Power of genetic epidemiology (association) study

25.10.2007

GE02 day 4 part 3

Yurii Aulchenko

Erasmus MC Rotterdam

Hypothesis testing

- If P-value is less then or equal to some pre-specified threshold (α), then you say that you reject the null hypothesis
- Usually (for single test) α is taken to be 0.05
- This means that if you do many tests and null is true, you are going to reject the true hypothesis in 5% of tests!

Type 1 (α) and 2 (β) errors

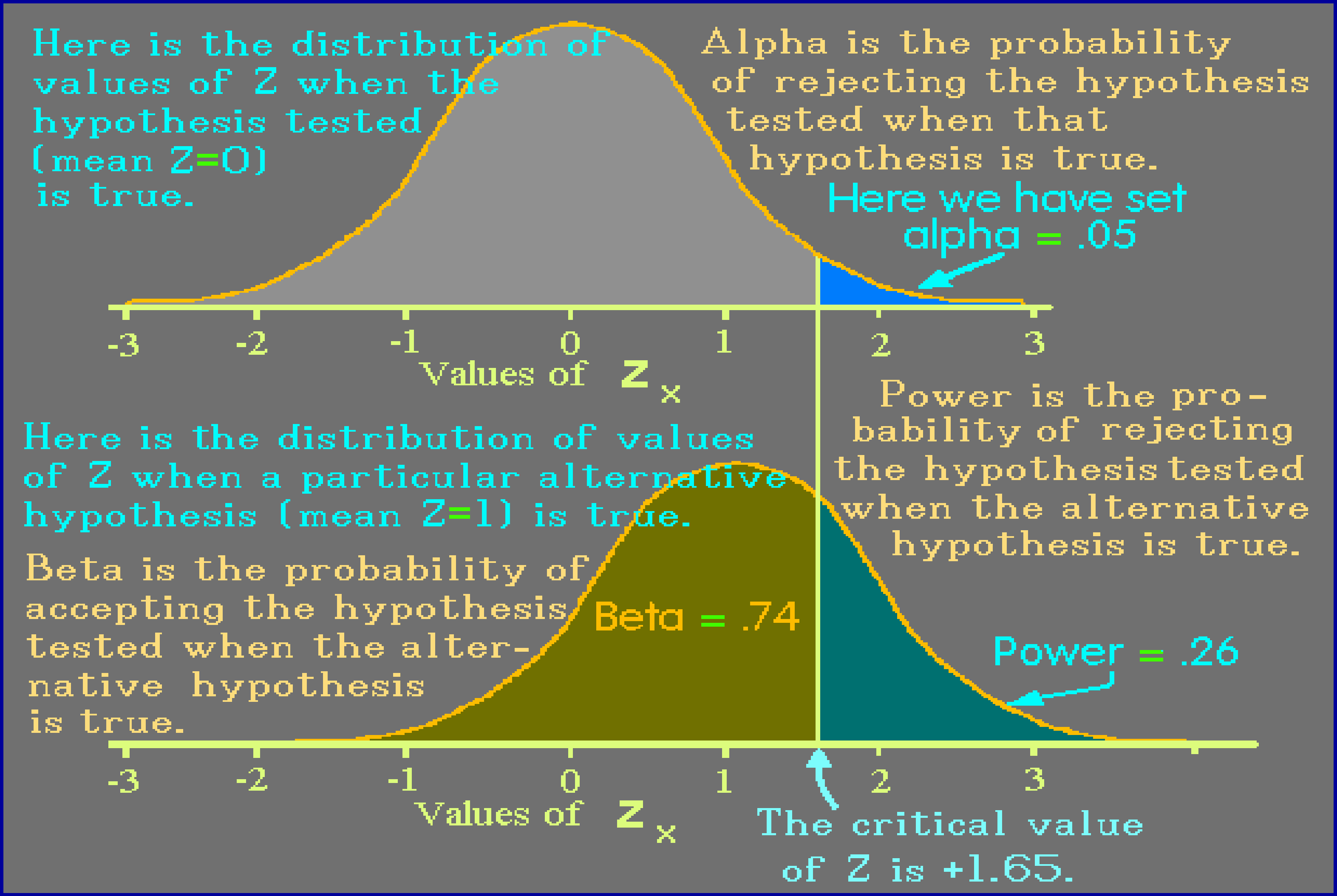
	Reality: null hypothesis is	
Test says	True	False
Reject	α	$1 - \beta$
Accept	$1 - \alpha$	β

