Meta-analysis and Power

Benjamin Neale

March 8th 2011

Boulder Workshop

Slide of Contents

- Meta-analysis
  - Motivation
  - Principles
  - Approaches
  - Practical using METAL
- Power
  - Concept
  - Importance
  - Practical using GPC

Why do we want to do meta-analysis?

You're meant to volunteer answers
[really, I don't have any slides on this,
so please help]

Meta-analysis motivations

Goal: Combine Data from Multiple Sources

Test statistics

Estimates + standard error

P-values

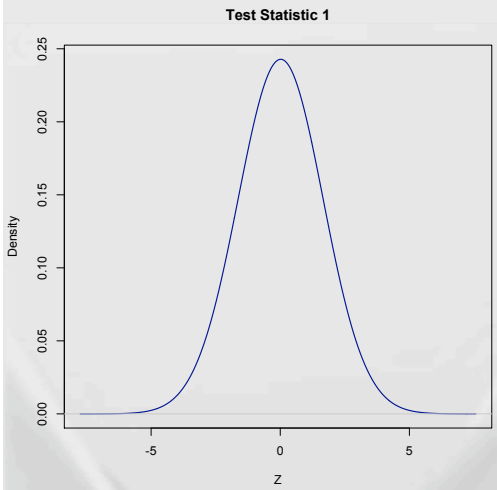
Test statistic combination

- Z-score combination
- A little bit of variance/covariance rules:
- $Cov(X,X) = Var(X)$
- $Var(X+Y) = Var(X) + Var(Y) + 2*cov(X,Y)$
- $Var(k*X) = k^2*Var(X)$
- $Var(X+X) = Var(X) + Var(X) + 2*cov(X,X)$
  $= 4*Var(X)$
- $Var(2*X) = 2^2Var(X) = 4*Var(X)$

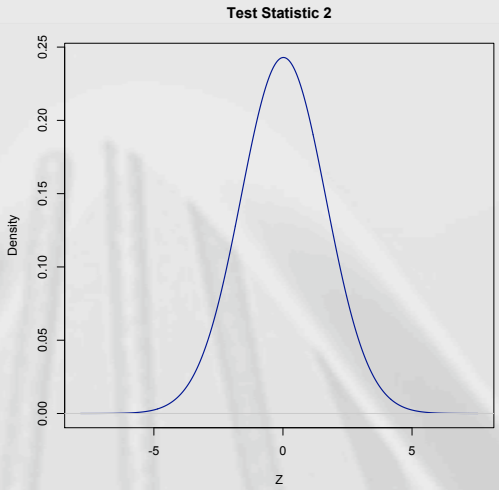
So we have Z-scores

- We can combine Z-scores for meta-analysis
- We want to weight the contribution of each dataset by the amount of 'information'
- Information in this context largely reflects sample size

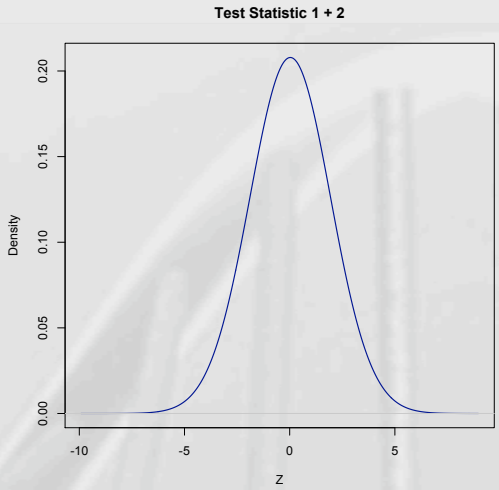
Two samples which are the same size



+



=



Var(T1) = 1

Var(T2) = 1

Var(T1 + T2) = ?

T1 and T2 are 10,000 test statistics simulated under the null [no effect]
More on the null later

Why do we want to know what the variance of T1 + T2 is?

Why variance matters

Test Statistic	# P < 0.05	# P < 0.01	# P < 0.001
T1	513	104	12
T2	512	105	12
T1 + T2 ($\sigma^2 = 1$)	1643	694	217
T1 + T2 ($\sigma^2 = \text{sqrt}(2)$)	508	103	12

If we assume the wrong variance of the test statistic we dramatically inflate our distribution of results

Weighted Z meta-analysis

$$Z^* = \sum_{i=1}^{m} \sqrt{\frac{w_i}{w_t}} Z_i$$

Meta Z

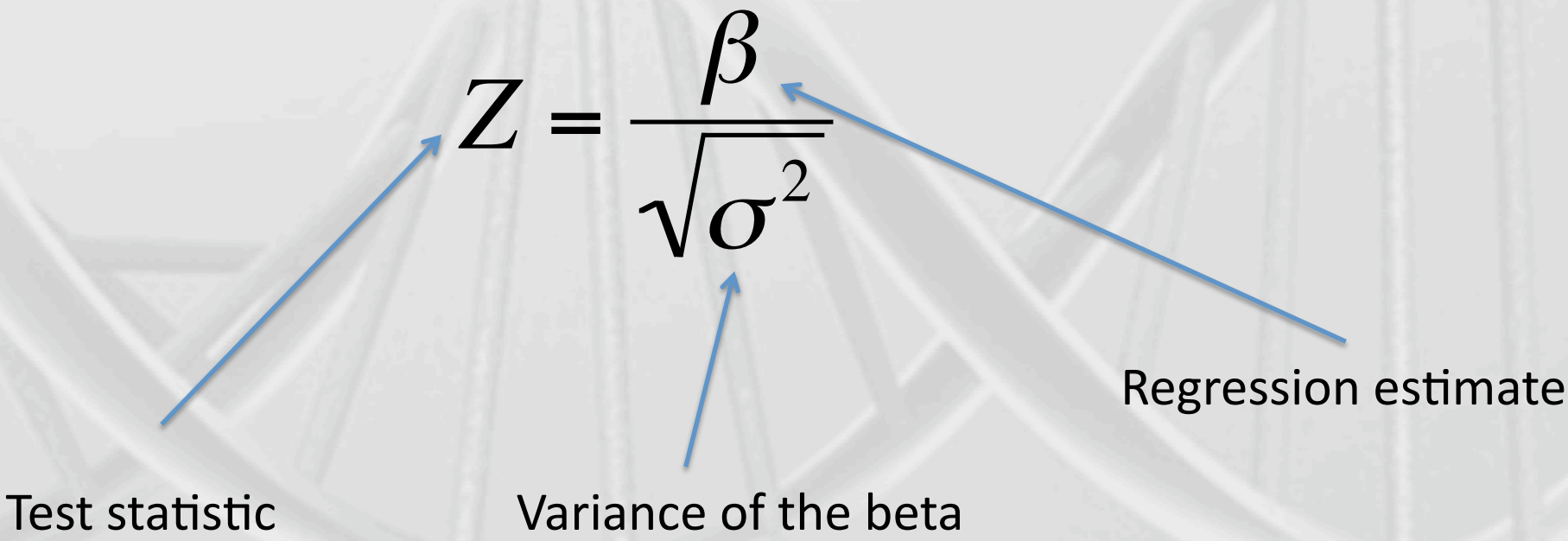
Sum from 1 to m
Where m is # Z's

w_i is each test weight
w_t is the sum of the weights

Each study Z

We can work on betas too…

- For regression we can get a test statistic from the beta



For meta-analysis

$$\beta^* = \frac{\sum_{i=1}^{m} \frac{\beta_i}{\sigma_i^2}}{\sum_{i=1}^{m} \frac{1}{\sigma_i^2}} \qquad SE^* = \sqrt{\frac{1}{\sum_{i=1}^{m} \frac{1}{\sigma_i^2}}}$$

B_i = study beta; σ^2 = variance of each individual betas

Warning: for Beta meta-analysis to be appropriate, the constituent data need to be measured on the same phenotype and the same scale

For example, meta-analysis of height in inches and height in centimeters would not be appropriate on beta and standard errors, and so we have to transform

We can also assess blind to direction

- Fisher's method:

$$\chi^2_{2k} = -2\sum_{i=1}^{k}\log_e(p_i)$$

- Sum of χ^2

$$\chi^2_k = \sum_{i=1}^{k}\chi^2_i$$

These give similar, though not Identical answers for P-values

K = # of tests; p_i are the p-values of each study;

These methods can be less efficient than Z-score meta-analysis

Though it depends on model and heterogeneity etc.

What about practical details?

Strand, strand, strand

- DNA is a double helix
- A pairs with T and C pairs with G
- There are two strands:

ATCTGGTACTCCAT Strand 1

TAGACCATGAGGTA Strand 2

Strand, strand, strand

- What about SNPs?

ATCTGGT[A/C]CTCCAT Strand 1

TAGACCA[T/G]GAGGTA Strand 2

Strand, strand, strand

- What's the big/annoying problem?

ATCTGGT[A/T]CTCCAT Strand 1

TAGACCA[T/A]GAGGTA Strand 2

Strand, strand, strand

- Two ambiguous SNP types
  - A/T and G/C
  - All others are resolved
- How to check?
  - Allele frequencies [know your population]
  - LD [if you have raw data]
- PLINK and METAL can re-orient strand
  - Remember the ambiguous ones!

