Hypothesis testing

25.10.2005
GE02: day 4 part 1

Yurii Aulchenko
Erasmus MC Rotterdam

Statistic

- An experiment leads to data (observed)
- There is some expectation (null hypothesis) about what the data could be
- “Statistic” is a quantity which measures how (un)likely the deviation of observed from the expected is, under the null hypothesis. The more is deviation, the higher is statistics
- Examples of a statistic:
  Z, Wald, Score, χ^2, Likelihood Ratio Test...

P-value

- P-value is probability that the (observed or even more extreme) deviation of observed from expected could have occurred purely by chance

- Given a test statistic used, it is computed as the probability that the test statistic has a value "**at least as extreme**" as the observed value

Example: association study

- Association study between a disease and an A → T polymorphism
- Null hypothesis is that T has equal frequency in cases and controls
- We observe that the frequency of T in cases is 0.47 and the frequency of T in controls is 0.40

Two-sided P-value

We do not know *a priori*, is T or A involved in the disease

- Alternative hypothesis: frequency of T is not equal in cases and controls (T may be more frequent in cases or more frequent in controls)
- Though we observe that T is 7% <u>more</u> frequent in cases, the P-value is the probability that frequency of T would <u>differ</u> by ≥ 7% between cases and controls
- That is P-value is probability under null that T would be 7% <u>more</u> frequent in cases OR that T is 7% <u>less</u> frequent in cases
- This is called 2-sided P-values

One-sided P-value

We know that T disrupts the product of the gene or have other very strong evidence that T is a “bad”variant

- Alternative hypothesis: T is more frequent in cases
- P-value is probability that T would be $\geq$ 7% more frequent by chance
- This is called one-sided P

Hypothesis testing

If P-value is less then or equal to some pre-specified threshold, then you say that

- “At this threshold, I do not believe that this deviation from expected is purely by chance”
- “Therefore I reject the null hypothesis”

- Usually the threshold is taken to be 0.05 (5%)

P-value: an example

- Experiment: four heads in 20 trials
- Null hypothesis is that the coin is “fair”
  - P(head)=P(tail)=0.5
- Alternative hypothesis is that it is **not** fair
  - *A priori* we do not know in what way (heads or tails) it is not fair
- We see that deviation from expected (10) is 6
- P-value is the chance that a fair coin in 20 trials would deviate from expected (10) by 6 or more

Solution: 4 out of 20, Binomial

P-value = $P(k=0) + P(k=1) + P(k=2) + P(k=3) +$
$P(k=4) + P(k=16) + P(k=17) + P(k=18) +$
$P(k=19) + P(k=20) =$

$½^{20} + 20\ ½^{20} + 190\ ½^{20} + 1140\ ½^{20} +$
$4845\ ½^{20} + 4845\ ½^{20} + 1140\ ½^{20} + 190\ ½^{20} +$
$20\ ½^{20} + ½^{20} = 6196\ ½^{20} =$

0.012

...finally!..

Because it is symmetric:

P-value = $2 \times [P(k=0) + P(k=1) + P(k=2) + P(k=3) + P(k=4)]$

P-value using Normal app-n (score test)

$$P(k \le x) = \Phi\left(\frac{(x+0.5)-\mu}{\sigma}\right) = \Phi\left(\frac{(x+0.5)-np_0}{\sqrt{np_0(1-p_0)}}\right)$$

$$2 \cdot P(k \le 4) = 2 \cdot \Phi\left(\frac{4-10+0.5}{\sqrt{20 \cdot 0.5 \cdot 0.5}}\right) =$$

$$2 \cdot \Phi(-2.46) = 2 \cdot (1-\Phi(2.46)) =$$

$$2 \cdot (1-0.993) = 0.014$$

Score (Z) test

With (Yate's) continuity correction

$$Z = \begin{cases} \dfrac{|O - E - 0.5|}{\sqrt{Var_0}}, & if\ O > E \\ \dfrac{|O - E + 0.5|}{\sqrt{Var_0}}, & if\ O < E \end{cases}$$

