Homework 2 – Association Statistics

Jacob Nisnevich

April 11, 2016

1 Association Study at a Single SNP

1.1 Calculating the Non-Centrality Parameter

500 individuals	Allele Frequency		
Relative Frequency	0.05	0.2	0.4
1.5	1.52396	2.70666	3.178209
2.0	2.756259	4.767313	5.423261
3.0	4.697591	7.784989	8.451543

Table 1: 500 Indivduals

1000 individuals	Allele Frequency		
Relative Frequency	0.05	0.2	0.4
1.5	2.155205	3.827795	4.494666
2.0	3.897938	6.741999	7.66965
3.0	6.643397	11.00964	11.95229

Table 2: 1000 Indivduals

1.2 Calculating the power

500 individuals	Allele Frequency		
Relative Frequency	0.05	0.2	0.4
1.5	0.331664	0.7723779	0.8884346
2.0	0.7870708	0.9975024	0.9997332
3.0	0.9969058	1	1

Table 3: 500 Indivduals

1000 individuals	Allele Frequency		
Relative Frequency	0.05	0.2	0.4
1.5	0.5774172	0.9691072	0.9943728
2.0	0.9736868	0.9999991	1
3.0	0.9999986	1	1

Table 4: 1000 Indivduals

1.3 Calculating the number of individuals

	Allele Frequency		
Relative Frequency	0.05	0.2	0.4
1.5	1690	536	389
2.0	517	173	134
3.0	178	65	55

Table 5: 80% Power

1.4 R Code

```r
#!/usr/bin/env Rscript

p_a_plus <- function(p, gamma) {
  return((gamma * p) / ((gamma - 1) * p + 1))
}

non_centrality <- function(p, gamma, N) {
  p_a_plus <- p_a_plus(p, gamma)
  p_a_minus <- p
  p_a <- mean(c(p_a_plus, p_a_minus))

  return((p_a_plus - p_a_minus) / (sqrt(2 / N) * sqrt(p_a * (1 - p_a))))
}

power <- function(non_centrality_param, alpha) {
  return(pnorm(qnorm(alpha / 2) + non_centrality_param) + 1 - pnorm(-1 * qnorm(alpha / 2) +
      non_centrality_param))
}

minor_allele_frequencies <- c(0.05, 0.2, 0.4)
relative_risks <- c(1.5, 2, 3)
alpha <- 0.05
n_vals <- c(500, 1000)

cat("Part 1\n\n")

for (N in n_vals) {
  for (gamma in relative_risks) {
    for (p in minor_allele_frequencies) {
      non_centrality_param <- non_centrality(p, gamma, N)
      cat(non_centrality_param)
      cat("\n")
    }
  }
  cat("\n")
}

cat("Part 2\n\n")

for (N in n_vals) {
  for (gamma in relative_risks) {
    for (p in minor_allele_frequencies) {
      non_centrality_param <- non_centrality(p, gamma, N)
      power_val <- power(non_centrality_param, alpha)
      cat(power_val)
      cat("\n")
    }
  }
  cat("\n")
}

cat("Part 3\n\n")
```

```
52
53 for (gamma in relative_risks) {
54   for (p in minor_allele_frequencies) {
55     for (N in (1:3000)) {
56       non_centrality_param <- non_centrality(p, gamma, N)
57       power_val <- power(non_centrality_param, alpha)
58
59       if (power_val > 0.8) {
60         cat(N)
61         cat("\n")
62         break
63       }
64     }
65   }
66 }
```

2 Unbalanced Cases and Controls

2.1 3 Times as Many Cases as Controls

For two studies to have equivalent power, they must have equivalent non-centrality parameters. It is also known that $N^+ = 3N^-$ Therefore:

$$\lambda_A \cdot \sqrt{N} = \lambda_A \cdot \sqrt{\frac{2N^+N^-}{N^+ + N^-}}$$

$$\sqrt{N} = \sqrt{\frac{2N^+N^-}{N^+ + N^-}}$$

$$N = \frac{2N^+N^-}{N^+ + N^-}$$

$$N = \frac{2(3N^-)(N^-)}{3N^- + N^-}$$

$$N = \frac{6(N^-)^2}{4N^-}$$

$$N = \frac{3}{2}N^-$$

$$\frac{2}{3}N = N^-$$

By substition we can also see that $N^+ = 2N$. As the total size of the study is $\dfrac{N^+}{2} + \dfrac{N^-}{2}$, the final total in terms of N is $\dfrac{1}{3}N + N = \dfrac{4}{3}N$.

2.2 Unlimited Number of Controls

Once again, we can set the two non-centrality parameters equal to compute the size of a study required for equivalent power. Skipping some simplification from the privous part we get the following limit:

$$N = \lim_{N^- \to \infty} \frac{2N^+N^-}{N^+ + N^-}$$

$$N = 2N^+$$