Metal How To

Metal Documentation:
http://genome.sph.umich.edu/wiki/Metal_Documentation

Results files examples

- Metal is flexible
  - It can run fixed effects meta-analysis
  - Heterogeneity tests
  - Effect size, Sample Size, or Weighted meta-analysis

METAL

- Requires results files
- 'Driver' file
  - Describes the input files
  - Defines meta-analysis strategy
  - Names output file

Practical

Steps

1. Check format of results files
    1. Ensure all necessary columns are available
    2. Modify files to include all information
2. Prepare driver file
    1. Ensure headers match description
    2. Crosscheck each results file matches Process name
3. Run metal

Step 1: Preparing Input Files

- PLINK results files are readable by metal
  - For CMH and logistic regression, we'll have to specify that the OR is reported
- I've made a second set of results
  - res_for_metal2.txt
  - Simulated using final.cmh.cmh as a seed

Columns METAL uses

- SNP
- OR
- SE [for standard error meta-analysis]
- P-value [for Z-score meta-analysis]
- If we had two samples of different sizes we would have to add an N/weight column

Meta-analysis running

- We will run meta-analysis based on effect size and on test statistic
- For the weights of test statistic, I've assumed that the sample sizes are the same
  - METAL defaults to weight of 1 when no weight column is supplied

To run metal

- Copy over files from /faculty/ben/2011/meta_analysis_power/
- cp final.cmh.cmh
  - [or keep it where you generated the data]
  - Results file 1
- cp res_for_metal2.cmh
  - New results file generated
- cp metal_run_file
  - Driver file for Metal
- cp reformat.sh
  - A little reformatting script to change the header

Step 2: driver file: meta_run_file

```
MARKER  SNP
ALLELE  A1 A2
PVALUE  P
EFFECT  log(OR)
STDERR  SE

PROCESS final.cmh.cmh
PROCESS res_for_metal2.txt

OUTFILE meta_res_Z .txt
ANALYZE

CLEAR
SCHEME STDERR
PROCESS final.cmh.cmh
PROCESS res_for_metal2.txt

OUTFILE meta_res_SE .txt
ANALYZE
```

Specifies columns in file

Processes two results files

Conducts Z-based meta-analysis from Test statistic
Output file naming

Clears workspace
Changes meta-analysis scheme to beta + SE
Processes two results file

Output file naming

OR is the column in the file
log(OR) tells METAL to take the log of the OR

Running metal

- metal < metal_run_file
- metal is the command
- metal_run_file is the driver file
- This will output information on the running of METAL things to standard out [the terminal]
- It will spawn 4 files:
  - 2 results files: meta_res_Z1.txt and meta_res_SE1.txt
  - 2 info files: meta_res_Z1.txt.info and meta_res_SE1.txt.info

Output you'll see

- Overview of METAL commands
- Any errors
  - There are some—we'll chat :)
- And your best hit from meta-analysis

Results files

- head meta_res_Z1.txt

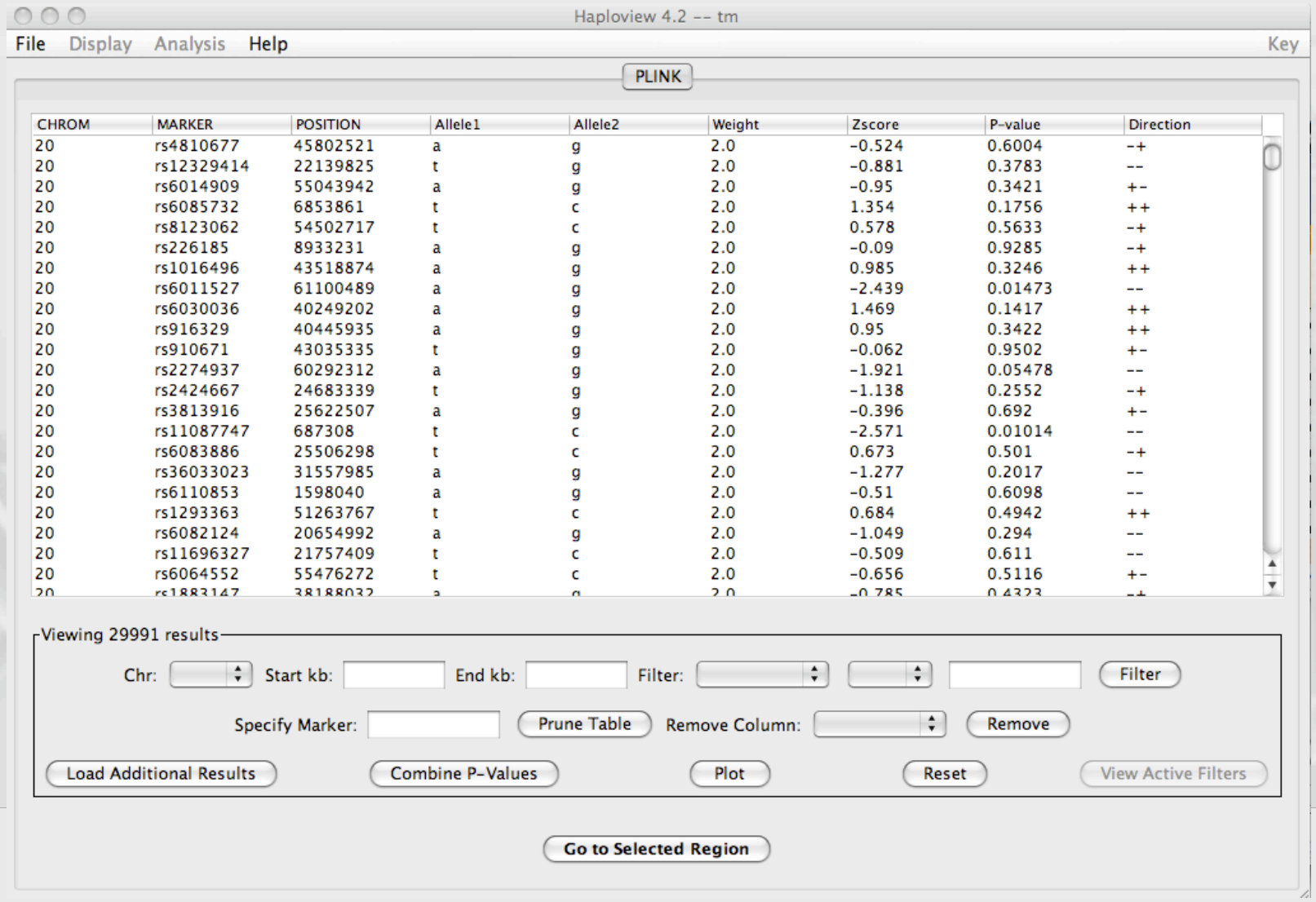
MarkerName	Allele1	Allele2	Weight	Zscore	P-value	Direction
rs4810677	a	g	2.00	-0.524	0.6004	-+
rs12329414	t	g	2.00	-0.881	0.3783	--
rs6014909	a	g	2.00	-0.950	0.3421	+-
rs6085732	t	c	2.00	1.354	0.1756	++
rs8123062	t	c	2.00	0.578	0.5633	-+
rs226185	a	g	2.00	-0.090	0.9285	-+
rs1016496	a	g	2.00	0.985	0.3246	++
rs6011527	a	g	2.00	-2.439	0.01473	--
rs6030036	a	g	2.00	1.469	0.1417	++

To load into Haploview

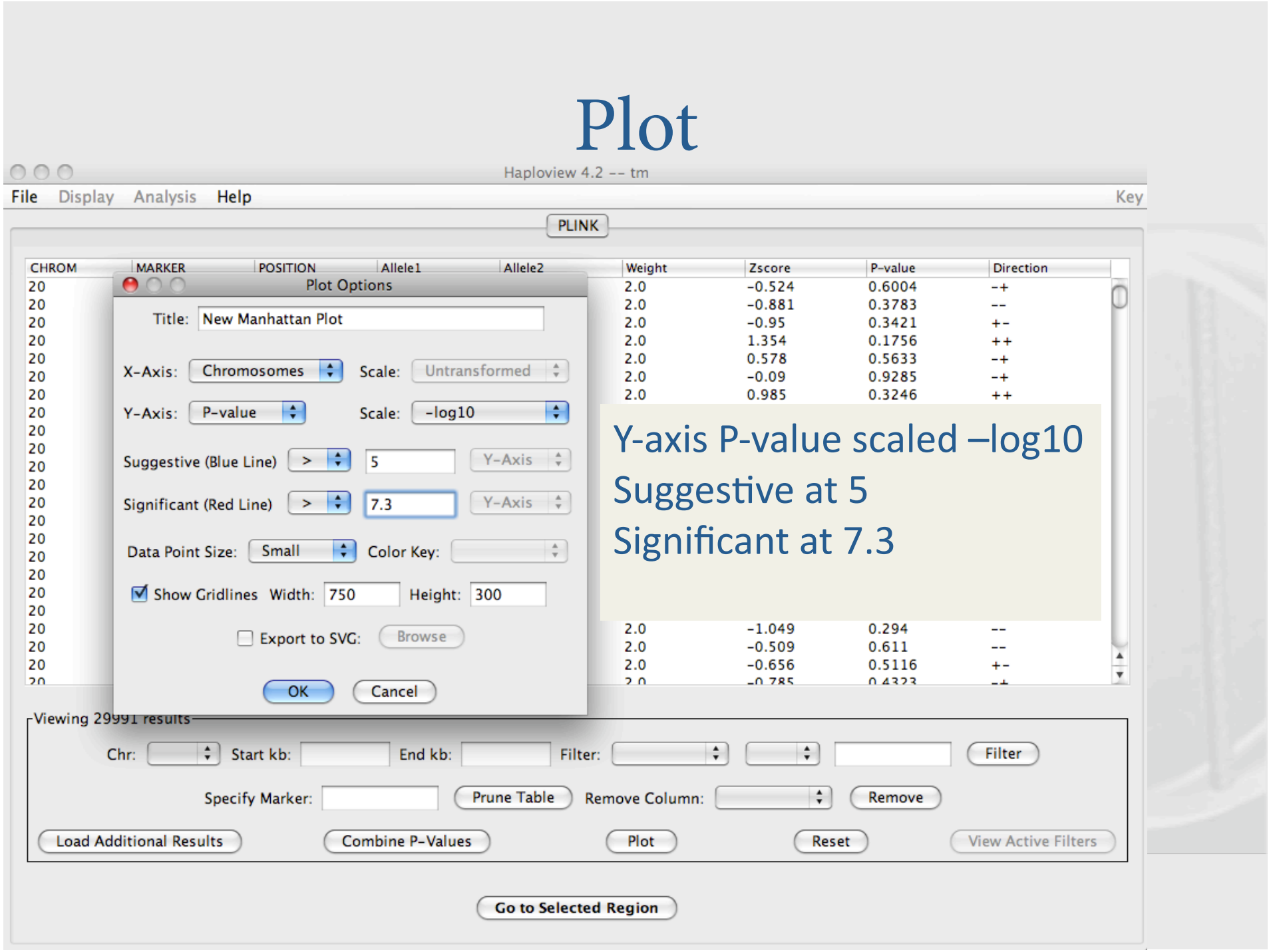
- We have to change the header
- In the same directory run:
- ./reformat.sh
  - This changes 1st column name to SNP
- We can then load the meta-analysis results files into haploview
  - Same as before but load in the meta_res_Z1.txt
  - Make sure to include the bim file that Jeff used earlier

