p8119_hw2_jsg2145

Jared Garfinkel

10/26/2020

```
tibble("Chapter" = c(7, 8, 9), "Problems" = c("1, 2, 3, 4", "7", "1, 2, 3"))
```

```
## # A tibble: 3 x 2
##   Chapter Problems
##     <dbl> <chr>
## 1       7 1, 2, 3, 4
## 2       8 7
## 3       9 1, 2, 3
```

Chapter 7

Exercise 1

We test for a codominant mode of inheritance using an odds ratio.

```
data = tibble(GM = c("present", "not present"), nsubjects = c(293, 4627), cases = c(23, 1343))
data
```

```
## # A tibble: 2 x 3
##   GM          nsubjects cases
##   <chr>           <dbl> <dbl>
## 1 present           293    23
## 2 not present      4627  1343
```

```
tidy_dat = tibble(exposed = c(23, 293-23), unexposed = c(1343, 4627-1343))
tidy_dat
```

```
## # A tibble: 2 x 2
##   exposed unexposed
##     <dbl>     <dbl>
## 1      23      1343
## 2     270      3284
```

```
OR = 23*(4627-1343)/(293-23)/1343
OR
```

```
## [1] 0.2083009
```

The odds ratio is 0.208.

```
var_log_OR = 1/23 + 1/1343 + 1/270 + 1/3284
var_log_OR
```

```
## [1] 0.04823107
```

```
SE = sqrt(var_log_OR)
lower = exp(log(OR)-SE*dnorm(0.975))
lower
```

```
## [1] 0.1972585
```

```
upper = exp(log(OR)+SE*dnorm(0.975))
upper
```

```
## [1] 0.2199615
```

The confidence interval is (0.197, 0.22). This indicates that there is evidence to suggest that there is a reduced odds of having the exposure in the diseased compared to the non-diseased.

Problem 2

Part a

Compute the test statistics for the additive model and the dominant model and compare.

```
df = tibble(disease = c("acne_patient", "control"), GG = c(66, 99), GA = c(43, 15), AA = c(4, 0))
df
```

```
## # A tibble: 2 x 4
##   disease         GG    GA    AA
##   <chr>        <dbl> <dbl> <dbl>
## 1 acne_patient    66    43     4
## 2 control         99    15     0
```

```
dom_df = tibble(disease = c("acne patient", "control"), `AA or GA` = (pull(df, AA) + pull(df, GA)), GG =
dom_df
```

```
## # A tibble: 2 x 3
##   disease      `AA or GA`    GG
##   <chr>             <dbl> <dbl>
## 1 acne patient         47    66
## 2 control              15    99
```

```
chisq.test(dom_df[,-1]) # p = <0.001
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  dom_df[, -1]
## X-squared = 21.702, df = 1, p-value = 3.184e-06
```

```
allele_df = tibble(disease = c("acne patient", "control"), A = c(2*4+83, 2*0+15), G = c(2*66+43, 2*99+15
allele_df
```

```
## # A tibble: 2 x 3
##   disease          A     G
##   <chr>        <dbl> <dbl>
## 1 acne patient    91   175
## 2 control         15   213
```

```
chisq.test(allele_df[,-1]) # p < 0.001
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  allele_df[, -1]
## X-squared = 53.991, df = 1, p-value = 2.014e-13
```

These two tests yield similar answers. They both show a high degree of significance in the dominant and additive models.

Part b

```
var_log_OR = 1/47+1/66+1/15+1/99
log_OR = log(47*99/66/15)
lower = exp(log_OR-sqrt(var_log_OR)*dnorm(0.975))
upper = exp(log_OR+sqrt(var_log_OR)*dnorm(0.975))
```

The confidence interval of the dominant odds ratio is (4.324, 5.109). This makes sense since the chi squared tests yield significant results.

Part c

```
df
```

```
## # A tibble: 2 x 4
##   disease         GG    GA    AA
##   <chr>        <dbl> <dbl> <dbl>
## 1 acne_patient    66    43     4
## 2 control         99    15     0
```

```
rec_df = tibble(disease = pull(df, disease), `GG or GA` = pull(df, GG) + pull(df, GA), AA = pull(df, AA
rec_df
```

```
## # A tibble: 2 x 3
##   disease      `GG or GA`    AA
##   <chr>             <dbl> <dbl>
## 1 acne_patient        109     4
## 2 control             114     0
```

```
fisher.test(rec_df[,-1]) # p = 0.0597
```

```
##
##  Fisher's Exact Test for Count Data
##
## data:  rec_df[, -1]
## p-value = 0.05977
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  0.000000 1.485382
## sample estimates:
## odds ratio
##          0
```

This Fisher exact test shows the recessive model is marginally significant at p = 0.0597. However, this is much greater than the significance levels for either the dominant model or the additive model.

Problem 3

Modes of inheritance: Dominant, codominant, recessive

```
df_dat = tibble(disease = c("case", "control"), x0 = c(500, 521), x1 = c(350, 270), x2 = c(120, 130), t
df_dat
```

```
## # A tibble: 2 x 5
##   disease    x0    x1    x2 total
##   <chr>   <dbl> <dbl> <dbl> <dbl>
## 1 case      500   350   120   970
## 2 control   521   270   130   921
```

$$K = \gamma_2 p_D^2 + f_0(1 - p_D^2)$$

The odds ratios are relative to a baseline, x0.

```
OR0 = 1
OR1 = 350*521/500/270
OR2 = 120*521/500/130
OR_df = cbind(OR0, OR1, OR2)
round(OR_df, 3)
```

```
##      OR0   OR1   OR2
## [1,]   1 1.351 0.962
```

The odds ratio for x = 1 is 1.351 and for x = 2, OR = 0.962.

