1. (a) First of all, the sample size (= 40) is too small to use Large-sample Wald Confidence Interval so I use R and SAS to calculate those intervals. $\alpha = 0.05$

$D = \hat{\pi}_1 - \hat{\pi}_2 = \frac{n_{11}}{n_{1+}} - \frac{n_{21}}{n_{2+}} = \frac{21}{23} - \frac{8}{17} \fallingdotseq 0.442$.

95% Wald Confidence interval for D is

$(0.1787184, 0.7061921)$.

$\therefore$ D is bigger than 1 and the confidence interval does not contain '1' so the probability of the cancer controlled by surgery is bigger than the probability of the cancer controlled by Radiation therapy

$RR = \frac{\hat{\pi}_1}{\hat{\pi}_2} = \frac{n_{11}/n_{1+}}{n_{21}/n_{2+}} \fallingdotseq 1.94$.

95% Wald Confidence interval for RR is

$(1.1538, 3.2626)$.

$\therefore$ RR and the all values in the confidence interval are bigger than 1 so the probability of the cancer controlled by surgery is bigger than the probability of the cancer controlled by Radiation therapy

$\hat{\theta}(OR) = \frac{\hat{\pi}_1/(1-\hat{\pi}_1)}{\hat{\pi}_2/(1-\hat{\pi}_2)} = \frac{\frac{n_{11}}{n_{1+}} / \frac{n_{12}}{n_{1+}}}{\frac{n_{21}}{n_{2+}} / \frac{n_{22}}{n_{2+}}} = \frac{n_{11}n_{22}}{n_{12}n_{21}} = \frac{21 \times 9}{8 \times 2} = 11.813$.

95% Wald Confidence interval for θ is

$(2.0835, 66.9727)$.

$\therefore$ Odds ratios and the all values in the confidence interval are bigger than 1 so the probability of the cancer controlled by surgery is much bigger than the probability of the cancer controlled by radiation therapy.

(b) Estimated expected frequency :

$\hat{\mu}_{11} = n\hat{\pi}_{1+}\hat{\pi}_{+1} = n \times \frac{n_{1+}}{n}\frac{n_{+1}}{n} = \frac{n_{1+}n_{+1}}{n} = \frac{23 \times 29}{40} = 16.675$

$\hat{\mu}_{12} = \frac{n_{1+}n_{+2}}{n} = \frac{23 \times 11}{40} = 6.325$, $\hat{\mu}_{21} = \frac{n_{2+}n_{+1}}{n} = \frac{17 \times 29}{40} = 12.325$, $\hat{\mu}_{22} = \frac{n_{2+}n_{+2}}{n} = \frac{17 \times 11}{40} = 4.675$.

Pearson χ^2 statistic for testing $H_0: \theta = 1$ vs $H_1: \theta \neq 1$:

$X^2 = \Sigma(n_{ij} - \hat{\mu}_{ij})^2/\hat{\mu}_{ij} = (21-16.675)^2/16.675 + (2-6.325)^2/6.325 + (8-12.325)^2/12.325$
$+ (9-4.675)^2/4.675 \doteq 1.122 + 2.957 + 1.518 + 4.001$
$= 9.598 \sim \chi^2_1$.

LR χ^2 statistic for testing $H_0: \theta = 1$ vs $H_1: \theta \neq 1$:

$G^2 = 2\Sigma n_{ij}\log(n_{ij}/\hat{\mu}_{ij}) = 2\{21 \cdot \log(\frac{21}{16.675}) + 2 \cdot \log(\frac{2}{6.325}) + 8 \cdot \log(\frac{8}{12.325}) + 9 \cdot \log(\frac{9}{4.675})\}$
$\doteq 2 \times 4.978 = 9.956 \sim \chi^2_1$.

If $\mu_{ij} < 5$, it is poor to use G^2 in testing. We have $\hat{\mu}_{22} = 4.675$
so it can be better for us to do Fisher's exact test.

$\chi^2_{1,0.05} = 3.84$. so the both X^2 and G^2 show that we can reject the null hypothesis.

(c) $n_{11} = 21$

$P(21) = \frac{\binom{23}{21}\binom{17}{8}}{\binom{40}{29}} \doteq 0.0027$ where $\max(0, 12) \leq 21 \leq \min(23, 29)$

p-value = 0.0034.

p-value = 0.0034 < 0.05 so we can reject the null hypothesis.