Make sure you ran reformat.sh otherwise Haploview will not work

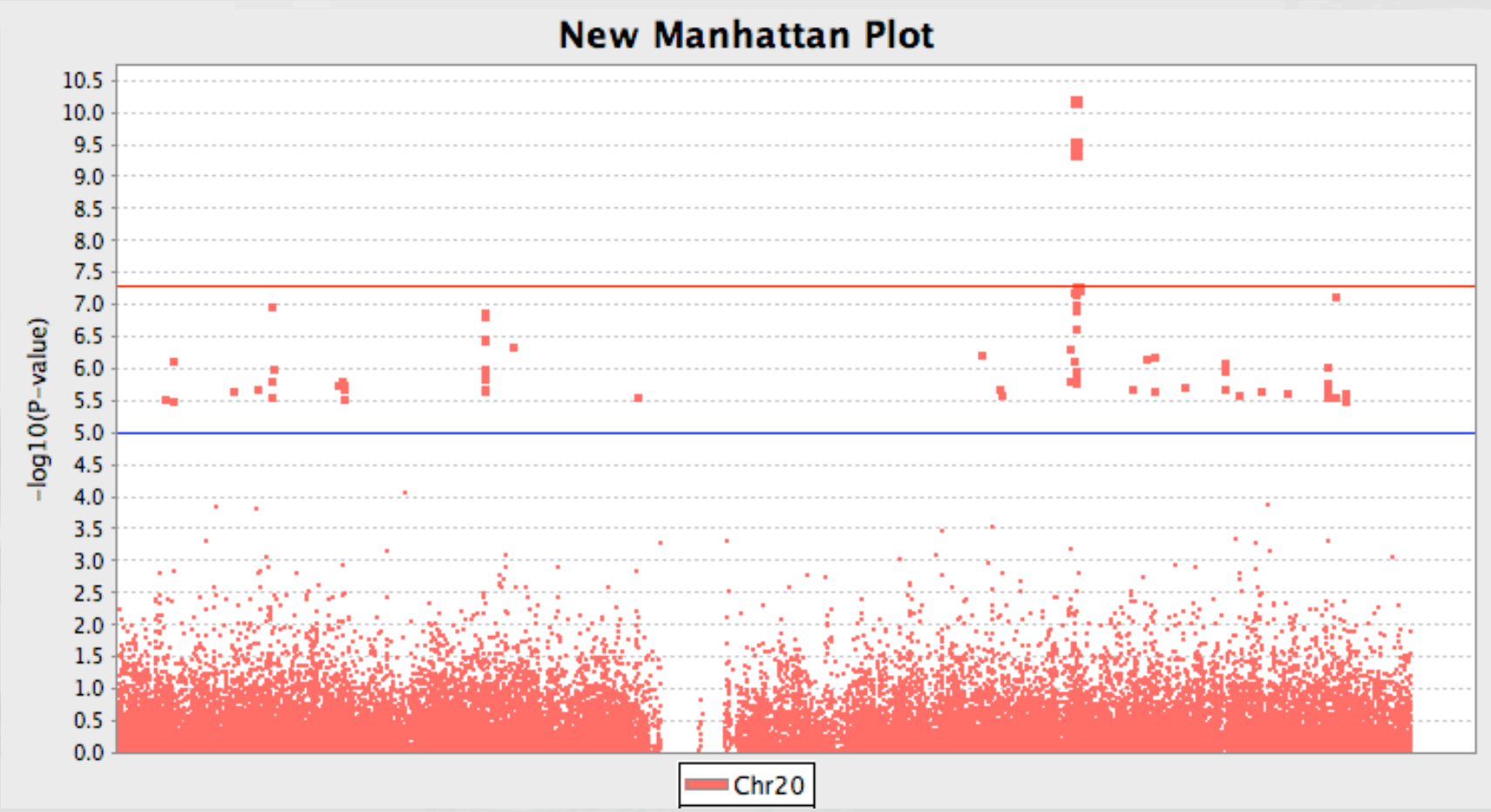
What you should see



Plot


Haploview 4.2 -- tm
File Display Analysis Help Key
PLINK
CHROM MARKER POSITION Allele1 Allele2 Weight Zscore P-value Direction
Plot Options
Title: New Manhattan Plot
X-Axis: Chromosomes Scale: Untransformed
Y-Axis: P-value Scale: -log10
Suggestive (Blue Line) > 5 Y-Axis
Significant (Red Line) > 7.3 Y-Axis
Data Point Size: Small Color Key:
Show Gridlines Width: 750 Height: 300
Export to SVG: Browse
OK Cancel
Y-axis P-value scaled –log10
Suggestive at 5
Significant at 7.3
Viewing 29991 results
Chr: Start kb: End kb: Filter: Filter
Specify Marker: Prune Table Remove Column: Remove
Load Additional Results Combine P-Values Plot Reset View Active Filters
Go to Selected Region

New results


New Manhattan Plot
-log10(P-value)
10.5
10.0
9.5
9.0
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
Chr20

Power!

Importance of power calculation

- Help design studies that are likely to succeed
  - Determine the minimum sample size necessary to achieve the desired level of statistical power (usually > 80%), for a given effect size
  - Determine the minimum effect size that can be detected with adequate statistical power, for a fixed sample size

Importance of power calculation

- Help design studies that are likely to succeed
  - Determine the minimum sample size necessary to achieve the desired level of statistical power (usually > 80%), for a given effect size
  - Determine the minimum effect size that can be detected with adequate statistical power, for a fixed sample size

Usually obligatory for grant applications

Hypothesis Testing

- SNP testing hypotheses:
  - h_o (null hypothesis) is $\beta=0$
  - h_a (alternative hypothesis) is $\beta\neq0$
    - Two-sided test, where $\beta > 0$ or $\beta < 0$ are one-sided
- Null hypothesis usually assumes no effect
- Alternative hypothesis is the idea being tested

Summary of Possible Results

Statistics

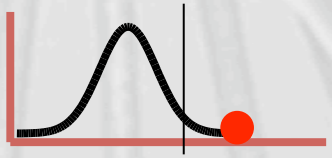
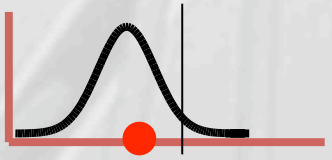
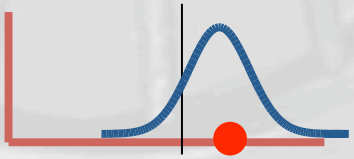
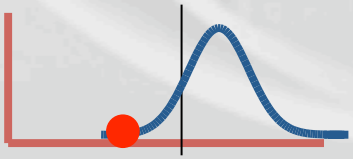
Reality	Rejection of H_0	Non-rejection H_0
H_0 true	α	$1-\alpha$
H_A true	$1-\beta$	β

α=type 1 error rate

β=type 2 error rate

$1-\beta$=statistical power

STATISTICS

REALITY	Rejection of H_0	Non-rejection of H_0
H_0 true	Type I error at rate α 	Nonsignificant result $(1-\alpha)$ 
H_A true	Significant result $(1-\beta)$ 	Type II error at rate β 

Power

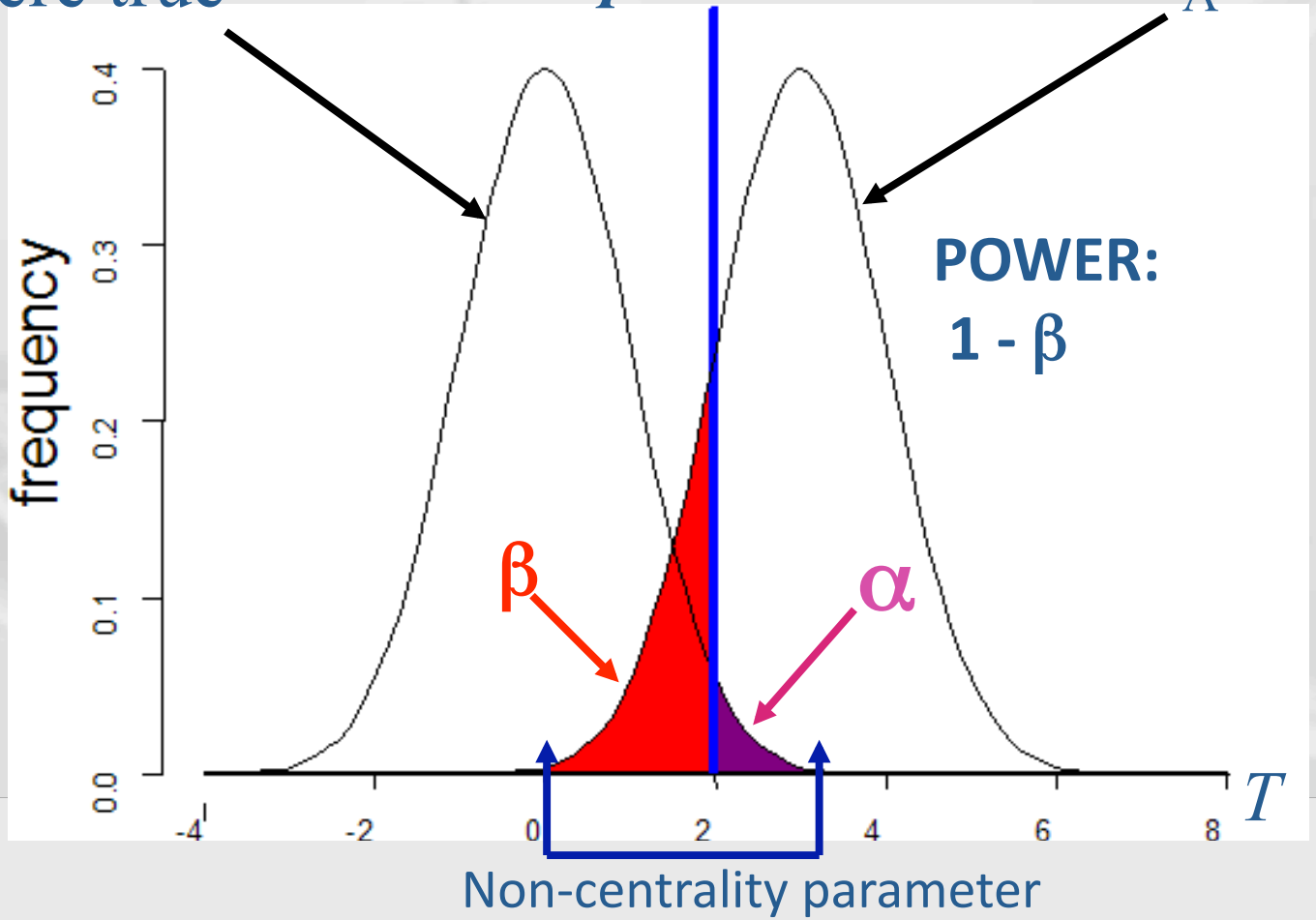
- The probability of rejection of a false null-hypothesis depends on:
  - the significance criterion (α)
  - the sample size (N)
  - the effect size (Δ)