Type 1 error

Power

Type 2 error

Power explained



Computing β and power

If expectation of the test statistics under alternative is X than

- Type 2 error is

$$\beta = \Phi(T - X)$$

- Power = (1 - β) is

$$1-\beta = 1 - \Phi(T - X) = \Phi(X - T)$$

Test statistics expected under alternative

- Binomial experiment with n and p
- If under null $p_0=0.5$ and under alternative $p_1=0.6$ then expected Z test statistics is
  - n=10 : $E[Z]=(6-5)/\sqrt{(10\ ½\ ½)} = 0.63$
  - n=40 : $E[Z]=(24-20)/(½\sqrt{(40)}) = 1.26$
  - n=160 : $E[Z]=(96-80)/(½\sqrt{(160)}) = 2.53$
  - n=640 : $E[Z]=(384-320)/(½\sqrt{(640)}) = 5.06$

Logic of power computations

- You wish to achieve significance of α and power of $(1 - \beta)$ [and hence $\beta = (1 - \text{power})$]
- Under null, the statistics is described by Standard Normal
- The cut-off statistics value is determined by α

- Given sample size, you can compute the expected test statistics under alternative
- What is the sample size which gives you right β?

Example

- If under null $p_0=0.5$ and under alternative $p_1=0.6$ then expected test statistics is
- $\alpha=0.05$ (threshold T=1.96) and power desired is 80% (thus $\beta<0.2$)
  - n=10 : $E[Z]=0.63$: $1-\beta = \Phi(0.63-1.96) = \Phi(-1.33) = 0.09$
  - n=40 : $E[Z]=1.26$: $1-\beta = \Phi(1.26-1.96) = \Phi(-0.7) = 0.24$
  - n=160 : $E[Z]=2.53$: $1-\beta = \Phi(2.53-1.96) = \Phi(0.57) = 0.72$
  - n=640 : $E[Z]=5.06$: $1-\beta = \Phi(5.06-1.96) = \Phi(3.1) = 0.999$

Genetic power calculator

http://pngu.mgh.harvard.edu/~purcell/gpc/

Genetic Power Calculator

S. Purcell & P. Sham, 2001-2005

This site provides automated power analysis for variance components (VC) quantitative trait locis (QTL) linkage and association tests in sibships, and other common tests. It is currently under construction - suggestions, comments to *Shaun Purcell*. If you use this site, please reference the following Bioinformatics a

```
Purcell S, Cherny SS, Sham PC. (2003) Genetic Power Calculator:
design of linkage and association genetic mapping studies of complex
traits. Bioinformatics, 19(1):149-150.
```

Modules

VC QTL linkage for sibships	Notes
VC QTL association for sibships	Notes
VC QTL linkage for sibships conditional on trait	Notes
MPIC: Multipoint Polymorphism Information Content	Notes
TDT for discrete traits	Notes
Case-control for discrete traits	Notes
TDT for threshold-selected quantitative traits	Notes
Case-control for threshold-selected quantitative traits	Notes
Probability Function Calculator	Notes

GRR(Aa) = P(disease|Aa)/P(disease|aa)

GRR(AA) = P(disease|AA)/P(disease|aa)

Power of the test that p ≠ ½

- Some anticipated p

$$\frac{|np - \frac{n}{2}|}{\sqrt{n \cdot 0.5 \cdot 0.5}} \geq Z_\alpha + Z_\beta$$

$$n = \frac{\left(Z_\alpha + Z_\beta\right)^2}{\left(2p - 1\right)^2}$$

- For α =0.05 and power = 1 – β = 80%
  - Z_α = 1.96, Z_β = 0.84

Example

- A TDT test is performed
- Expected ratio of segregation of the susceptibility allele is 0.57
- How many informative meioses we need?

