Stat 344 – PS05

Luca Seazzu

July 01, 2023

Problem 5.26

5.30

```
n <- c(1997, 906, 904, 32)

LL <- function(theta, x){
  x[1] * log(.25 * (2 + theta)) +  x[2] * log(.25 * (1-theta)) + x[3] * log(.25 * (1-theta)) + x[4] * l
}

maxLik(LL, start = .5 , x = n)
```

```
## Warning in log(0.25 * (theta)): NaNs produced
```

```
## Maximum Likelihood estimation
## Newton-Raphson maximisation, 3 iterations
## Return code 8: successive function values within relative tolerance limit (reltol)
## Log-Likelihood: -4074.879 (1 free parameter(s))
## Estimate(s): 0.0357123
```

e.

```
o <- n; o
```

```
## [1] 1997  906  904   32
```

```
sn <- sum(n); sn
```

```
## [1] 3839
```

```
theta <- 0.0357123

probs <- c(.25*(2 + theta), (.25 * (1-theta)), (.25 * (1-theta)), (.25 * theta))

e <- probs * sn

G <- 2 * sum(o * log(o / e)); G
```

```
## [1] 2.018721
```

```
1 - pchisq(G, df = 2)
```

```
## [1] 0.3644519
```

Data did not fit the model well with a goodness-of-fit p-value of .36

f.

```
theta.hat <- 0.0357123

pval_minus_critical <- function(theta0) {
  2 * (LL(theta.hat, x) - LL(theta0, x)) - qchisq(.95, df = 1)
}

lo <- uniroot(pval_minus_critical, c(0, theta.hat)) |> value()
hi <- uniroot(pval_minus_critical, c(theta.hat, 1)) |> value()

c(lo,hi)
```

5.33

```
bin.pois <- function() {

rdata <- data.frame(V1 = rpois(1000, lambda = 15))

breaks <-c(0, 5, 10, 15, 20, 25, Inf)

rdata <- rdata |>
  mutate(binned = cut(V1, breaks = breaks))
tally(~binned, data = rdata)

count_dat <- data.frame(tally(~binned, data = rdata))
count_dat <- count_dat |>
mutate(probs = diff(ppois(breaks, lambda = 10)),
e = sum(count_dat$Freq) * probs )

G_pois <- 2 * sum( count_dat$Freq * log( count_dat$Freq / count_dat$e))

1 - pchisq(G_pois, df = nrow(count_dat) - 1 - 2)
}

bin.pois()
```

```
## [1] 0
```

5.36

```
Fusion1m <- merge(FUSION1, Pheno, by = "id", all = FALSE)

genotype <- tally(~ t2d + genotype, data = Fusion1m)

Gdose <- tally(~ t2d + Gdose, data = Fusion1m)
```

a. T is the most common allele among cases than among controls.

b.

```
chisq.test(genotype)
```

```
##
##  Pearson's Chi-squared test
##
## data:  genotype
```

```
## X-squared = 18.306, df = 2, p-value = 0.0001059
```

Based on the chi-squared test, there is an association between the SNP and type 2 diabetes.

c. knowing this marker was just one part of a larger study with over 300,000 SNPs changes how to interpret these results because