"The probability of detecting a given effect size in a population from a sample of size N, using significance criterion α"

Standard Case

Sampling distribution if H_0 were true

Sampling distribution if H_A were true

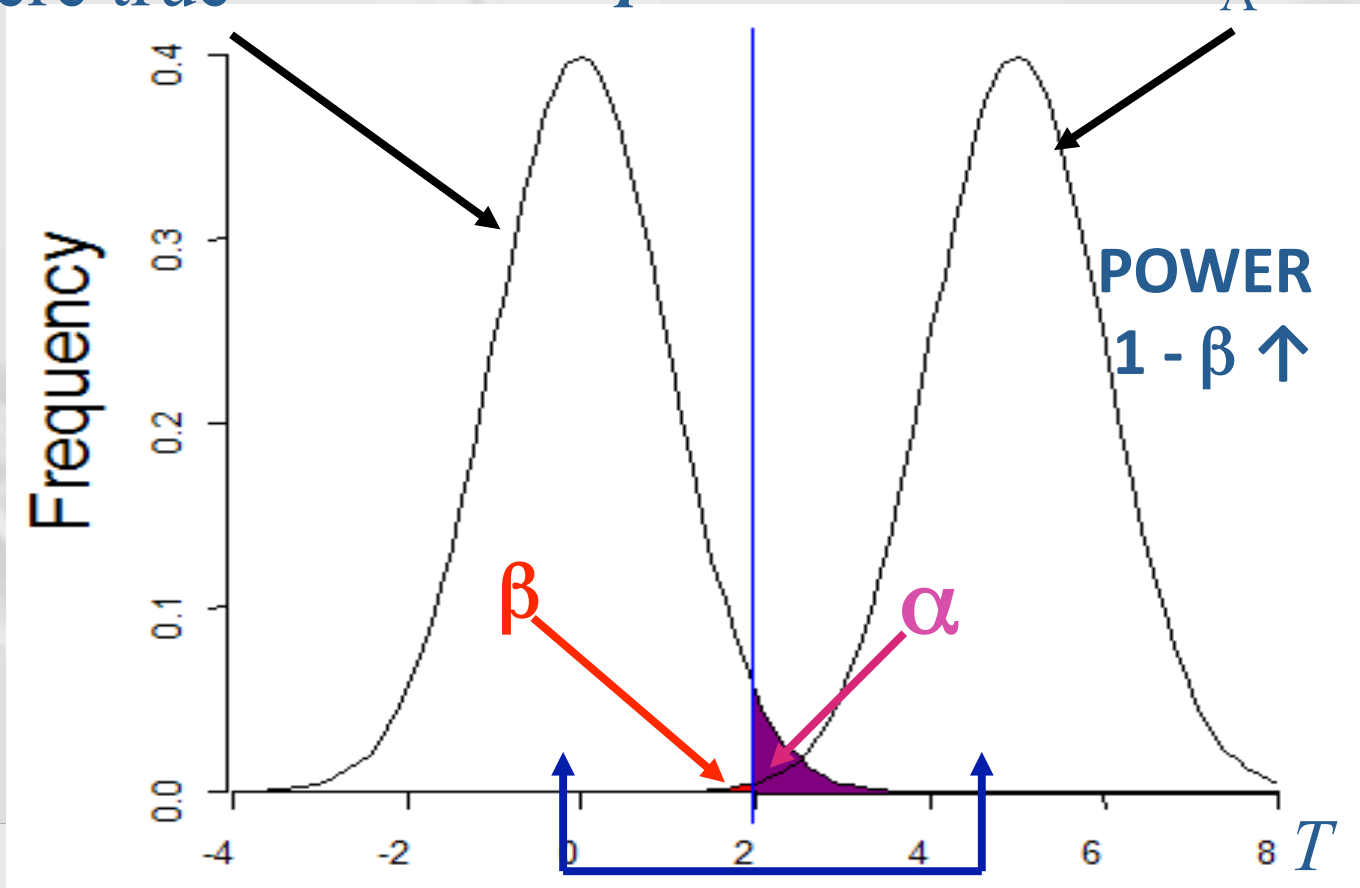


Increased effect size

Sampling distribution if H_0 were true

alpha 0.05

Sampling distribution if H_A were true

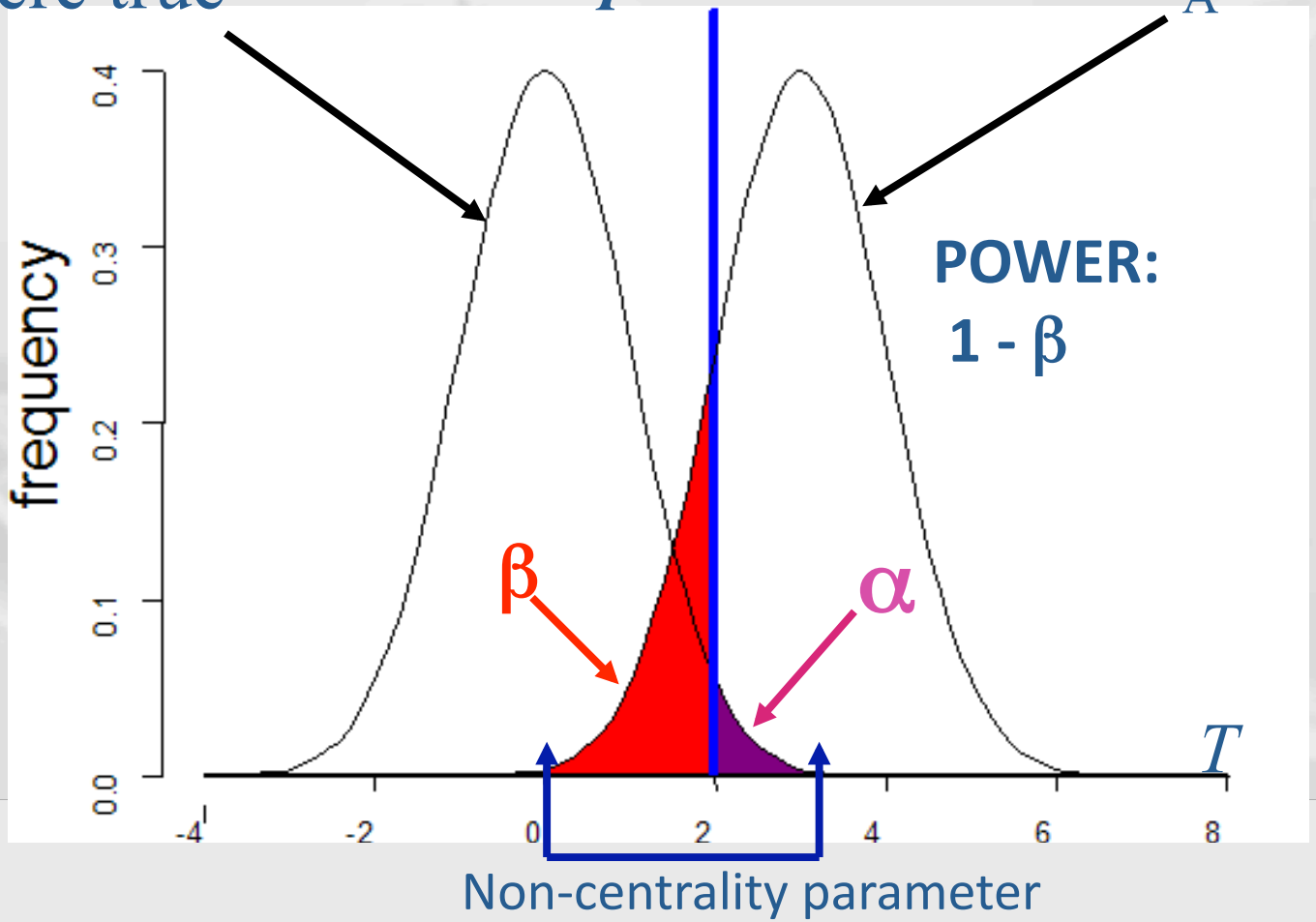


Non-centrality parameter

Standard Case

Sampling distribution if H_0 were true

Sampling distribution if H_A were true

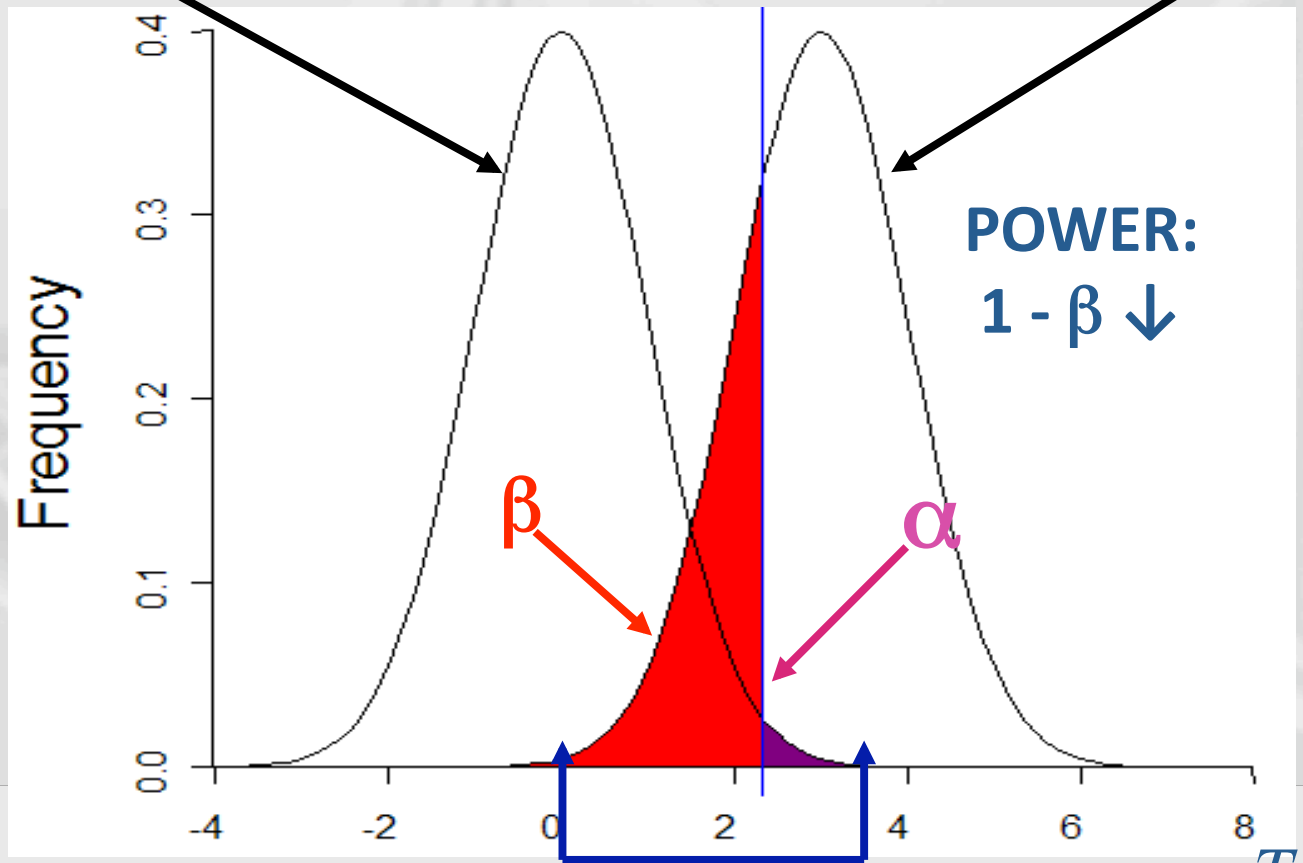


More conservative α

Sampling distribution if H_0 were true

alpha 0.01

