Exercises for Lecture 3 and 4

Rei Monden

2024-12-12

Exercise 3-1

(a) The difference of proportions is $\frac{62.4}{1{,}000{,}000} - \frac{1.3}{1{,}000{,}000} = 0.0000611$.

(b) The relative risk is $\frac{62.4}{1.3} = 48$.

From this, we can say that the estimated probability of gun-related deaths in the US is 48 times that in Britain. From the above-shown results, we can say that the relative risk is more useful as a measure to describe how different both proportions are. The difference of proportions is misleadingly small, wrongly implying that the difference between the US and Britain is extremely small.

Exercise 3-2

(a) Yes: $\frac{P(\text{Women get small-cell lung cancer})}{P(\text{Men get small-cell lung cancer})} = 1.7$.

(b) Denote
$\pi_1 = P(\text{Women experiencing invasive breast cancer who are taking the drug})$
$\pi_2 = P(\text{Women experiencing invasive breast cancer who are not taking the drug})$.

  (i) $\pi_1 = 0.55 \times \pi_2$. Therefore, $\frac{\pi_1}{\pi_2} = 0.55$.
  (ii) $\frac{1}{0.55} = 1.82$.

Exercise 3-3

(a) The quoted interpretation is about the relative risk. Therefore, the term **probability** should be replaced by **odds**.

(b) For females, proportion $= \frac{2.9}{(1+2.9)} = 0.744$.
Odds for males $= \frac{2.9}{11.4} = 0.254$, and therefore the proportion is $= 0.254/(1 + 0.254) = 0.203$.
Therefore, $RR = \frac{0.744}{0.203} = 3.7$.

Exercise 3-4

(a) The sample size is 1362, the 1st row total is 290, and the 1st column total is 168.
Therefore, $35.8 = (290)(168)/1362$.

(b) $df = (3-1)(3-1) = 4$, P-value $= P(\chi^2_4 > 73.4) < .001$. At 5% significance level, we conclude that the observed data are extremely unlikely in case *income* and *happiness* were independent. We therefore reject the null hypothesis of independence.

(c) The expected counts were computed assuming that *Happiness* and *Income* were independent (i.e., H_0). Under H_0, these two standardized residuals indicate that the independence model predicts much higher counts than what was observed. For instance, in this sample, the observed count in the first cell is 2.93 standard errors smaller than the estimated count.
Since H_0 was rejected in question (b), we expect (but have no evidence yet!) that *fewer* people will be predicted in these cells assuming that *Happiness* and *Income* are not independent (i.e., H_1).

(d) Similar to question (c), but now the independence model is most likely *underpredicting* the true frequencies.

Exercise 3-5

For any "reasonable" significance test, whenever H_0 is false, the test statistic tends to be larger and the P-value tends to be smaller as the sample size increases. Even if H_0 is just slightly false, the P-value will be small if the sample size is large enough.

In this case, as long as the variables are not independent in the population, then most likely $(\hat{\pi}_{ij} - \hat{\pi}_{i+}\hat{\pi}_{+j})$ will be different from 0. The alternative formula for X^2 is then, essentially, n times a non-zero quantity. Thus, increasing the sample size (for fixed sample observed proportions) will increase X^2 with no bound.

Because significance tests can be sensitive to even tiny deviations from H_0 (which may be deemed of little *practical* importance), many statisticians think we learn more by estimating parameters using confidence intervals instead of conducting significance testing.