Without correction

$$Z = \frac{|O - E|}{\sqrt{Var_0}}$$

Where Var_0 is the variance expected under Null

With 2-tailed P, Z>1.96 give P<0.05

P-value using Normal (Wald test)

$$P(k \le x) = \Phi\left(\frac{(x+0.5)-\mu}{\sigma}\right) = \Phi\left(\frac{(x+0.5)-np_0}{\sqrt{n\hat{p}(1-\hat{p})}}\right)$$

$$2 \cdot P(k \le 4) = 2 \cdot \Phi\left(\frac{4-10+0.5}{\sqrt{20 \cdot 0.2 \cdot 0.8}}\right) =$$

$$2 \cdot \Phi(-3.07) = 2 \cdot (1-\Phi(3.07)) =$$

$$2 \cdot (1-0.999) = 0.002$$

Wald test

With (Yate's) continuity correction

$$Z = \begin{cases} \frac{|O - E - 0.5|}{\sqrt{Var_1}}, & if\ O > E \\ \frac{|O - E + 0.5|}{\sqrt{Var_1}}, & if\ O < E \end{cases}$$

Without correction

$$Z = \frac{|O - E|}{\sqrt{Var_1}}$$

Where Var_0 is the variance expected under Null

Chi-squared distribution

- Chi-squared with *m* d.f. is sum of *m* squared standard Normal
- Consider several classes *i = 1, 2, … m*
- in each class expected (E_i) and observed (O_i) is known
- Chi-squared statistic (with and w/o cont-ty corr.)

$$\sum_{i=1}^{m} \frac{(O_i - E_i \pm 0.5)^2}{E_i} \infty \chi^2_{m-1}$$

$$\sum_{i=1}^{m} \frac{(O_i - E_i)^2}{E_i} \infty \chi^2_{m-1}$$

Coin tossing with n = 20, k = 4

- class one: observed 4 heads, expected 10
- class two: observed 16 tails, expected 10

$$\sum_{i=1}^{2} \frac{(O_i - E_i \pm 0.5)^2}{E_i} = \frac{(4-10+0.5)^2}{10} + \frac{(16-10-0.5)^2}{10}$$

= 3.025 x 2 = 6.05 (P-value = 0.014)

Exactly the same as Score test without correction!

Critical values of χ_1^2

- For 1 d.f. can be computed by squaring Normal values:
  - P=0.05 : 3.84 (=1.96^2)
  - P=0.01 : 6.63 (=2.57^2)

- More degrees of freedom:
  - Use Tables

Likelihood Ratio Test

- Two hypothesis considered
  - Null, H_0: the parameter **p** has some fixed value
    - $\mathbf{p = p_0}$
  - Alternative, H_1: the parameter **p** has a value which maximises the likelihood of getting the observed data
    - **p = p'**

- Let the probability of the data, computed under
  - Null $L(K \mid p_0)$
  - Alternative $L(K \mid \hat{p})$

Likelihood Ratio Test

- Then $$LRT = 2\log_e\left[\frac{L(K \mid \hat{p})}{L(K \mid p_0)}\right] \propto \chi_1^2$$

More general:

- Hierarchical hypotheses are considered
- The number of degrees of freedom is the difference in number of parameters under estimation

Likelihood Ratio Test

- Coin tossing with n = 20, k = 4
  - H_0: p_0 = ½
    - $L_0 = C(20,4)\ (1/2)^{20}$
  - H_1: p’= 4/20
    - $L_1 = C(20,4)\ (4/20)^4\ (16/20)^{16}$

$$LRT = 2\log_e\left[\frac{L(K\mid\hat{p})}{L(K\mid p_0)}\right] =$$

$$2\log_e\left[\frac{0.2^4\cdot 0.8^{16}}{(1/2)^{20}}\right] = 2\log_e[47.22] = 2\cdot 3.85 = 7.7$$

P-value = 0.005

Many P's to choose from…

In coin tossing with n = 20, k = 4

P-value is equal to

- 0.014 (Z = χ^2 -test with continuity correction)
- **0.012 (exact binomial)**
- 0.007 (Z = χ^2 -test without continuity correction)
- 0.005 (Likelihood Ratio Test)
- **0.002** (Wald test with continuity correction)
- **0.001** (Wald test without continuity correction)

Do DO continuity correction!

When numbers are small, large numbers approximation may work poor

Importance of large numbers

If $n = 200$, $k = 82$

P-value is equal to

- **0.013 (exact binomial)**
- 0.013 (Z-test with continuity correction)
- 0.012 (Wald test with continuity correction)
- 0.011 (Z-test without continuity correction)
- 0.011 (LRT)
- 0.010 (Wald test without continuity correction)