Sampling distribution if H_A were true

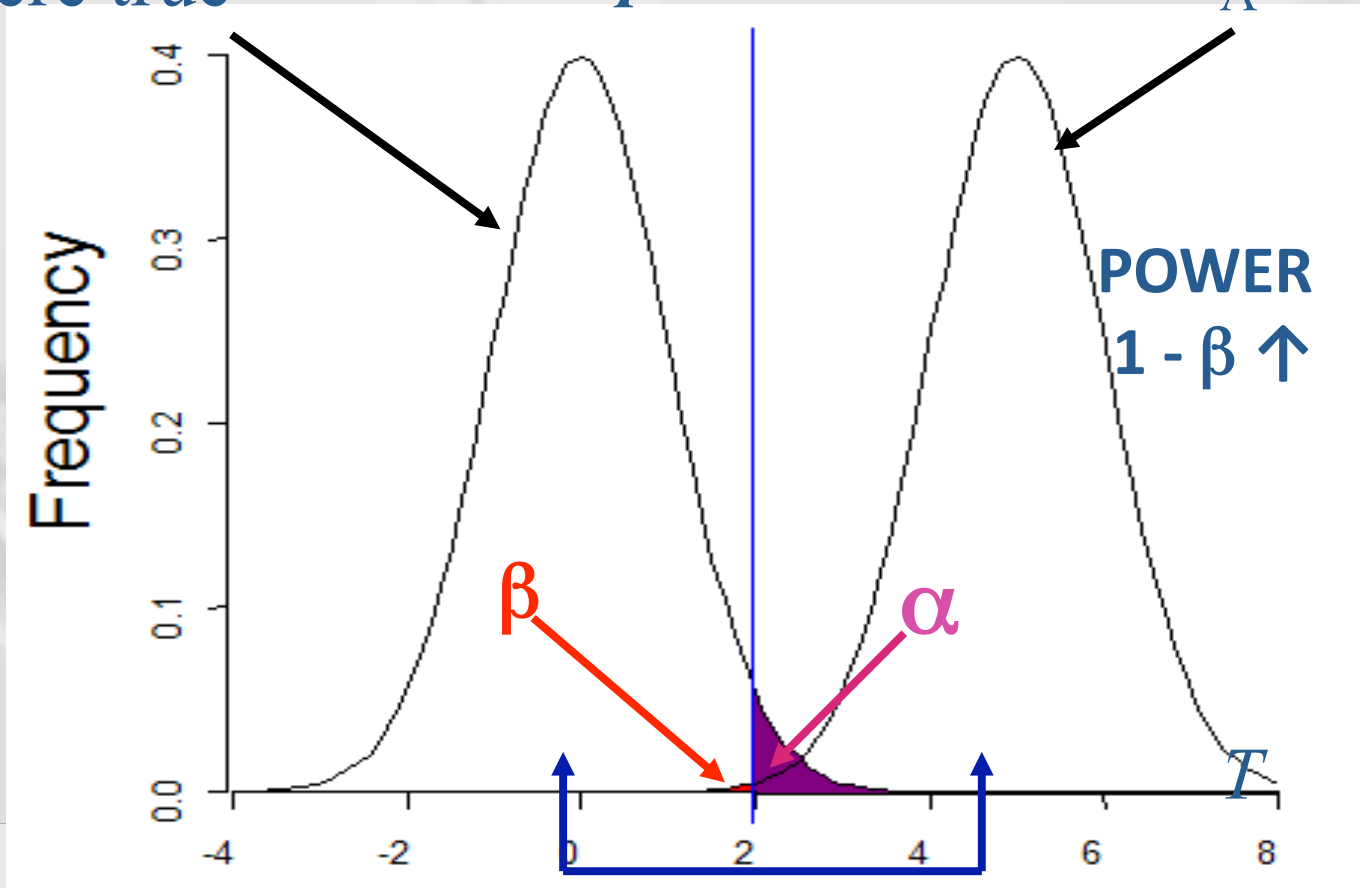


Increased sample size

Sampling distribution if H_0 were true

alpha 0.05

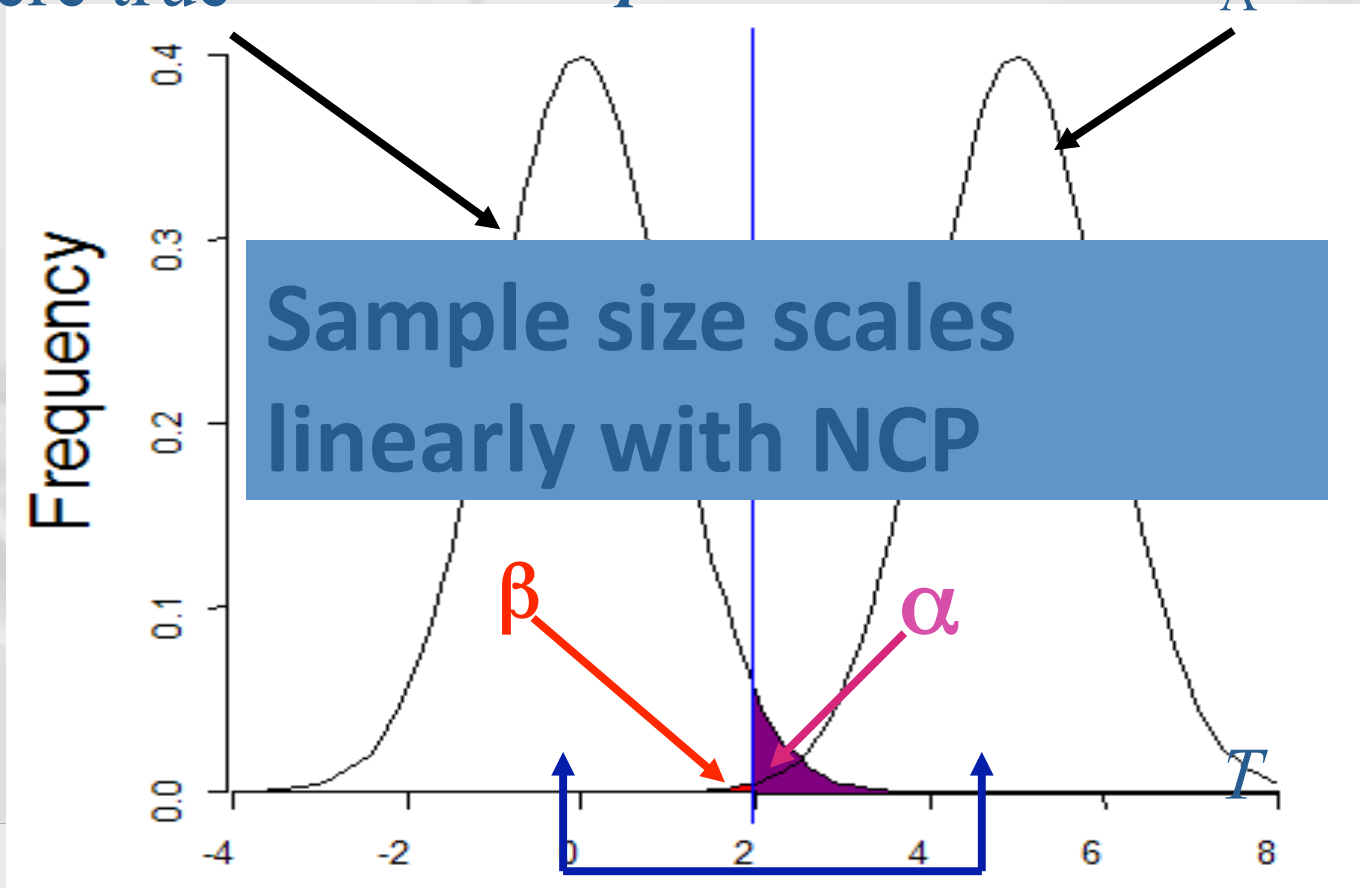
Sampling distribution if H_A were true



Increased sample size

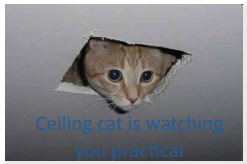
Sampling distribution if H_0 were true

Sampling distribution if H_A were true



Short practical on GPC

- Genetic Power Calculator is an online resource for carrying out basic power calculations
- For our 1st example we will use the probability function calculator to play with power
- http://pngu.mgh.harvard.edu/~purcell/gpc/
- Google GPC purcell



