Mathematical statistics
Week 6/a: Commonly used hypothesis tests

Péter Elek and Ádám Reiff

October 22, 2013

1. Steps of hypothesis testing

Steps of hypothesis testing.

1. Formulate H_0 and H_1.
2. Find an appropriate test statistic with a known distribution under H_0.
3. Choose a significance level α, the fixed probability of type I error (generally 0.01, 0.05 or 0.1).
4. Calculate the critical value of the test statistic.
5. Calculate the value of the test statistic in the sample.
6. Decide: reject or do not reject H_0.
7. (Give the p-value of the test.)

2. Commonly used hypothesis tests

2.1. Tests of population mean

Large sample test of population mean.

- $H_0 : \mu = \mu_0$
- Test statistic: $z = \sqrt{n}\frac{\overline{x}-\mu_0}{s^*}$
- Under H_0, $z \overset{A}{\sim} N(0,1)$.
- Two-sided test: $H_0 : \mu = \mu_0$ and $H_1 : \mu \neq \mu_0$
  - We reject H_0 at significance level α if $|z| > z_{1-\alpha/2}$.

- The p-value of the test is $p = 2 * (1 - \Phi(|z|))$.

- One-sided test with $H_0 : \mu = \mu_0$ and $H_1 : \mu > \mu_0$
  - We reject H_0 at significance level α if $z > z_{1-\alpha}$.
  - The p-value of the test is $p = 1 - \Phi(z)$.
- One-sided test with $H_0 : \mu = \mu_0$ and $H_1 : \mu < \mu_0$
  - We reject H_0 at significance level α if $z < z_{\alpha}$.
  - The p-value of the test is $p = \Phi(z)$.
- Rule of thumb for large sample approximation: $n > 60$ or $n > 120$ (although it is not a theorem)

Small sample test of population mean in normal samples.

- Assume that X is normally distributed with unknown mean and variance.
- $H_0 : \mu = \mu_0$
- Test statistic: $t = \sqrt{n}\frac{\overline{x}-\mu_0}{s^*}$
- Under H_0, $t \sim t_{n-1}$.
- Two-sided test: $H_0 : \mu = \mu_0$ and $H_1 : \mu \neq \mu_0$
  - We reject H_0 at significance level α if $|t| > t_{n-1;1-\alpha/2}$.
- One-sided test with $H_0 : \mu = \mu_0$ and $H_1 : \mu > \mu_0$
  - We reject H_0 at significance level α if $t > t_{n-1;1-\alpha}$.
- One-sided test with $H_0 : \mu = \mu_0$ and $H_1 : \mu < \mu_0$
  - We reject H_0 at significance level α if $t < t_{n-1;\alpha}$.
- If σ is known, we could use a normal test instead of a t-test. But this is rarely done in practice.

Note: relationship between hypothesis tests and confidence intervals.

- Suppose we have constructed an $1-\alpha$ confidence interval for the population mean (by either of the methods above).
- If μ_0 does not lie within this interval, then $H_0 : \mu = \mu_0$ is rejected against a two-sided alternative at significance level α.
- If μ_0 lies within this interval, then $H_0 : \mu = \mu_0$ is not rejected against a two-sided alternative at significance level α.

Example 1.

- The targeted pH-value of drinking water is 8. We have 17 sample elements, with a sample mean of 7.92 and a corrected sample standard deviation 0.16.
- Is there enough evidence at the 5% level that the pH-value deviated from the target value?
- (Assume that pH-measurements follow a normal distribution.)

Solution.

- $H_0 : \mu = 8$ and $H_1 : \mu \neq 8$
- Under H_0: $\sqrt{n}\frac{\overline{x}-8}{s^*} \sim t_{n-1}$.
- So our test statistic is $\frac{\overline{x}-8}{s^*/\sqrt{n}} = -2.05$, and the critical values are $t_{16;0.025} = -2.12$ and $t_{16;0.975} = 2.12$.
- Therefore we do not reject H_0 (marginally though) at 5% level. We would have rejected it at 10%.

Example 2: paired samples.

- We would like to test the effect of a new drug on blood pressure. We have a sample of 100 patients, and their average blood pressure before taking the drug was 155.2, while their average blood pressure after taking the drug was 137.2. The (corrected) standard deviation of the *change* in the blood pressure of the patients was 25.
- Is there enough evidence that the drug changes the blood pressure?

Solution.

- $H_0 : \mu = 0$ and $H_1 : \mu \neq 0$
- If we denote the *changes* by $x_1, \ldots, x_{100}$, we have $\overline{x} = 137.2 - 155.2 = -18$ and $s^* = 25$.
- Test statistic: $z = \sqrt{100} * \frac{-18}{25} = -7.2$, and hence the p-value is $p = 2 * (1 - \Phi(7.2)) = 0.0000$, so there is evidence for the change of the blood pressure at all conventional levels.