2. (a) $\hat{\mu}_{11} = \frac{29 \times 76}{273} \doteqdot 8.07$ $\hat{\mu}_{12} = \frac{29 \times 108}{273} \doteqdot 11.47$ $\hat{\mu}_{13} = \frac{29 \times 89}{273} \doteqdot 9.45$

$\hat{\mu}_{21} = \frac{137 \times 76}{273} \doteqdot 38.14$ $\hat{\mu}_{22} = \frac{137 \times 108}{273} \doteqdot 54.20$ $\hat{\mu}_{23} = \frac{137 \times 89}{273} \doteqdot 44.66$

$\hat{\mu}_{31} = \frac{48 \times 76}{273} \doteqdot 13.36$ $\hat{\mu}_{32} = \frac{48 \times 108}{273} \doteqdot 18.99$ $\hat{\mu}_{33} = \frac{48 \times 89}{273} \doteqdot 15.65$

$\hat{\mu}_{41} = \frac{59 \times 76}{273} \doteqdot 16.43$ $\hat{\mu}_{42} = \frac{59 \times 108}{273} \doteqdot 23.34$ $\hat{\mu}_{43} = \frac{59 \times 89}{273} \doteqdot 19.23$

$H_0: \pi_{ij} = \pi_{i+}\pi_{+j}$ for all i, j vs $H_1: \pi_{ij} \neq \pi_{i+}\pi_{+j}$

i) Use Pearson χ^2 statistic,

$X^2 = \Sigma(n_{ij} - \hat{\mu}_{ij})^2/\hat{\mu}_{ij} = \frac{(9-8.07)^2}{8.07} + \frac{(11-11.47)^2}{11.47} + \cdots + \frac{(27-19.23)^2}{19.23}$ (In SAS, 8.8709)

$\doteqdot 8.878. \sim \chi^2_6$. ($\because$ mean of χ^2_6 is 6, sd is about 3.464)

$\therefore$ I can't reject the null hypothesis that educational aspirations and family income are independent.

ii) Use the likelihood ratio χ^2 statistic,

$G^2 = 2\Sigma n_{ij} \log(n_{ij}/\hat{\mu}_{ij}) = 2\{9 \times \log(\frac{9}{8.07}) + 11 \times \log(\frac{11}{11.47}) + \cdots + 27 \times \log(\frac{27}{19.23})\}$

$\doteqdot 2 \times 4.471 = 8.942. \sim \chi^2_6$. (In SAS, 8.9165)

$\therefore$ I can't reject the null hypothesis that educational aspirations and family income are independent.

(b) $r_{11} = (n_{11} - \hat{\mu}_{11})/\sqrt{\hat{\mu}_{11}(1-\hat{\pi}_{1+})(1-\hat{\pi}_{+1})} = (9-8.07)/\sqrt{8.07 \times \frac{244}{273} \times \frac{197}{273}}$

$\doteqdot 0.41$.

$r_{43} = (n_{43} - \hat{\mu}_{43})/\sqrt{\hat{\mu}_{43}(1-\hat{\pi}_{4+})(1-\hat{\pi}_{+3})} = (27-19.23)/\sqrt{19.23 \times \frac{214}{273} \times \frac{184}{273}}$

$\doteqdot 2.44$.

$\therefore$ Residuals do not suggest any association pattern

$|r_{41}|, |r_{43}| > 2$. so the n_{41}, n_{43} can reject the H_0 when $\alpha = 0.05$

(c) $\hat{\theta}(OR) = \frac{9 \times 27}{9 \times 10} = 2.7$

$\therefore$ The family income estimated odds of some high school are 2.7 times ~~lower~~ higher than the family income estimated odds of college graduate. ✓

2-(b) Complete the table

	Family income			
Study Aspirations	low	middle	high	Total
Some high school	9 (8.07, 0.41)	11 (11.47, -0.19)	9 (9.45, -0.19)	29
High School Graduate	44 (38.14, 1.58)	52 (54.02, -0.54)	41 (44.66, -0.95)	137
Some college	13 (13.36, -0.13)	23 (18.99, 1.30)	12 (15.65, -1.24)	48
college graduate	10 (16.43, -2.11)	22 (23.34, -0.40)	27 (19.23, 2.44)	59
Total	76	108	89	273

3. (a) $\bar{u} = \sum_i u_i \hat{\pi}_{i+} = 1 \times \frac{29}{273} + 2 \times \frac{137}{273} + 3 \times \frac{48}{273} + 4 \times \frac{59}{273} \doteq 2.50$

$\bar{v} = \sum v_j \hat{\pi}_{+j} = 1 \times \frac{76}{273} + 2 \times \frac{108}{273} + 3 \times \frac{89}{273} \doteq 2.048$

