Jacob Miller STAT-784-Midterm

(2)

	no cancer	cancer	
low	350	680	1030
high	120	980	1100
	470	1660	2130

a) a "case-control study" is one that looks back in time to compare cases to controls

b) this study was likely run by choosing a set number of samples (2130) & seeing if they had high/low levels of exposure & whether they got cancer or not

c) $OR = \frac{\hat{\pi}_1/(1-\hat{\pi}_1)}{\hat{\pi}_2/(1-\hat{\pi}_2)} = \frac{350 \cdot 980}{120 \cdot 680} = 4.2$

d) Yes, the odds that an individual had cancer after a high level of exposure are more than 4 times the odds of an individual getting cancer after a low level of exposure

e) $\hat{\pi}_1 = 680/1030 = 0.66$ $\hat{\pi}_2 = 980/1100 = 0.89$

relative risk $= \pi_2/\pi_1 = 0.89/0.66 = 1.35$

∴ sample proportion for patients w/ cancer is 35% higher for high levels of exposure

$SE = \sqrt{1/n_{11} + 1/n_{12} + 1/n_{21} + 1/n_{22}} = \sqrt{1/350 + 1/680 + 1/120 + 1/980} = 0.117$

confidence interval $= \ln\hat{\theta} \pm z_{\alpha/2} \cdot SE$

$= \ln(4.2) \pm 1.96(0.117)$

$= 1.21, \ 1.66$

$e^{1.21} = 3.35$ $e^{1.66} = 5.26$

∴ We can be 95% confident that the odds of getting cancer after being exposed at a high level are between 3.35 & 5.26

Jacob Miller STAT-784-Midterm

③

Mode	count	prob
air	3	0.4
bus	3	0.2
car	5	0.3
train	4	0.1

a) $P(y_1, y_2, y_3, y_4) = \left(\frac{n!}{y_1! y_2! y_3! y_4!}\right)\left(\pi_1^{y_1} \pi_2^{y_2} \pi_3^{y_3} \pi_4^{y_4}\right)$

$= \left(\frac{15!}{3!3!5!4!}\right)\left(0.4^3 \cdot 0.2^3 \cdot 0.3^5 \cdot 0.1^4\right)$

$= \underline{0.00157}$ (very small)

b) $\mu_{bus} = 15 \cdot 0.2 = \underline{3 \text{ people}}$

c)

air	bus	car	train
28	50	35	47

$car = \frac{35}{160} = 0.21875$

∴ No, I do not believe the probability a representative would like to drive themselves is 0.3, because this sample size is large enough to get an accurate sample distribution