2.2. Two-sample test of population means

Testing the equality of means of two i.i.d. samples.

- Suppose we have two i.i.d. samples: $x_1, \ldots, x_{n_X}$ with population mean μ_X and population variance σ_X^2, and $y_1, \ldots, y_{n_Y}$ with population mean μ_Y and population variance σ_Y^2. Suppose also that both are independent of each other as well.
- Based on observing $\overline{x}$, s_x, $\overline{y}$ and s_y, construct a *large sample test* of the equality of the two means.

Two-sample test (cont.).

- $H_0 : \mu_X = \mu_Y$ and $H_1 : \mu_X \neq \mu_Y$.
- It can be shown (how?) that $\left(\overline{Y} - \overline{X}\right) \overset{A}{\sim} N\left((\mu_Y - \mu_X), \left(\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}\right)\right)$.
- In large samples σ_X can be replaced by s_X and σ_Y by s_y (why?).
- Hence the test statistic is: $z = \frac{\overline{y} - \overline{x}}{\sqrt{\frac{s_x^2}{n_X} + \frac{s_y^2}{n_Y}}} \overset{A}{\sim} N(0, 1)$
- and so a z-test can be performed in large samples.
- Note: There is also an exact small sample t-test in normal populations, where $\sigma_X = \sigma_Y$.

Example.

- Suppose that in 1970 the average height of 100 Stanford male students was 6 feet with a (corrected) standard deviation of 0.4 foot, while in 1990 the average height of 120 students was 6.1 with a standard deviation of 0.3 foot.
- Should we conclude that the mean height of Stanford male students changed in this period?

Solution.

- $H_0 : \mu_X = \mu_Y$ and $H_1 : \mu_X \neq \mu_Y$
- Large sample test statistic: $z = \frac{6.1 - 6}{\sqrt{\frac{0.4^2}{100} + \frac{0.3^2}{120}}} = 2.06$.
- Hence the p-value is $p = 2 * (1 - \Phi(2.06)) = 0.039$.
- So there is evidence of the change at the 5% level but not at the 1% level.

2.3. Large-sample test of population proportion

Large sample test of population proportion: example.

- p is the probability that the stock price increases after a large insider purchase.
- We want to test at the 10% level whether $p = 0.5$ or $p \neq 0.5$.
- In our sample, the stock price increased 327 times of 576 insider purchases .

Solution.

- $H_0 : p = 0.5$ and $H_1 : p \neq 0.5$
- Under $H_0 : \frac{\overline{x}-0.5}{\sqrt{0.5(1-0.5)/n}} \overset{A}{\sim} N(0,1)$ and also $\frac{\overline{x}-0.5}{\sqrt{\overline{x}(1-\overline{x})/n}} \overset{A}{\sim} N(0,1)$.
- We reject H_0 if $-z_{0.95} < \frac{\overline{x}-0.5}{\sqrt{0.5(1-0.5)/n}} < z_{0.95}$.
- In the sample $\overline{x} = \frac{327}{576} = 0.568$ and $n = 576$.
- So our test statistics is $\frac{\overline{x}-0.5}{\sqrt{0.5(1-0.5)/n}} = 3.264$, and the critical value is $z_{0.95} = 1.645$.
- Therefore we reject H_0 and the p-value is smaller than 0.001.

2.4. Test of variance in normal populations

Test of variance in normal populations: example.

- A filling machine of margarine should have $\sigma = 4$ grams.
- In a sample of $n = 10$, we have $s^* = 4.508$.
- Test at the 5% level whether the machine meets the standards.
- (Assume normality)

Solution.

- $H_0 : \sigma^2 = \sigma_0^2 = 16$ and $H_1 : \sigma^2 \neq 16$
- Under H_0: $\frac{(n-1)s^{*2}}{\sigma_0^2} \sim \chi^2_{n-1}$.
- We do not reject H_0 if $\chi^2_{n-1;0.025} < \frac{(n-1)s^{*2}}{\sigma_0^2} < \chi^2_{n-1;0.975}$.
- In the sample $s^* = 4.508$ and $n = 10$.

- So our test statistic is $\frac{(n-1)s^{*2}}{\sigma^2} = 11.431$, and the critical values are $\chi^2_{9;0.025} = 2.70$ and $\chi^2_{9;0.975} = 19.02$.
- Therefore we do not reject H_0.

3. Further remarks

Practical versus statistical significance.

- Statistical significance is important but we must also interpret the magnitude of the point estimates. We may call this the *practical significance* of the estimates (i.e. whether they are interesting in economic sense).
- In large samples we often find statistically significant point estimates which are not especially large.
- That is because our tests are *consistent*, i.e. they reject H_0 with probability approaching one as the sample size grows, provided that H_1 is true.

Material

Material.

- W Appendix C.6
- CB 8.1, 8.3.1 (except for Examples 8.3.7-8.3.8), 8.3.4 (except Example 8.3.30)
  - (only to the extent covered in the course)