$\hat{\rho} = \sum_{i,j} (u_i - \bar{u})(v_j - \bar{v}) \hat{\pi}_{ij} / \sqrt{[\sum_i (u_i - \bar{u})^2 \hat{\pi}_{i+}][\sum_j (v_j - \bar{v})^2 \hat{\pi}_{+j}]}$

$\doteq 0.097 / \sqrt{0.895 \times 0.602} \doteq 0.132$

$H_0: \rho = 0$ vs $H_1: \rho \neq 0$

$M^2 = (273-1)\hat{\rho}^2 = 272 \times (0.132)^2 \doteq 272 \times 0.017 = 4.739 \sim \chi^2_1$ (In SAS, 4.7489)

p-value < 0.05 so we can reject the null hypothesis under $\alpha = 0.05$

$\hat{\rho} = 0.132$ and $M \doteq 2.177$ so there is an increasing trend.

(b) $\bar{u} = \sum_i u_i \hat{\pi}_{i+} = 10 \times \frac{29}{273} + 20 \times \frac{137}{273} + 30 \times \frac{48}{273} + 40 \times \frac{59}{273} \doteq 25.02$

$\bar{v} = \sum_j v_j \hat{\pi}_{+j} = (-1) \times \frac{76}{273} + 0 \times \frac{108}{273} + 1 \times \frac{89}{273} \doteq 0.048$

$\hat{\rho} = \sum_{ij} (u_i - \bar{u})(v_j - \bar{v}) \hat{\pi}_{ij} / \sqrt{[\sum_i (u_i - \bar{u})^2 \hat{\pi}_{i+}][\sum_j (v_j - \bar{v})^2 \hat{\pi}_{+j}]}$

$\doteq 0.97 / \sqrt{89.47 \times 0.602} \doteq 0.132$

The results do not change because $\hat{\rho}$ is same as $\hat{\rho}$ in (a).

(c) $\bar{u} = \sum_i u_i \hat{\pi}_{i+} = 1 \times \frac{29}{273} + 3 \times \frac{137}{273} + 5 \times \frac{48}{273} + 10 \times \frac{59}{273} \doteq 4.65$

$\bar{v} = \sum_j v_j \hat{\pi}_{+j} = 1 \times \frac{76}{273} + 5 \times \frac{108}{273} + 10 \times \frac{89}{273} \doteq 5.516$

$\hat{\rho} = \sum_{ij} (u_i - \bar{u})(v_j - \bar{v}) \hat{\pi}_{ij} / \sqrt{[\sum_i (u_i - \bar{u})^2 \hat{\pi}_{i+}][\sum_j (v_j - \bar{v})^2 \hat{\pi}_{+j}]}$

$\doteq 1.575 / \sqrt{8.99 \times 12.34} \doteq 0.149$

$H_0: \rho = 0$ vs $H_1: \rho \neq 0$

$M^2 = (273-1)\hat{\rho}^2 = 272 \times (0.149)^2 \doteq 6.038 \sim \chi^2_1$ (In SAS, 6.0867)

p-value < 0.05 so the results do not change.

4. (a) $\hat{\theta}_{AG(1)} = \frac{512\times19}{319\times89} \fallingdotseq 0.349$ $\hat{\theta}_{AG(2)} = \frac{353\times8}{207\times17} \fallingdotseq 0.803$ $\hat{\theta}_{AG(3)} = \frac{120\times391}{205\times202} \fallingdotseq 1.133$

$\hat{\theta}_{AG(4)} = \frac{138\times244}{279\times131} \fallingdotseq 0.921$ $\hat{\theta}_{AG(5)} = \frac{53\times299}{138\times94} \fallingdotseq 1.222$ $\hat{\theta}_{AG(6)} = \frac{22\times317}{351\times24} \fallingdotseq 0.828$

∴ The estimated odds ratios of the men who were admitted are 34.3%, 80.3%, 113.3%, 92.1%, 122.2%, and 82.8% of the estimated odds ratios of the women in the department 1.2.3.4.5. and 6 respectively.

(b) The sample AG marginal odds ratios. $\frac{1198\times1278}{1493\times557} \fallingdotseq 1.841$

∴ The estimated odds ratios of the men who were admitted are 184.1% of the estimated odds ratios of the women.

(c) They give such different indications of the AG association because of Simpson's paradox. For example, the conditional marginal odds ratios in the department 1 & 2 are low but about 72% of the men who admitted admitted department 1 & 2 while about 19% of the women who admitted admitted department 1 & 2.