GPC Power Practical

Genetic Power Calculator - Mozilla Firefox

File Edit View History Bookmarks Tools Help

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

This site provides automated power analysis for variance components (VC) quantitative trait locus (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 article:

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 and singletons	Notes
TDT for discrete traits	Notes
TDT and parenTDT with ascertainment	Notes
Case-control for discrete traits *(new output features)*	Notes
TDT for threshold-selected quantitative traits	Notes
Case-control for threshold-selected quantitative traits	Notes
Epistasis power calculator	Notes
Probability Function Calculator	Notes
Various miscellaneous utilities	

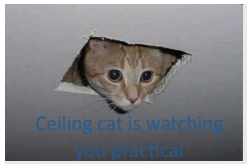
Click this link [no this isn't a banner ad]

Instructions for VC power calculations

All calculations are based upon formula derived in Sham et al (2000) [AJHG, 66, 1616-1630]. Users of this site who are unsure of the nature of the VC tests and power calculations are **strongly** advised to consult this article.

Parameters in probability function calculator

- Click on the link to probability function calculator
- 4 main terms:
- X: critical value of the chi-square
- P(X>x): Power
- df: degrees of freedom
- NCP: non-centrality parameter



GPC

Statistical Genetics Group - Mozilla Firefox

File Edit View History Bookmarks Tools Help

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

Probability Function Calculator

S. Purcell, 2000, 2005.

This site is designed to provide a calculator for the chi-squared and normal distributions. See below for notes on how to use the forms.

X	P(X>x)	df	NCP	
	?			Non-central chi-squared
?				Inverse non-central chi-squared
			?	NCP non-central chi-squared

x	P(X>x)	mean	SD	
	?			Normal cumulative
?				Inverse normal cumulative

X : critical value for statistic
P(X>x) : probability of being *above* the critical value
df : degrees of freedom
NCP : noncentrality parameter

For central chi-squared distribution, NCP = 0.
i.e. df=1, P(X>x) = 0.05, NCP = 0 gives X = 3.84146

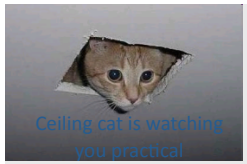
For standard normal distribution, mean = 0, sd = 1
i.e. mean = 0, sd = 1, P(X>x) = 0.025 give X = 1.95996

1) Fill in three
2) Click the button
3) Reveals the fourth

Find: exclude Next Previous Highlight all Match case

Bosoton: Wed 11:46 | UK: Wed 16:46 | Netherlands: Wed 17:46 | Hong Kong: Thu 00:46 | Los Angeles: Wed 08:46 | Done

start 11:46 AM



GPC

Statistical Genetics Group - Mozilla Firefox

File Edit View History Bookmarks Tools Help

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

Probability Function Calculator

S. Purcell, 2000, 2005.

This site is designed to provide a calculator for the chi-squared and normal distributions. See below for notes on how to use the forms.

X	P(X>x)	df	NCP	
3.85	?	1	10	Non-central chi-squared
?				Inverse non-central chi-squared
			?	NCP non-central chi-squared

x	P(X>x)	mean	SD	
	?			Normal cumulative
?				Inverse normal cumulative

X : critical value for statistic
P(X>x) : probability of being *above* the critical value
df : degrees of freedom
NCP : noncentrality parameter

For central chi-squared distribution, NCP = 0.
i.e. df=1, P(X>x) = 0.05, NCP = 0 gives X = 3.84146

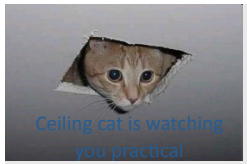
For standard normal distribution, mean = 0, sd = 1
i.e. mean = 0, sd = 1, P(X>x) = 0.025 give X = 1.95996

1) Fill in three
2) Click the button
3) Reveals the fourth

Find: exclude Next Previous Highlight all Match case

Bosoton: Wed 11:49 | UK: Wed 16:49 | Netherlands: Wed 17:49 | Hong Kong: Thu 00:49 | Los Angeles: Wed 08:49 | Done

start 11:49 AM



GPC

Statistical Genetics Group - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://pngu.mgh.harvard.edu/~purcell/cgi-bin/pdf.cgi

Most Visited Smart Bookmarks Free Hotmail Amazon.com: Statistic... The Arreat Summit - It... Share on Facebook

Gmail - GENEVA I... IBG workshop:2009:... IBG Index of /worksh... Wiley InterScien... Hulu - Videos Cypriots and U.N... Statistical Ge...

Probability Function Calculator

This site is designed to provide a calculator for the chi-squared and normal distributions. See below for notes on how to use the forms.

X	P(X>x)	df	NCP	
3.85	0.884957	1	10	Non-central chi-squared
?				Inverse non-central chi-squared
			?	NCP non-central chi-squared

X	P(X>x)	mean	SD	
	?			Normal cumulative
?				Inverse normal cumulative

X : critical value for statistic
P(X>x) : probability of being *above* the critical value
df : degrees of freedom
NCP : noncentrality parameter

For central chi-squared distribution, NCP = 0.
i.e. df=1, P(X>x) = 0.05, NCP = 0 gives X = 3.84146

For standard normal distribution, mean = 0, sd = 1
i.e. mean = 0, sd = 1, P(X>x) = 0.025 give X = 1.95996

1) Fill in three
2) Click the button
3) Reveals the fourth

Find: exclude Next Previous Highlight all Match case

Bosoton: Wed 11:50 UK: Wed 16:50 Netherlands: Wed 17:50 Hong Kong: Thu 00:50 Los Angeles: Wed 08:50 Done

start 62% 11:50 AM

Exercises

1) Find the power when NCP=5, degrees of freedom=1, and the critical X is 3.84
2) Find the NCP for power of .8, degrees of freedom=1 and critical X is 13.8

Answers

1) Power=0.608922, when NCP=5, degrees of freedom=1, and the critical X is 3.84
2) NCP=20.7613 when power of .8, degrees of freedom=1 and critical X is 13.8

Practical using GPC for association

Genetic Power Calculator - Mozilla Firefox

File Edit View History Bookmarks Tools Help

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

Most Visited Smart Bookmarks Free Hotmail Amazon.com: Statistic... The Arreat Summit - It... Share on Facebook

Gmail - Inbox (85... workshop:2009:... Index of /worksh... Wiley InterScien... Hulu - Videos Cypriots and U.N... Genetic Powe...

This site provides automated power analysis for variance components (VC) quantitative trait locus (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 article:

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 and singletons	Notes
TDT for discrete traits	Notes
TDT and parenTDT with ascertainment	Notes
Case-control for discrete traits *(new output features)*	Notes
TDT for threshold-selected quantitative traits	Notes
Case-control for threshold-selected quantitative traits	Notes
Epistasis power calculator	Notes
Probability Function Calculator	Notes
Various miscellaneous utilities	

These are all association tests

Each refers to a different study design

Instructions for VC power calculations

All calculations are based upon formula derived in Sham et al (2000) [AJHG, 66, 1616-1630]. Users of this site who are unsure of the nature of the VC tests and power calculations are **strongly** advised to consult this article.

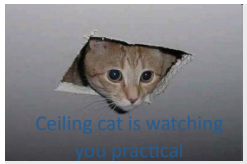
Find: exclude Next Previous Highlight all Match case

Bosoton: Wed 12:11 UK: Wed 17:11 Netherlands: Wed 18:11 Hong Kong: Thu 01:11 Los Angeles: Wed 09:11 Done

start 78% 12:11 PM

Question 1

- What case control sample size do we need to achieve genome-wide significance for an odds ratio of 1.2 in a multiplicative model and an allele frequency of 20% when we directly type the locus for a disease with 5% prevalence?



