C240424_3

YOUSEF IBRAHIM GOMAA MAHMOUD MABROUK

2024-12-16

Exercise 3-1

$\because \hat{\pi_1} = \frac{62.4}{1000000} = 0.0000624$, $\hat{\pi_2} = \frac{1.3}{1000000} = 0.0000013$

(a) $\therefore$ Difference of Proportions $= \hat{\pi_1} - \hat{\pi_2} = 0.0000611$

(b) $\therefore$ Relative Risk $= \frac{\hat{\pi_1}}{\hat{\pi_2}} = 48$

- Relative risk is more practical as it provides a better explanation for the association between the two probabilities, while as the difference yields an absolute value that is less meaningful, and may be interpreted differently due to its extremely small magnitude.
    - In this case, it is clear that the proportion of US citizens that had a gun-related death is 48 times than that of British citizens. (Relative Risk = 48)
    - However, this also means that the risk of US citizens experiencing a gun-related death is 0.00611% higher than British citizens. (Difference of Proportions = 0.0000611)

Exercise 3-2

(a) No, 1.7 is an odd ratio. As it signifies the ratios of the odds of an event occurring in one group to the odds of the same event occurring in another group, which is the event of lung cancer in men and women in this case. It does not necessarily mean that it is a relative risk.

(b) (i) Relative risk for those taking the drug compared to those taking placebo = 1 - 0.45 = 0.55 = 55% of the risk for the placebo group.
    - This means the risk for women taking the drug is 55% of the risk for women taking the placebo.

(ii) Relative risk for those taking placebo compared to those taking the drug $= \frac{1}{0.55} = 1.8181818$

Exercise 3-3

a. The word probability is not suitable for this statement as the odds ratio describe the relation between odds of an event happening in a group to another group.
The correct interpretation is: "The odds of survival for females were 11.4 times higher than the odds of survival for males."

b. $\because$ Odds of Survival for Females $= \frac{\pi_f}{1-\pi_f} = 2.9$

$\therefore pi_f = \frac{2.9}{2.9+1} = 0.7435897$ (1)

$\because$ Odd Ratio $= \frac{Odds\ for\ Females}{Odds\ for\ Males} = 11.4$, Odds of Survival for Males $= \frac{\pi_m}{1-\pi_m}$

$\therefore$ Odds for Males $= \frac{Odds\ for\ Females}{Odd\ Ratio} = \frac{2.9}{11.4} = 0.254386$

$\therefore pi_m = \frac{\frac{2.9}{11.4}}{(\frac{2.9}{11.4}+1)} = 0.2027972$ (2)

From (1) and (2):

- RR $= \frac{P_f}{P_m} \approx 3.6666667$

- Then the interpretation is as follows: "The probability of survival for females was 3.6666667 times that for males."

Exercise 3-4

(a) $\because \mu_{ij} = n\hat{\pi}_{i+}\hat{\pi}_{+j} = \frac{n_{i+}n_{+j}}{n}$

$\therefore \mu_{11} = \frac{(21+159+110)(21+53+94)}{(21+159+110+53+372+221+94+249+83)} = 35.7709251 \approx 35.8$
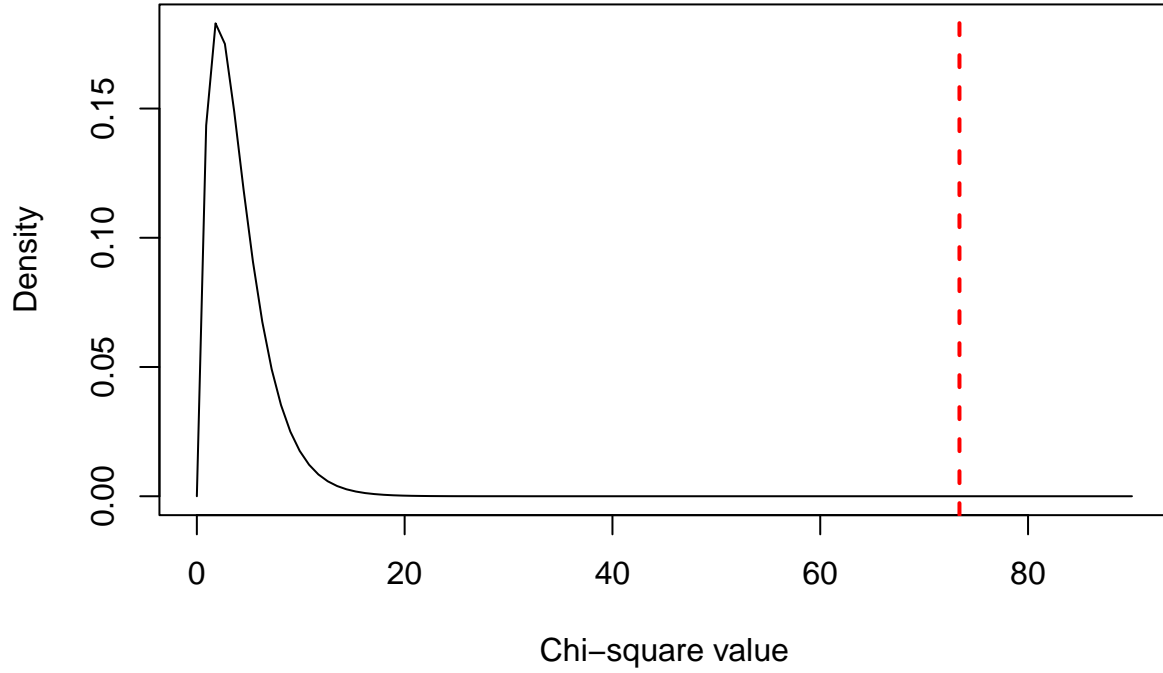
(b) $d_f = $ (r-1)(c-1) = (3-1)(3-1) = 4

