Name: CHAN King Yeung
SID: 1155119394

STAT 4006
Homework 2

Question 1

	Yes	No
Placebo	115.46	10452.54
Aspirin	113.54	10278.46

H_0: the incidence of heart attack is independent of aspirin intake

H_1: the incidence of heart attack is dependent of aspirin intake

$$X^2 = \frac{(158-115.46)^2}{115.46} + \frac{(10410-10452.54)^2}{10452.54} + \frac{(71-113.54)^2}{113.54} + \frac{(10321-10278.46)^2}{10278.46}$$

$$\approx 31.9589$$

$$\text{p-value} \approx 1.5747 \times 10^{-8}$$

Since p-value < 0.05, we reject H_0 at $\alpha = 0.05$. We conclude that the incidence of heart attack is NOT independent of aspirin intake.

Question 2

$E_{11} = \frac{6(8)}{16} = 3 < 5$, there are over 20% of $E_{ij} < 5$.

Thus, we use Fisher's exact test here.

	Small	Large
Fatal	x	$8-x$
not Fatal	$6-x$	$2+x$

$$\left|\frac{x}{8} - \frac{6-x}{8}\right| \geq \left|\frac{1}{8} - \frac{5}{8}\right|$$

$$|2x-6| \geq 4$$

$$\Rightarrow x \geq 5 \text{ or } x \leq 1$$

H_0: the frequency of fatal accidents is independent of the size of automobiles

H_1: the frequency of fatal accidents is dependent of the size of automobiles

$$\text{p-value} = 1 - \frac{\binom{6}{2}\binom{10}{6} + \binom{6}{3}\binom{10}{5} + \binom{6}{4}\binom{10}{4}}{\binom{16}{8}}$$

$$\approx 0.1189$$

Since p-value > 0.05, we do not reject H_0 at $\alpha = 0.05$. We cannot conclude that the frequency of fatal accidents is NOT independent of the size of automobiles

Question 3

	Fatal	Non-fatal	
None	1598	162526	164124
Seat belt	502	421360	421862

a) $$\hat{\gamma} = \frac{1598}{164124} - \frac{502}{421862}$$

$$\approx 0.0085$$

$$\hat{RR} = \frac{\frac{1598}{164124}}{\frac{502}{421862}}$$

$$\approx 8.1822$$

con't a) $$\hat{\theta} = \frac{1598(421360)}{502(162526)}$$

$$\approx 8.2528$$

$\hat{\gamma}$ suggests a small different between 2 groups, but both $\hat{RR}$ and $\hat{\theta}$ suggest a strong association between fatal and use seat belt; Since the probabilities of fatal in both "None" and "Seat belt" are close to zero, the relative risk approximately equals odd ratio.

b) $$CI(\hat{\gamma}): 0.0085 \pm 1.96\sqrt{\frac{1598(162526)}{(164124)^3} + \frac{502(421360)}{(421862)^3}}$$

$$\approx (0.008, 0.009)$$

Since CI does not contain zero, the association of not using seat belt with fatal is stronger than that of using seat belt.

$$CI(\hat{RR}): e^{\ln(8.1822) \pm 1.96\sqrt{\frac{1}{1598} - \frac{1}{164124} + \frac{1}{502} - \frac{1}{421862}}}$$

$$\approx (7.4027, 9.0438)$$

The risk of fatal is at least 740% higher for not using seat belt group.

$$CI(\hat{\theta}): e^{\ln(8.2528) \pm 1.96\sqrt{\frac{1}{1598} + \frac{1}{162526} + \frac{1}{502} + \frac{1}{421360}}}$$

$$\approx (7.4641, 9.1248)$$

The odds of fatal is at least 746% higher for not using seat belt group.

Question 4

	1	2	3	
1	178	183	108	469
2	570	648	442	1660
3	138	252	252	642
	886	1083	802	2771

a) $$C = 178(648+442+252+252) + 183(442+252) + 570(252+252) + 648(252)$$

$$= 861310$$

$$D = 108(570+648+138+252) + 442(138+252) + 183(570+138) + 648(138)$$

$$= 565032$$

con't a) $\hat{r} = \frac{861310 - 565032}{861310 + 565032}$

≈ 0.2077

there is a weak tendency for educational level increase as religious beliefs increase

b) $\bar{u} \approx 2.0624$ $\quad (n-1)Var(u) \approx 0.397$

$\bar{v} \approx 1.9697$ $\quad (n-1)Var(v) \approx 0.6082$

$(n-1)Cor = \frac{(1-\bar{u})}{2271}[178(1-\bar{v}) + 183(2-\bar{v}) + 108(3-\bar{v})]$

$+ \frac{(2-\bar{u})}{2271}[570(1-\bar{v}) + 648(2-\bar{v}) + 442(3-\bar{v})]$

$+ \frac{(3-\bar{u})}{2271}[138(1-\bar{v}) + 252(2-\bar{v}) + 252(3-\bar{v})]$

≈ 0.0833

$r = \frac{0.0833}{\sqrt{0.397(0.6082)}}$

≈ 0.1695

H_0: the religious beliefs is independent of the degree

H_1: the religious beliefs is dependent of the degree

$M^2 = 2270(0.1695)^2$

≈ 65.2177

critical value $= 3.841$

Since $M^2 > 3.841$, we reject H_0 at $\alpha = 0.05$.

c) the method in b) provides a more accurate test statistic by considering the ordering information of the data.

con't a) $\hat{\theta}_{AG(2)} = \frac{365(7)}{199(16)}$

≈ 0.8024

$\hat{\theta}_{AG(3)} = \frac{117(385)}{203(204)}$

≈ 1.0877

$\hat{\theta}_{AG(4)} = \frac{133(250)}{276(127)}$

≈ 0.9486

$\hat{\theta}_{AG(5)} = \frac{50(299)}{138(94)}$

≈ 1.1525

$\hat{\theta}_{AG(6)} = \frac{22(317)}{351(24)}$

≈ 0.8279

$\hat{\theta}_{AG} = \frac{1165(1281)}{1469(545)}$

≈ 1.864

for partial θ, the majority are less than 1, which means females are more likely to be admitted; for marginal θ, it is greater than 1, which means males are more likely to be admitted. "Gender" may be the effect modifier that controls over the association between "department" and "admission".

Question 5

X: number of COVID 19 newly confirmed patients in Hong Kong

Y: number of COVID 19 newly confirmed patients in China

Z: government policy (dummy variable)

Question 6

a) $\hat{\theta}_{AG(1)} = \frac{478(23)}{302(80)}$

≈ 0.455