Question 1

Genetic Power Calculator - Mozilla Firefox

File Edit View History Bookmarks Tools Help

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

This site provides automated power analysis for variance components (VC) quantitative trait locus (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 article:

```
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 and singletons	Notes
TDT for discrete traits	Notes
TDT and parenTDT with ascertainment	Notes
Case-control for discrete traits *(new output features)*	Notes
TDT for threshold-selected quantitative traits	Notes
Case-control for threshold-selected quantitative traits	Notes
Epistasis power calculator	Notes
Probability Function Calculator	Notes
Various miscellaneous utilities	

Click this link

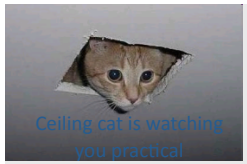
Instructions for VC power calculations

All calculations are based upon formula derived in Sham et al (2000) [AJHG, 66, 1616-1630]. Users of this site who are unsure of the nature of the VC tests and power calculations are **strongly** advised to consult this article.

Find: exclude Next Previous Highlight all Match case

Bosoton: Wed 12:11 UK: Wed 17:11 Netherlands: Wed 18:11 Hong Kong: Thu 01:11 Los Angeles: Wed 09:11 Done

start 78% 12:11 PM



Question 1

Statistical Genetics Group - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://pngu.mgh.harvard.edu/~purcell/gpc/cc2.html

Most Visited | Smart Bookmarks | Free Hotmail | Amazon.com: Statistic... | The Arreat Summit - It... | Share on Facebook

Gmail - Update: ... | IBG workshop:2009:... | IBG Index of /worksh... | Wiley InterScien... | Hulu - Videos | Cypriots and U.N... | Statistical Ge...

Genetic Power Calculator

Case - control for discrete traits

Allele frequency at the risk locus

High risk allele frequency (A)	:	0.2	(0 - 1)
Prevalence	:	.05	(0.0001 - 0.9999)
Genotype relative risk Aa	:	1.2	(>1)
Genotype relative risk AA	:	1.44	(>1)
D-prime	:	1	(0 - 1)
Marker allele frequency (B)	:	0.2	(0 - 1)
Number of cases	:	1000	(0 - 10000000)
Control : case ratio	:	1	(>0)

(1 = equal number of cases and controls)

☐ Unselected controls? (* see below)

User-defined type I error rate	:	5e-8	(0.00000001 - 0.5)
User-defined power: determine N (1 - type II error rate)	:	0.80	(0 - 1)

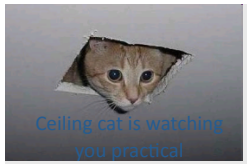
Process | Reset

Created by Shaun Purcell 24.Oct.2008

Find: exclude | Next | Previous | Highlight all | Match case

Bosoton: Wed 12:49 | UK: Wed 17:49 | Netherlands: Wed 18:49 | Hong Kong: Thu 01:49 | Los Angeles: Wed 09:49 | Stopped

start | 92% | 12:49 PM



Question 1

Statistical Genetics Group - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://pngu.mgh.harvard.edu/~purcell/gpc/cc2.html

Most Visited | Smart Bookmarks | Free Hotmail | Amazon.com: Statistic... | The Arreat Summit - It... | Share on Facebook

Gmail - Update: ... | IBG workshop:2009:... | IBG Index of /worksh... | Wiley InterScien... | Hulu - Videos | Cypriots and U.N... | Statistical Ge...

Genetic Power Calculator

Case - control for discrete traits

How common disease is

High risk allele frequency (A)	:	0.2	(0 - 1)
Prevalence	:	.05	(0.0001 - 0.9999)
Genotype relative risk Aa	:	1.2	(>1)
Genotype relative risk AA	:	1.44	(>1)
D-prime	:	1	(0 - 1)
Marker allele frequency (B)	:	0.2	(0 - 1)
Number of cases	:	1000	(0 - 10000000)
Control : case ratio	:	1	(>0)

(1 = equal number of cases and controls)

☐ Unselected controls? (* see below)

User-defined type I error rate	:	5e-8	(0.00000001 - 0.5)
User-defined power: determine N (1 - type II error rate)	:	0.80	(0 - 1)

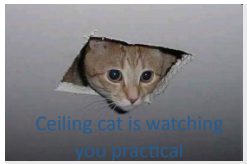
Process Reset

Created by Shaun Purcell 24.Oct.2008

Find: exclude Next Previous Highlight all Match case

Bosoton: Wed 12:49 | UK: Wed 17:49 | Netherlands: Wed 18:49 | Hong Kong: Thu 01:49 | Los Angeles: Wed 09:49 | Stopped

start | 2 F. | 2 W | 2 M. | 3 M. | 5 M. | RG... | ma... | silv... | 92% | 12:49 PM



Question 1

Statistical Genetics Group - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://pngu.mgh.harvard.edu/~purcell/gpc/cc2.html

Most Visited | Smart Bookmarks | Free Hotmail | Amazon.com: Statistic... | The Arreat Summit - It... | Share on Facebook

Gmail - Update: ... | IBG workshop:2009:... | IBG Index of /worksh... | Wiley InterScien... | Hulu - Videos | Cypriots and U.N... | Statistical Ge...

Genetic Power Calculator

Case - control for discrete traits

High risk allele frequency (A)	:	0.2	(0 - 1)
Prevalence	:	.05	(0.0001 - 0.9999)
Genotype relative risk Aa	:	1.2	(>1)
Genotype relative risk AA	:	1.44	(>1)
D-prime	:	1	(0 - 1)
Marker allele frequency (B)	:	0.2	(0 - 1)
Number of cases	:	1000	(0 - 10000000)
Control : case ratio	:	1	(>0)

(1 = equal number of cases and controls)

☐ Unselected controls? (* see below)

User-defined type I error rate	:	5e-8	(0.00000001 - 0.5)
User-defined power: determine N (1 - type II error rate)	:	0.80	(0 - 1)

Process Reset

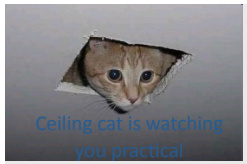
Created by Shaun Purcell 24.Oct.2008

Find: exclude Next Previous Highlight all Match case

Bosoton: Wed 12:49 | UK: Wed 17:49 | Netherlands: Wed 18:49 | Hong Kong: Thu 01:49 | Los Angeles: Wed 09:49 | Stopped

start | 92% | 12:49 PM

This is the relative risk—not the odds ratio. The OR is approximately equivalent to the RR for small values of RR.



Question 1

Statistical Genetics Group - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://pngu.mgh.harvard.edu/~purcell/gpc/cc2.html

Most Visited | Smart Bookmarks | Free Hotmail | Amazon.com: Statistic... | The Arreat Summit - It... | Share on Facebook

Gmail - Update: ... | IBG workshop:2009:... | IBG Index of /worksh... | Wiley InterScien... | Hulu - Videos | Cypriots and U.N... | Statistical Ge...

Genetic Power Calculator

Case - control for discrete traits

High risk allele frequency (A)	:	0.2	(0 - 1)
Prevalence	:	.05	(0.0001 - 0.9999)
Genotype relative risk Aa	:	1.2	(>1)
Genotype relative risk AA	:	1.44	(>1)
D-prime	:	1	(0 - 1)
Marker allele frequency (B)	:	0.2	(0 - 1)
Number of cases	:	1000	(0 - 10000000)
Control : case ratio	:	1	(>0)

(1 = equal number of cases and controls)

Unselected controls? (* see below)

User-defined type I error rate	:	5e-8	(0.00000001 - 0.5)
User-defined power: determine N (1 - type II error rate)	:	0.80	(0 - 1)

Process Reset

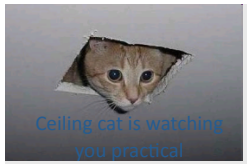
Created by Shaun Purcell 24.Oct.2008

Find: exclude Next Previous Highlight all Match case

Bosoton: Wed 12:49 | UK: Wed 17:49 | Netherlands: Wed 18:49 | Hong Kong: Thu 01:49 | Los Angeles: Wed 09:49 | Stopped

start | 92% | 12:49 PM

Risk of the AA genotype. Note that the model of risk is defined by the relationship between Aa and AA. We have a multiplicative model because 1.44 = 1.2*1.2.



Question 1

Statistical Genetics Group - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://pngu.mgh.harvard.edu/~purcell/gpc/cc2.html

Genetic Power Calculator

Case - control for discrete traits

High risk allele frequency (A)	:	0.2	(0 - 1)
Prevalence	:	.05	(0.0001 - 0.9999)
Genotype relative risk Aa	:	1.2	(>1)
Genotype relative risk AA	:	1.44	(>1)
D-prime	:	1	(0 - 1)
Marker allele frequency (B)	:	0.2	(0 - 1)
Number of cases	:	1000	(0 - 10000000)
Control : case ratio	:	1	(>0)

(1 = equal number of cases and controls)

☐ Unselected controls? (* see below)

User-defined type I error rate	:	5e-8	(0.00000001 - 0.5)
User-defined power: determine N (1 - type II error rate)	:	0.80	(0 - 1)

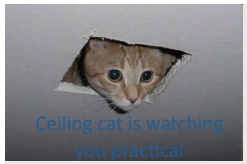
Process Reset

Created by Shaun Purcell 24.Oct.2008

Find: exclude Next Previous Highlight all Match case

Bosoton: Wed 12:49 UK: Wed 17:49 Netherlands: Wed 18:49 Hong Kong: Thu 01:49 Los Angeles: Wed 09:49 Stopped

The LD statistic D' which represents recombination patterns historically. D' + allele frequency at the typed locus information yields r^2



Question 1

Statistical Genetics Group - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://pngu.mgh.harvard.edu/~purcell/gpc/cc2.html

Genetic Power Calculator

Case - control for discrete traits

Sample size for cases

High risk allele frequency (A)	:	0.2	(0 - 1)
Prevalence	:	.05	(0.0001 - 0.9999)
Genotype relative risk Aa	:	1.2	(>1)
Genotype relative risk AA	:	1.44	(>1)
D-prime	:	1	(0 - 1)
Marker allele frequency (B)	:	0.2	(0 - 1)
Number of cases	:	1000	(0 - 10000000)
Control : case ratio	:	1	(>0)

(1 = equal number of cases and controls)

☐ Unselected controls? (* see below)

User-defined type I error rate	:	5e-8	(0.00000001 - 0.5)
User-defined power: determine N (1 - type II error rate)	:	0.80	(0 - 1)