Calculation in R for Question (b):

```r
x <- c(21, 159, 110, 53, 372, 221, 94, 249, 83)
curve(dchisq(x, df=4), from=0, to=90,
      xlab="Chi-square value", ylab="Density",
      main="Chi-square Distribution (df=4)")

chisqr <- chisq.test(matrix(x, ncol=3),
correct = FALSE)

chi_sq_result <- round(chisqr$statistic, 1)
abline(v = chi_sq_result, col = "red", lwd = 2, lty = 2)
```

Chi−square Distribution (df=4)


```
chisqr
```

```
##
##   Pearson's Chi-squared test
##
## data:  matrix(x, ncol = 3)
## X-squared = 73.352, df = 4, p-value = 0.000000000000004444
```

Since p-value is extremely close to 0 ($P(X^2 > 73.4) < 0.05$), there is strong evidence to reject the null hypothesis.

(c) $\because r_{ij} = \dfrac{n_{ij} - \hat{\mu}_{ij}}{\sqrt{\hat{\mu}_{ij}(1-\hat{\pi}_{i+})(1-\hat{\pi}_{+j})}}$

$\therefore r_{11} = \dfrac{21-35.8}{\sqrt{35.8(1-\frac{21+159+110}{21+159+110+53+372+221+94+249+83})(1-\frac{21+53+94}{21+159+110+53+372+221+94+249+83})}} = $ -2.9778153

$\therefore r_{33} = \dfrac{83-129.5}{\sqrt{129.5(1-\frac{94+249+83}{21+159+110+53+372+221+94+249+83})(1-\frac{110+221+83}{21+159+110+53+372+221+94+249+83})}} = $ -5.9081714

(d) $\because r_{ij} = \dfrac{n_{ij} - \hat{\mu}_{ij}}{\sqrt{\hat{\mu}_{ij}(1-\hat{\pi}_{i+})(1-\hat{\pi}_{+j})}}$

$\therefore r_{13} = \dfrac{110-88.1}{\sqrt{88.1(1-\frac{21+159+110}{21+159+110+53+372+221+94+249+83})(1-\frac{110+221+83}{21+159+110+53+372+221+94+249+83})}} = $ 3.1523302

$\therefore r_{31} = \dfrac{94-52.5}{\sqrt{52.5(1-\frac{94+249+83}{21+159+110+53+372+221+94+249+83})(1-\frac{21+53+94}{21+159+110+53+372+221+94+249+83})}} = $ 7.3791352

The results are a bit off due to rounding of expected cell counts.

Calculation in R for Question (c), (d):

```
chisq.test(matrix(x, ncol=3),
correct = FALSE)$stdres
```

```
##              [,1]       [,2]        [,3]
## [1,] -2.9731729 -4.403194  7.3676655
## [2,] -0.9472192  0.224221  0.5948871
## [3,]  3.1442766  2.906749 -5.9070232
```

Exercise 3-5

Because the $\hat{\pi}_{ij}$ values are fixed, that means as n increases, the observed and expected counts scale proportionally, but the chi-squared statistic (X^2) grows because it multiplies the squared deviations by n. For example, assuming a 2x2 table with observed proportions and expected proportions under the null hypothesis.
For small sample size (n = 100):

```
n_small <- 100
pi_ij <- c(0.25, 0.15, 0.35, 0.25)
pi_i_plus <- c(0.4, 0.6)
pi_plus_j <- c(0.6, 0.4)

expected_11 <- pi_i_plus[1] * pi_plus_j[1]
expected_12 <- pi_i_plus[1] * pi_plus_j[2]
expected_21 <- pi_i_plus[2] * pi_plus_j[1]
expected_22 <- pi_i_plus[2] * pi_plus_j[2]

X2_small <- n_small * ((pi_ij[1] - expected_11)^2 / expected_11 +
                       (pi_ij[2] - expected_12)^2 / expected_12 +
                       (pi_ij[3] - expected_21)^2 / expected_21 +
                       (pi_ij[4] - expected_22)^2 / expected_22)
X2_small
```

```
## [1] 0.1736111
```

But for a large sample size (n = 1000000):

```
n_large <- 1000000
X2_large <- n_large * ((pi_ij[1] - expected_11)^2 / expected_11 +
                       (pi_ij[2] - expected_12)^2 / expected_12 +
                       (pi_ij[3] - expected_21)^2 / expected_21 +
                       (pi_ij[4] - expected_22)^2 / expected_22)
X2_large
```

```
## [1] 1736.111
```

and this does not mean that the higher value signifies the strength of group association, but instead the degree of evidence of its existence.