The confidence intervals are as follows:

```
CI_OR = function(df = df_dat, x = x1, a, b, c, d) {
  b = pull(df, x0)[[1]]
  d = pull(df, x0)[[2]]
  a = pull(df, x)[[1]]
  c = pull(df, x)[[2]]
  var_log_OR = 1/a+1/b+1/c+1/d
  log_OR = log(a*d/b/c)
  lower = exp(log_OR-sqrt(var_log_OR)*dnorm(0.975))
  upper = exp(log_OR+sqrt(var_log_OR)*dnorm(0.975))
  CI = cbind(lower, upper)
  return(CI)
}

CI_x0 = CI_OR(x = "x0")
CI_x1 = CI_OR(x = "x1")
CI_x2 = CI_OR(x = "x2")
```

The confidence intervals are x0: 1 (0.978, 1.022), x1: 1.351 (1.317, 1.385), and x2: 0.962 (0.929, 0.996).

```
df_dat
```

```
## # A tibble: 2 x 5
##   disease    x0    x1    x2 total
##   <chr>   <dbl> <dbl> <dbl> <dbl>
## 1 case      500   350   120   970
## 2 control   521   270   130   921
```

```
dom_df = tibble(disease = pull(df_dat, disease), x0 = pull(df_dat, x0), `x1 or x2` = pull(df_dat, x1) +
chisq.test(dom_df[,-1]) # p = 0.032
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  dom_df[, -1]
## X-squared = 4.5976, df = 1, p-value = 0.03202
```

```
rec_df = tibble(disease = pull(df_dat, disease), `x0 or x1` = pull(df_dat, x0) + pull(df_dat, x1), x2 =
chisq.test(rec_df[,-1]) # 0.2932
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  rec_df[, -1]
## X-squared = 1.105, df = 1, p-value = 0.2932
```

```r
chisq.test(df_dat[,-1]) # p = 0.01951
```

```
##
##  Pearson's Chi-squared test
##
## data:  df_dat[, -1]
## X-squared = 9.8915, df = 3, p-value = 0.01951
```

The dominant and additive model tests are significant, but the recessive model test is not. This indicates that a dominant or additive model may be more appropriate for this data.

Problem 4

Definitions:

$$n_{DSL} = 2(z_{(1-\beta)} + z_{(1-\alpha/2)})^2 p_D(1-p_D)/\Delta_D^2$$

$$\Delta_D = (p_{D|cases} - p_{D|controls})$$

Given:

$\gamma_1 = 1.3$

It follows that under an additive model, $\gamma_2 = 2 * \gamma_1 - 1 = 1.6$

So, $f_1 = 1.3 * f_0$ and $f_2 = 1.6 * f_0$

Since $\sum_i(f_i) = 1$,

$(1.3 + 1.6 + 1) * f_0 = 1, f_0 = 1/3.9 = 0.26 =$ p(disease | i copies of the allele)

$f_1 = 1.3 * 0.26 = 0.338$

$f_2 = 1.6 * 0.26 = 0.416$

risk of colon cancer (K) = .04

power = 0.8

$p_D = 0.55$

So,

$q_{cases} = f_2 g_2/K = 0.416 * 0.55^2/0.04 = 3.146$

$q_{controls} = (1 - f_2)g_2/Q = 0.184$

q = $\frac{r*q_{case}+s*q_{control}}{n}$ = (3.146 + 0.184)/2 = 1.665

$\Delta_D = (q_{cases} - q_{controls})$ = 3.146 - 0.184 = 2.962

Assume: $\alpha = 0.05$

r = s = $\frac{2*(z_{(1-\beta)}+z_{(1-\alpha/2)})^2*q(1-q)}{\Delta^2}$

Chapter 8

Problem 7

$\hat{\lambda}$

```
data = "5.112124234 0.827057943 3.158134984 3.395351358 0.056900096 0.878446231 4.955161751 0.127185994
  str_replace_all(., " ", ",")
data2 = data %>%
  str_split_fixed(., ",", n = 20) %>%
  as.numeric()

median = median(data2)

lambda = 0.4549/median

lambda
```

```
## [1] 0.7477657
```

```
lambda*data2
```

```
##  [1] 3.82267127 0.61844558 2.36154509 2.53892736 0.04254794 0.65687198
##  [7] 3.70530010 0.09510533 0.83405087 1.10021341 0.03183799 0.62301715
## [13] 0.29135442 0.06599828 0.00602476 0.15413107 0.03917213 0.01557086
## [19] 1.08113749 0.14605442
```

```
qchisq(.95, 1)
```

```
## [1] 3.841459
```

```
which(lambda*data2 > qchisq(.95, 1))
```

```
## integer(0)
```

There is no evidence for admixture since the inflation factor adjusted chi-squared values for the null markers are less than the chi-squared with 1 df.

The genomic adjustment factor is 0.748.

The marker of interest is not associated with affection status in the alleles test or the trend test adjusted for genetic control.

Chapter 9

Problem 1

$(x-y)^2/(x+y)$

```r
(78-46)^2/(78+46)
```

```
## [1] 8.258065
```

Confirmed.

Problem 2

The alternative hypothesis of a TDT is that a marker is both linked and associated with a disease locus underlying the trait.

A rejection of the null in a case-control or cohort study does not necessarily mean an association with a disease locus because of issues with population substructure.

Problem 3

The TDT is conditioned on the parental genotypes. The null distribution is computed using the distribution of the offspring genotypes conditional on parental genotypes and offspring traits.