Stat 205 Quiz 9

Take-Home Quiz due by beginning of class, Thursday April 21

A case-control study was designed to examine risk factors for cervical dysplasia (Becker et al., 1994); the women in the study were patients aged 18 to 40. There were $n_1 = 175$ cases with cervical dysplasia and $n_2 = 308$ controls without. Each woman was tested for human papilloma virus (HPV) and either positive (HPV+) or negative (HPV-):

	dysplasia	no dysplasia
HPV+	164	130
HPV-	11	178
total	175	308

Let p_1 be the probability of HPV+ among the cases and p_2 be the probability of HPV+ among the controls.

(a)

$$\hat{p}_1 = \frac{164}{175} = 0.937, \quad \hat{p}_2 = \frac{130}{308} = 0.422,$$

 $\hat{p}_1 - \hat{p}_2 = 0.515, \quad \frac{\hat{p}_1}{\hat{p}_2} = 2.22, \quad \hat{\theta} = \frac{\hat{p}_1/(1-\hat{p}_1)}{\hat{p}_2/(1-\hat{p}_2)} = 20.4.$

The probability of being HPV+ increases by *one half* when dysplasia is present. One is twice as likely to have HPV when dysplasia is present. The odds of being HPV+ increase by a factor of 20 when dysplasia is present.

(b)

$$\tilde{p}_1 = \frac{164+1}{175+2} = 0.932, \quad \tilde{p}_2 = \frac{130+1}{308+2} = 0.423,$$
$$SE_{\tilde{p}_1-\tilde{p}_2} = \sqrt{\frac{\tilde{p}_1(1-\tilde{p}_1)}{n_1+2} + \frac{\tilde{p}_2(1-\tilde{p}_2)}{n_2+2}} = 0.0338.$$
$$\tilde{p}_1 - \tilde{p}_2 \pm 1.96SE_{\tilde{p}_1-\tilde{p}_2} = (0.443, 0.576).$$

We reject that $H_0: p_1 = p_2$ at the 5% level because the 95% confidence interval does not include zero.

(c)

$$\log \hat{\theta} = \log \frac{164 \times 178}{130 \times 11} = \log 20.4 = 3.02$$

$$SE_{\log\hat{\theta}} = \sqrt{\frac{1}{y_1} + \frac{1}{n_1 - y_1} + \frac{1}{y_2} + \frac{1}{n_2 - y_2}}$$
$$= \sqrt{\frac{1}{164} + \frac{1}{11} + \frac{1}{130} + \frac{1}{178}}$$
$$= 0.332.$$

The 95% confidence interval for $\log \theta$ is given by

$$3.02 \pm 1.96 \times 0.332 = (2.37, 3.67).$$

The 95% confidence interval for θ is given by

$$(e^{2.37}, e^{3.67}) = (10.6, 39.1).$$

The odds of dysplasia increase by 10 to 40 times for HPV+ versus HPV-. *Also*, the odds of HPV+ increase 10 to 40 times for those with dysplasia versus those without.

- (d) We reject $H_0: \theta = 1$ at the 5% level because the 95% confidence interval does not include one. This agrees with part (b).
- (e) There is a significant association between dysplasia and HPV; we reject no association at the 5% level in parts (b) and (d).