$$n = \frac{\left(Z_\alpha + Z_\beta\right)^2}{\left(2p-1\right)^2} =$$

$$\frac{\left(1.96+0.84\right)^2}{\left(2\cdot 0.57-1\right)^2} = \frac{7.84}{0.0196} = 400$$

Test for association

- Assume that number of cases = number of controls = n
- Frequency in controls is p
- Expected frequency in cases is p_1

Power of the test that $p_1 \neq p$

- Expected Z is

$$\frac{2np_1 - 2np}{\sqrt{4n\frac{p+p_1}{2}\left(1-\frac{p+p_1}{2}\right)}} \geq Z_\alpha + Z_\beta$$

$$n = \frac{\left(2p(1-p_1) + p_1(2-p_1) - p^2\right)\left(Z_\alpha + Z_\beta\right)^2}{4(p_1 - p)^2}$$

Example

- Frequency in controls is 0.2
- Anticipated frequency in cases is 0.25
- What sample # of cases is needed at alpha of 0.05 and power of 0.8?

$$n = \frac{\left(2p(1-p_1) + p_1(2-p_1) - p^2\right)\left(Z_\alpha + Z_\beta\right)^2}{4(p_1 - p)^2}$$

$$n = \frac{\left(2 \cdot 0.2 \cdot 0.75 + 0.25 \cdot (2 - 0.25) - 0.2^2\right) \cdot \left(1.96 + 0.84\right)^2}{4 \cdot (0.25 - 0.2)^2} = 547$$

Putting it all together…

- Study of Sladek (Nature 2007) reports novel locus for T2D, HHEX
- At SNP rs7923837, frequency of the susceptibility allele (G) was 0.67 in cases and 0.62 in controls

- What sample size is needed for a confirmation study for this SNP? (power = **90%**, nominal P=0.05)
- What was the power to detect this locus in the study of Sladek? (Assuming 400K SNPs, 600 cases and 600 controls)
- What sample size needed to discover this gene in a GWA with 400K SNPs with 90% power?

Sample size for conformation

- Power of allelic test:

$$n = \frac{\left(2p(1-p_1) + p_1(2-p_1) - p^2\right)\left(Z_\alpha + Z_\beta\right)^2}{4(p_1 - p)^2}$$

- Za = 1.96, Zb = 1.28

$$n = \frac{\left(2 \cdot 0.62 \cdot 0.33 + 0.67 \cdot (2 - 0.67) - 0.62^2\right) \cdot \left(1.96 + 1.28\right)^2}{4 \cdot (0.67 - 0.62)^2} = 961$$

- Thus: 1000 cases, 1000 controls!

Power to discover with 600:600

Required numbers

- Desired nominal P = 0.05/400,000 = 1.25 x 10^{-7}
- Corresponding Za = 5.29
- E[Z] =

  $(0.67*1200-0.62*1200)/ \sqrt{[2400\ 0.645\ (1-0.645)]} = 2.56$

Genome-wide power?

- $P(Z>5.29) = \Phi(2.56 - 5.29) = \Phi(-2.73) =$ **0.003 (!)**

Sample size to have 90% GWA power

- Power of allelic test:

$$n = \frac{\left(2p(1-p_1) + p_1(2-p_1) - p^2\right)\left(Z_\alpha + Z_\beta\right)^2}{4(p_1 - p)^2}$$

- Za = 5.29, Zb = 1.28

$$n = \frac{\left(2\cdot 0.62\cdot 0.33 + 0.67\cdot(2-0.67) - 0.62^2\right)\cdot\left(5.29+1.28\right)^2}{4\cdot(0.67-0.62)^2} = 3{,}953$$

- Thus: 4000 cases, 4000 controls!

Power of QTL study here???