0.1, 0.05, 0.01.
  - If we do not reject H_0, we should omit the phrase "we accept H_0".
- The fixed α is called the *significance level* or *size* of the test.

Example.

- H_0 : A particular drug does not change the blood pressure ($\mu_0 = 0$)
- H_1 : It reduces the blood pressure ($\mu_0 < 0$)
- We have a random sample of $n = 25$ and obtained an average change of blood pressure $\overline{x} = -12$. (For simplicity assume that the sample is normally distributed and $\sigma = 20$ is known.)
- Is this enough evidence that the drug decreases blood pressure?

Solution.

- We know: if H_0 is true, then $\overline{X} \sim N\left(\mu_0, \frac{\sigma^2}{n}\right) \sim N\left(0, \frac{20^2}{25}\right)$.
- Set $\alpha = 0.05$. This is the probability of rejection of H_0 (in favor of H_1) if it is in fact true (i.e. if $\overline{X} \sim N\left(0, \frac{20^2}{25}\right)$).
- Draw a figure about $N\left(0, \frac{20^2}{25}\right)$, and mark the bottom 5%.
- If H_0 is true:
  - $\Pr\left(\frac{\overline{X}-0}{20/\sqrt{25}} < z_{0.05}\right) = 0.05$,
  - so $\Pr\left(\overline{X} < 0 - 1.645\frac{20}{\sqrt{25}}\right) = \Pr\left(\overline{X} < -6.58\right) = 0.05$.
- Our *decision rule*: we reject H_0 if $\overline{x} < -6.58$.
  - -6.58 is the critical value.
  - Then the probability of type I error is exactly 5%.
- For us, $\overline{x} = -12$, so we reject H_0.

Type II errors and the power of this test.

- Probability of type II error (not rejecting H_0 if it is false) depends on the true value of μ.
- If the true value is μ, then
  - $\overline{X} \sim N\left(\mu, \frac{20^2}{25}\right)$,
  - so $\beta(\mu) = \Pr\left(\overline{X} > -6.58\right) = \Pr\left(\frac{\overline{X}-\mu}{20/\sqrt{25}} > \frac{-6.58-\mu}{20/\sqrt{25}}\right) = 1-\Phi\left(\frac{-6.58-\mu}{4}\right)$.
- Values of β for different true values of μ :
  - If $\mu = -10$, then $\beta = 1 - \Phi(0.855) = 0.196$.
  - If $\mu = -5$, then $\beta = 1 - \Phi(-0.395) = 0.654$.
  - If $\mu = -2$, then $\beta = 1 - \Phi(-1.145) = 0.874$.
  - If $\mu = -1$, then $\beta = 1 - \Phi(-1.395) = 0.918$.
  - Hence $\lim_{\mu\to 0} \beta(\mu) = 1 - \alpha = 0.95$,
  - and $\lim_{\mu\to -\infty} \beta(\mu) = 0$.
- $\pi(\mu) = 1 - \beta(\mu)$ is called the *power function* of the test.

One-tailed and two-tailed tests.

- Types of alternative hypotheses:
  - Two-sided: $H_1 : \mu \neq \mu_0$
  - One-sided: $H_1 : \mu < \mu_0$
  - One-sided: $H_1 : \mu > \mu_0$
- Typical rejection regions of the tests:
  - Two-tailed test: $|T| > c$ for $H_1 : \mu \neq \mu_0$
  - One-tailed test: $T < c$ for $H_1 : \mu < \mu_0$
  - One-tailed test: $T > c$ for $H_1 : \mu > \mu_0$
- Unless you have a good reason to do otherwise, use a two-sided alternative and hence a two-tailed test.

Example (cont.).

- In the blood pressure example, let $H_1 : \mu \neq 0$. Construct a test with $\alpha = 0.05$.
- Solution:
  - Under H_0 we still have $\overline{X} \sim N\left(0, \frac{20^2}{25}\right)$,
  - but now we need upper and lower thresholds: $\Pr\left(z_{0.025} < \frac{\overline{X}-0}{20/\sqrt{50}} < z_{0.975}\right) = 0.95$
  - or $\Pr\left(-z_{0.975}\frac{20}{\sqrt{50}} < \overline{X} < z_{0.975}\frac{20}{\sqrt{50}}\right) = 0.95$
  - and hence $\Pr\left(-7.84 < \overline{X} < 7.84\right) = 0.95$.
  - So our decision rule: we reject H_0 if $|\overline{x} - 0| > 7.84$.
  - The probability of type I error will be $\alpha = 0.05$.

p-values.

- *p-value*: the largest significance level at which we could carry out the test and still fail to reject H_0
- Hence, we reject H_0 at all significance levels larger than the p-value, and do not reject H_0 for levels smaller than the p-value.
- Back to the example:
  - In the one-sided case: If H_0 is true, we would have $z = \frac{\overline{x}-0}{20/\sqrt{25}} = -3 = z_{0.0013}$. Hence $p = 0.0013$ is the p-value of the test.
  - In the two-sided case the p-value is $2 * 0.0013 = 0.0026$.
  - Hence our results are significant even at the 1% level.