Hypothesis testing

27.10.2005
GE02: day 3 part 1

Yurii Aulchenko
Erasmus MC Rotterdam

Statistic

- … generally, is a quantity which measures deviation of observed from expected
- In a series 100 coin tosses 59 tails are obtained. Possible statistics are
  - Absolute deviation from expected number of tails:
    - |#observed – #expected| = |59 – 50| = 9
  - Absolute deviation from expected on standard normal scale:
    - |observed – expected| / std.dev. =

      $|59 - 50| / \sqrt{(100\ ½\ ½)} = 9 / 5 = 1.8$
  - … etc.

P-value

- … is probability that test statistics has a value “at least as extreme” as the observed value

P-value: example

- For previous experiment, (59 tails in 100 tosses) this probability is
  - Statistics: abs. deviation from expected number of tails:
    - |#observed – #expected| = |59 – 50| = 9
  - P-value
    - P(|#observed – #expected|≥9) = [binomial distribution]

      P(#tails ≥ 59 or #tails≤41) = **0.089**
  - Statistics: abs. deviation from expected on standard normal scale:
    - |observed – expected| / std.dev. = 1.8
    - P(x≤-1.8 or x≥1.8) = [normal distribution]

      $\Phi(x \leq -1.8 \text{ or } x \geq 1.8) = 2\,(1 - P(x \geq 1.8)) = 2\;0.9641 - 1 =$ **0.0718**

Hypothesis testing

- If P-value is less then or equal to some pre-specified threshold (α), then the null hypothesis (e.g. fair coin) is rejected
- Usual α is taken to be 0.05

Why absolute deviation?

- Experiment: 100 coin tosses
- Outcome: 59 tails
- Statistic: absolute deviation from expected number of tails

  $|\#observed - \#expected| = |59 - 50| = 9$

- P-value:

  $P(|\#observed - \#expected| \geq 9) = P(\#tails \geq 59 \text{ or } \#tails \leq 41)$

- Why not just $P(\#tails \geq 59)$?
  - Not guilty until proven!
  - The reason is that under null we assume that the coin is fair, and in this case, it can deviate to any side from the expected

Not always absolute deviation

- Polymorphism: A → T
  - Testing involvement of the polymorphism into disease

    Test: T is more OR less frequent in cases
  - If we know that T disrupts the product of the gene or have some other evidence that T is a "bad"variant

    Test: T is more frequent in cases

Task

- Test if coin is fair given
  - One tail out of ten
  - Two tails out of ten
  - Use both Normal and Binomial for the test

Answer: 1 out of 10

- Binomial

  P(deviation ≥ 4) =

  $P(k{=}9) + P(k{=}10) + P(k{=}1) + P(k{=}0) =$

  $10\ \tfrac{1}{2}^{10} + \tfrac{1}{2}^{10} + 10\ \tfrac{1}{2}^{10} + \tfrac{1}{2}^{10} = 22\ \tfrac{1}{2}^{10} = 0.021$

- Normal

  $\Phi(\text{deviation} \geq 4/\sqrt{(10\ \tfrac{1}{2}\ \tfrac{1}{2})}) =$

  $\Phi(x \geq 2.53 \text{ OR } x \leq -2.53) =$

  $2\ (1 - 0.9943) = 0.011$

Answer: 2 out of 10

- Binomial

  P(deviation ≥ 3) =

  $$= P(k=8) + P(k=9) + P(k=10) + P(k=2) + P(k=1) + P(k=0)$$

  $$45\, \tfrac{1}{2}^{10} + 10\, \tfrac{1}{2}^{10} + \tfrac{1}{2}^{10} + 45\, \tfrac{1}{2}^{10} + 10\, \tfrac{1}{2}^{10} + \tfrac{1}{2}^{10} =$$

  $$112\, \tfrac{1}{2}^{10} = 0.109$$

- Normal

  $\Phi(\text{deviation} \geq 3/\sqrt{(10\ \tfrac{1}{2}\ \tfrac{1}{2})}) =$

  $$\Phi(x \geq 1.9 \text{ OR } x \leq -1.9) =$$

  $$2\,(1 - 0.9713) = 0.0574$$

Chi-squared distribution

- Chi-squared with m d.f. is sum of m squared standard Normal
- Consider several classes $i = 1, 2, \ldots m$
- in each class expected (E_i) and observed (O_i) is known
- Chi-squared statistics

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

- is distributed as chi-squared test with the number of degrees of freedom (m – 1)

Chi-squared distribution

- Coin tossing with $n = 20$, with $k = 4$
  - class one: observed 4 heads, expected 10
  - class two: observed 16 tails, expected 10

$$\chi_1^2 = \sum_{i=1}^{2} \frac{(O_i - E_i)^2}{E_i} = \frac{(4-10)^2}{10} + \frac{(16-10)^2}{10}$$

$$= 3.6 \ x \ \ 2 = 7.2 \ (P\text{-value} = 0.0073)$$

Critical values of χ_1^2

- If chi-squared at 1 d.f. $\geq 3.84 \Rightarrow P \leq 0.05$
- If chi-squared at 1 d.f. $\geq 6.63 \Rightarrow P \leq 0.01$
- If chi-squared at 1 d.f. $\geq 7.88 \Rightarrow P \leq 0.005$

Likelihood Ratio Test

- Two hypothesis considered
  - Null, H_0: $p = 0.5$
  - Alternative, H_1: p has a value maximizing the observed data
- Let $P_0(K)$ and $P_1(K)$ are the probabilities of the data under null and alternative

Likelihood Ratio Test

- Then

$$LRT = 2\log_e\left[\frac{P_1(K)}{P_0(K)}\right] \infty \chi_1^2$$

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: $p = ½$
    - $P_0(K) = C(20,4)\ (1/2)^{20}$
  - H_0: $p = 4/20$
    - $P_1(K) = C(20,4)\ (4/20)^4\ (16/20)^{16}$

$$LRT = 2\log_e\left[\frac{P_1(K)}{P_0(K)}\right] = 2\log_e\left[\frac{0.2^4 \cdot 0.8^{16}}{(1/2)^{10}}\right] = 2\log_e[47.22] = 2 \cdot 3.85 = 7.7$$

P-value = 0.0055

Importance of large numbers

- In coin tossing with $n = 20$, with $k = 4$
- P-value is equal to
  - 0.018 (exact binomial)
  - 0.007 (chi-squared approximation)
  - 0.006 (LRT)
- When numbers are small, Normal approximation may work poor, hence Z-, χ^2- and LRT-test are not behaving very good

Testing HWE, using chi2

$$X^2 = \Sigma_{i=1,m} \frac{(E_i - O_i)^2}{E_i}$$

- *number of degrees of freedom is*

  (number of genotypes – number of alleles)

genotype	No.	Expected	Expected	(o-e)2/e	
DD	2	0,01	1,1	0,73	
ND	17	0,19	18,8	0,17	
NN	81	0,8	80,1	0,01	P-value

Excel functions

- Cumulative binomial P($x \le k$)

    =binomdist(k,n,p,1)

- Cumulative standard normal $\Phi(x \le k)$

    =normdist(k,0,1,1)

- Chi-squared with m d.f., $\chi_m^2(k)$

    =chidist(k,m)