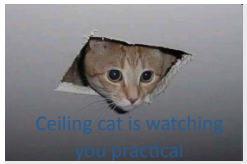
Process Reset

Created by Shaun Purcell 24.Oct.2008

Find: exclude Next Previous Highlight all Match case

Bosoton: Wed 12:49 UK: Wed 17:49 Netherlands: Wed 18:49 Hong Kong: Thu 01:49 Los Angeles: Wed 09:49 Stopped

start 92% 12:49 PM



Question 1

Statistical Genetics Group - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://pngu.mgh.harvard.edu/~purcell/gpc/cc2.html

Most Visited | Smart Bookmarks | Free Hotmail | Amazon.com: Statistic... | The Arreat Summit - It... | Share on Facebook

Gmail - Update: ... | IBG workshop:2009:... | IBG Index of /worksh... | Wiley InterScien... | Hulu - Videos | Cypriots and U.N... | Statistical Ge...

Genetic Power Calculator

Case - control for discrete traits

Ratio of Controls to Cases

High risk allele frequency (A)	:	0.2	(0 - 1)
Prevalence	:	.05	(0.0001 - 0.9999)
Genotype relative risk Aa	:	1.2	(>1)
Genotype relative risk AA	:	1.44	(>1)
D-prime	:	1	(0 - 1)
Marker allele frequency (B)	:	0.2	(0 - 1)
Number of cases	:	1000	(0 - 10000000)
Control : case ratio	:	1	(>0) (1 = equal number of cases and controls)
		☐	Unselected controls? (* see below)
User-defined type I error rate	:	5e-8	(0.00000001 - 0.5)
User-defined power: determine N (1 - type II error rate)	:	0.80	(0 - 1)

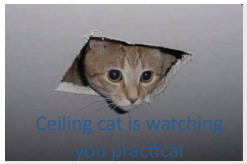
Process | Reset

Created by Shaun Purcell 24.Oct.2008

Find: exclude | Next | Previous | Highlight all | Match case

Bosoton: Wed 12:49 | UK: Wed 17:49 | Netherlands: Wed 18:49 | Hong Kong: Thu 01:49 | Los Angeles: Wed 09:49 | Stopped

start | 92% | 12:49 PM



Question 1

Statistical Genetics Group - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://pngu.mgh.harvard.edu/~purcell/gpc/cc2.html

Genetic Power Calculator

Case - control for discrete traits

High risk allele frequency (A)	:	0.2	(0 - 1)
Prevalence	:	.05	(0.0001 - 0.9999)
Genotype relative risk Aa	:	1.2	(>1)
Genotype relative risk AA	:	1.44	(>1)
D-prime	:	1	(0 - 1)
Marker allele frequency (B)	:	0.2	(0 - 1)
Number of cases	:	1000	(0 - 10000000)
Control : case ratio	:	1	(>0)

(1 = equal number of cases and controls)

☐ Unselected controls? (* see below)

User-defined type I error rate	:	5e-8	(0.00000001 - 0.5)
User-defined power: determine N (1 - type II error rate)	:	0.80	(0 - 1)

Process Reset

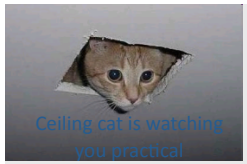
Created by Shaun Purcell 24.Oct.2008

Find: exclude Next Previous Highlight all Match case

Bosoton: Wed 12:49 UK: Wed 17:49 Netherlands: Wed 18:49 Hong Kong: Thu 01:49 Los Angeles: Wed 09:49 Stopped

Genome-wide significance threshold

We'll learn about this later in the session



Question 1

Statistical Genetics Group - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://pngu.mgh.harvard.edu/~purcell/gpc/cc2.html

Most Visited | Smart Bookmarks | Free Hotmail | Amazon.com: Statistic... | The Arreat Summit - It... | Share on Facebook

Gmail - Update: ... | IBG workshop:2009:... | IBG Index of /worksh... | Wiley InterScien... | Hulu - Videos | Cypriots and U.N... | Statistical Ge...

Genetic Power Calculator

Case - control for discrete traits

High risk allele frequency (A)	:	0.2	(0 - 1)
Prevalence	:	.05	(0.0001 - 0.9999)
Genotype relative risk Aa	:	1.2	(>1)
Genotype relative risk AA	:	1.44	(>1)
D-prime	:	1	(0 - 1)
Marker allele frequency (B)	:	0.2	(0 - 1)
Number of cases	:	1000	(0 - 10000000)
Control : case ratio	:	1	(>0) (1 = equal number of cases and controls)
		☐	Unselected controls? (* see below)
User-defined type I error rate	:	5e-8	(0.00000001 - 0.5)
User-defined power: determine N (1 - type II error rate)	:	0.80	(0 - 1)

Process Reset

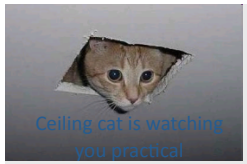
Created by Shaun Purcell 24.Oct.2008

Find: exclude Next Previous Highlight all Match case

Bosoton: Wed 12:49 | UK: Wed 17:49 | Netherlands: Wed 18:49 | Hong Kong: Thu 01:49 | Los Angeles: Wed 09:49 | Stopped

start | 92% | 12:49 PM

Power level—what we're interested in observing



Question 1

Statistical Genetics Group - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://pngu.mgh.harvard.edu/~purcell/gpc/cc2.html

Most Visited | Smart Bookmarks | Free Hotmail | Amazon.com: Statistic... | The Arreat Summit - It... | Share on Facebook

Gmail - Update: ... | IBG workshop:2009:... | IBG Index of /worksh... | Wiley InterScien... | Hulu - Videos | Cypriots and U.N... | Statistical Ge...

Genetic Power Calculator

Case - control for discrete traits

High risk allele frequency (A)	:	0.2	(0 - 1)
Prevalence	:	.05	(0.0001 - 0.9999)
Genotype relative risk Aa	:	1.2	(>1)
Genotype relative risk AA	:	1.44	(>1)
D-prime	:	1	(0 - 1)
Marker allele frequency (B)	:	0.2	(0 - 1)
Number of cases	:	1000	(0 - 10000000)
Control : case ratio	:	1	(>0) (1 = equal number of cases and controls)
		☐ Unselected controls? (* see below)	
User-defined type I error rate	:	5e-8	(0.00000001 - 0.5)
User-defined power: determine N (1 - type II error rate)	:	0.80	(0 - 1)

Process Reset

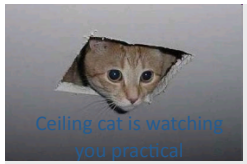
Click here to process

Created by Shaun Purcell 24.Oct.2008

Find: exclude Next Previous Highlight all Match case

Bosoton: Wed 12:49 | UK: Wed 17:49 | Netherlands: Wed 18:49 | Hong Kong: Thu 01:49 | Los Angeles: Wed 09:49 | Stopped

start | 92% | 12:49 PM



Answer 1

Genetic Power Calculator - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://pngu.mgh.harvard.edu/~purcell/cgi-bin/cc2k.cgi

Most Visited Smart Bookmarks Free Hotmail Amazon.com: Statistic... The Arreat Summit - It... Share on Facebook

Gmail - Update: ... workshop:2009:... Index of /worksh... Wiley InterScien... Hulu - Videos Cypriots and U.N... Genetic Powe...

Alpha	Power	N cases for 80% power
0.1	0.4374	2803
0.05	0.3177	3559
0.01	0.1377	5296
0.001	0.0355	7742
5e-08	3.674e-05	17958

Case-control statistics: general 2 df test (BB versus Bb versus bb)
Sample NCP = 6.216

Alpha	Power	N cases for 80% power
0.1	0.716	1240
0.05	0.6002	1550
0.01	0.3609	2233
0.001	0.1451	3163
5e-08	0.0007464	6920

Case-control statistics: allelic 1 df test (B versus b)
Sample NCP = 6.224

Scroll to the bottom for answer

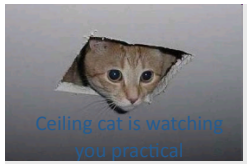
Alpha	Power	N cases for 80% power
0.1	0.8024	993
0.05	0.7037	1260
0.01	0.4677	1876
0.001	0.2131	2743
5e-08	0.001557	6362

Controls are selected (i.e. screened for not being a case)

Find: exclude Next Previous Highlight all Match case

Bosoton: Wed 13:02 UK: Wed 18:02 Netherlands: Wed 19:02 Hong Kong: Thu 02:02 Los Angeles: Wed 10:02 Done

start 2 F.. 2 W.. 2 M.. 2 M.. 2 M.. RG... mac... silv... 94% 1:02 PM



Answer 1

Genetic Power Calculator - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://pngu.mgh.harvard.edu/~purcell/cgi-bin/cc2k.cgi

Most Visited Smart Bookmarks Free Hotmail Amazon.com: Statistic... The Arreat Summit - It... Share on Facebook

Gmail - Update: ... IBG workshop:2009:... IBG Index of /worksh... Wiley InterScien... Hulu - Videos Cypriots and U.N... Genetic Powe...

Alpha	Power	N cases for 80% power
0.1	0.4374	2803
0.05	0.3177	3559
0.01	0.1377	5296
0.001	0.0355	7742
5e-08	3.674e-05	17958

Case-control statistics: general 2 df test (BB versus Bb versus bb)
Sample NCP = 6.216

Alpha	Power	N cases for 80% power
0.1	0.716	1240
0.05	0.6002	1550
0.01	0.3609	2233
0.001	0.1451	3163
5e-08	0.0007464	6920

Case-control statistics: allelic 1 df test (B versus b)
Sample NCP = 6.224

Scroll to the bottom for answer

Alpha	Power	N cases for 80% power
0.1	0.8024	993
0.05	0.7037	1260
0.01	0.4677	1876
0.001	0.2131	2743
5e-08	0.001557	6362

Controls are selected (i.e. screened for not being a case)

Find: exclude Next Previous Highlight all Match case

Bosoton: Wed 13:02 UK: Wed 18:02 Netherlands: Wed 19:02 Hong Kong: Thu 02:02 Los Angeles: Wed 10:02 Done

start 2 F.. 2 W.. 2 M.. 2 M.. 2 M.. RG... mac... silv... 94% 1:02 PM

6,362 case samples required: total sample size 12,724

Questions on your own

- For the same model as above, find the total sample size required for a TDT
  - Hint: use TDT for discrete traits
  - Try for different effect sizes and models (e.g. dominance)
- What is the effect of degrading LD in case-control data?
  - Change the D' and keep allele freq the same
  - Change allele freq and keep D' the same
- How well does the additive model capture a dominance only effect?
- Should you use 2x population controls vs 1x screened controls
  - For a prevalence of 5% and for a prevalence of 25%?

Answers

- Additive
  - Total case number for CC: 6,362
  - Total case number for TDT: 7,079
- Dominance only
  - RR: 1; 1; 1.44
  - 30,595 cases for CC
  - 33,950 cases for TDT