

Homework 2 - Association Statistics

COMSCI/HUMGEN 124/224

Due: April 12th, 2016, 6 pm

All assignments must be submitted by a hard copy to TA's office by 6 pm on the day they are due.

Problem 1 - Association Study at a Single SNP

Part A - Calculating the Non-Centrality Parameter

Use the formulas described in Lectures 2 & 3 to compute the non-centrality parameters. Compute the non-centrality parameters for minor allele frequencies 0.05, 0.2 and 0.4, for relative risks of 1.5, 2.0 and 3.0, for total individual numbers in the cases and controls of 500 and 1000. Assume that disease prevalence (F) is very low. You must compute the non-centrality parameter using R and show a transcript of your code and results.

You can enter the results into these tables and include them in the homework submission,

500 individuals		Allele frequency		
		0.05	0.2	0.4
Relative risk	1.5			
	2.0			
	3.0			

Table 1: 500 Individuals

1000 individuals		Allele frequency		
		0.05	0.2	0.4
Relative risk	1.5			
	2.0			
	3.0			

Table 2: 1000 Individuals

Part B - Calculating the Power

Now compute the power of these studies assuming a p-value threshold of 0.05. You must compute the power using R and show a transcript of your code and results. You should re-use the R code you wrote for computing non-centrality parameter in Part A.

You can enter the results into these tables and include them in the homework submission,

500 individuals		Allele frequency		
		0.05	0.2	0.4
Relative risk	1.5			
	2.0			
	3.0			

Table 3: 500 Individuals

1000 individuals		Allele frequency		
		0.05	0.2	0.4
Relative risk	1.5			
	2.0			
	3.0			

Table 4: 1000 Individuals

Part C - Calculating the number of individuals

Using the same relative risks and minor allele frequencies as in Part A and B, compute the number of individuals needed to achieve 80% power for each pair of relative risk and minor allele frequency. You should use the R code you wrote for Part B, and try different values of the number of individuals to achieve 80% power roughly (79% ~ 81%).

You can enter the results into these tables and include them in the homework submission,

		Allele frequency		
		0.05	0.2	0.4
Relative risk	1.5			
	2.0			
	3.0			

Table 5: 80% power

Problem 2 - Unbalanced Cases and Controls

Part A

Assume that you have a N total individuals in a balanced case and control study (i.e. $N/2$ case individuals and $N/2$ control individuals). The non-centrality parameter for this study is $\lambda_A \sqrt{N}$.

On the other hand, if the number of cases and controls are not equal, the non-centrality parameter is different. If there are $N^+/2$ cases and $N^-/2$ controls, the non centrality parameter is $\lambda_A \sqrt{\frac{2(N^+N^-)}{N^++N^-}}$.

Now assume you are designing a study with three times the number of cases as controls. How large does your study have to be (as a factor of N) so that you achieve the same power as a balanced study with N individuals?

Part B

Assume that you have $N^+/2$ cases and an unlimited number of controls. What is the size of a balanced study (i.e. equal numbers of cases and controls) that would have equivalent power? Show your derivation. (Hint: First solve for the noncentrality parameter if you have a very large number of controls. Try using 1,000,000.)

Part C (Grad Students Only)

Derive the non-centrality parameter for unbalanced cases and controls above. Describe the precise approximation assumption you need to make.