STAT 3004: Assignment 4

Cheng Li
2 Dec 2020

---

1 Problems 1

1.1 (13.19)

The estimated $OR = \frac{52/300}{66/753} = 2.0$

1.2 (13.20)

We use the Woolf formula to obtain a 95% CI for the OR. A 95% CI for ln(OR) is given by: $ln(\hat{OR}) \pm 1.96 * \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}$ where a, b, c, d are 52, 300, 66, 753. Therefore the CI of ln(OR) is (0.295, 1.069). As a consequence, the 95% CI is (1.3,2.9).

2 Problem 2

2.1 (13.80)

We use the chi-square test for $2 * 2$ tables.

$$\chi^2 = \frac{3000 * [|34 * 1432 - 68 * 1466| - 1500]^2}{102 * 2898 * 1500 * 1500}$$
$$= 11.05 \sim \chi_1^2 \; under \; H_0,$$

since $\chi^2_{1,0.999} = 10.83 < 11.05$, it follows that $p < 0.001$. Thus, raloxifene significantly reduces the risk of new fractures among women with no preexisting fractures.

2.2 (13.81)

We can get that:

$$RR = \frac{34/1500}{68/1500} = 0.5,$$

The se of lnRR is $\sqrt{\frac{b}{an_1} + \frac{d}{cn_2}} = 0.2068$. A 95% CI for lnRR is (-1.099, -0.288). So a 95% CI for RR is (0.33, 0.75).

2.3 (13.82)

We use the Mantel Haenszel test. We have the test statistic:

$$X_{MH}^2 = \frac{(|O-E|-0.5)^2}{V}$$
$$O = 34 + 103 = 137$$
$$E = \frac{102 * 1500}{3000} + \frac{273 * 700}{1500} = 178.4$$
$$V = \frac{102 * 2898 * 1500 * 1500}{3000^2 * 2999} + \frac{273 * 1227 * 700 * 800}{1500^2 * 1499} = 80.26$$

So, the test statistic is:

$$X_{MH}^2 = 20.84 \sim \chi_1^2 \; under \; H_0$$

And the p-value is less than 0.001. Thus, raloxifene significantly reduces the risk of new fractures in both groups combined.

2.4 (13.83)

In the total population, there are 3000 women without preexisting fractures and 1500 women with preexisting fractures. Thus, the standardized risk ratio is

$$SRR = \frac{34/1500 * 3000 + 103/700 * 1500}{68/1500 * 3000 + 170/800 * 1500}$$
$$= 0.63$$

2.5 (13.84)

No. Randomization was preformed after stratification by preexisting fractures. Thus it is unlikely to be a confounder.

3 Problem 3

3.1 (14.53)

A censored observation in 1992 would indicate that a woman was determined not to have breast cancer at the 1992 follow-up questionnaire, but no further information was obtain from her after that time.

3.2 (14.54)

For each group, the 10-year survival probability is estimated by $\hat{S}(10) = \Pi(1 - \frac{d_j}{S_{j-1}})$.
For the current users, this estimate is $(1-\frac{0}{200})*(1-\frac{3}{199})*(1-\frac{2}{194})*(1-\frac{4}{190})*(1-\frac{2}{185})*(1-\frac{2}{133}) = 0.93$
For the never users, this estimate is $(1-\frac{0}{1000})*(1-\frac{3}{988})*(1-\frac{9}{975})*(1-\frac{7}{944})*(1-\frac{5}{914})*(1-\frac{9}{716}) = 0.963$.
Since we are interested in the incidence rate of breast cancer, we subtract each of these estimates from 1 to obtain estimated 10-year incidence rates of 0.070 for the current users and 0.039 for the never users.

3.3 (14.55)

The log-rank test.

3.4 (14.56)

For our log-rank test, we have the test statistic:

$$X_{LR}^2 = \frac{(|O-E|-0.5)^2}{Var_{LR}} \sim \chi_1^2 \; under \; H_0.$$

Where

$$\begin{aligned} O = & \; observed \; number \; of \; failures \; in \; the \; current \; users \; group \\ E = & \; expected \; number \; of \; failures \; in \; the \; current \; users \; group \\ = & \sum_{i=1}^{6} E_i = \sum_{i=1}^{6} \frac{(a_i+b_i)(c_i+d_i)}{N_i} = \frac{(d_{i1}+d_{i2}) * S_{i-1,1}}{S_{i-1,1}+S_{i-1,2}} \end{aligned}$$

d_{i1}, d_{i2} = number of persons who fail in the i-th year in the current- and never-users group, respectively, S_{i1}, S_{i2} = number of persons who survive up to the i-th year in the current- and never-users groups, respectively.

$$\begin{aligned} Var_{LR} &= \sum_{i=1}^{6} V_i \\ &= \sum_{i=1}^{6} \frac{(a_i+b_i)(c_i+d_i)(a_i+c_i)(b_i+d_i)}{N_i^2(N_i-1)} \\ &= \sum_{i=1}^{6} \frac{(d_{i1}+d_{i2})(S_{i-1,1}+S_{i-1,2}-d_{i1}-d_{i2})(S_{i-1,1})(S_{i-1,2})}{(S_{i-1,1}+S_{i-1,2})^2(S_{i-1,1}+S_{i-1,2}-1)} \end{aligned}$$

We can get the p-value is $0.049 < 0.05$. So we can reject the null hypothesis.