1.

(a)

Difference of proportions

```
> prop.test(c(21,8), c(23, 17), conf.level=0.95, correct=FALSE)
        2-sample test for equality of proportions without continuity
correction

data:  c(21, 8) out of c(23, 17)
X-squared = 9.5981, df = 1, p-value = 0.001948
alternative hypothesis: two.sided
95 percent confidence interval:
 0.1787184 0.7061921
sample estimates:
   prop 1    prop 2
0.9130435 0.4705882
```

The difference of proportions D = 0.9130 - 0.4705 = 0.4424; and Large sample 95% Wald CI for D is [0.1787, 0.7061]

The difference of success probability (that is, the probability to control the cancer) is 0.4424. Since the CI has only positive values, one can conclude that surgery has better control on cancer than radiation therapy.

Relative Risk

```
> riskratio(c(104, 10933, 189, 10845), method="wald", conf=0.95,
correct=FALSE)

$data
         Outcome
Predictor  Disease1 Disease2 Total
  Exposed1        9        8    17
  Exposed2        2       21    23
  Total          11       29    40

$measure
         risk ratio with 95% C.I.
Predictor  estimate    lower    upper
  Exposed1 1.000000       NA       NA
  Exposed2 1.940217 1.153809 3.262623

$p.value
         two-sided
Predictor  midp.exact fisher.exact chi.square
  Exposed1         NA          NA         NA
  Exposed2 0.00305813  0.003444149  0.0019478

$correction
[1] FALSE

attr(,"method")
[1] "Unconditional MLE & normal approximation (Wald) CI"
```

treatment(surgery and radiation therapy) are dependent.

Likelihood Ratio

```
> GTest(CancerControl)

        Log likelihood ratio (G-test) test of independence without
correction

data:  CancerControl
G = 9.9552, X-squared df = 1, p-value = 0.001604
```

Likelihood ratio χ^2 test statistic G^2 =9.9552 with df=(2-1)*(2-1)=1, and p-value 0.0016.

Since the p-value is small enough, one can reject the H_0: $\pi_{ij} = \pi_{i+}\pi_{+j}$. That is, we can say that there is association between the cancer control and type of treatment.

(c)

```
> fisher.test(CancerControl, alternative = "two.sided")

        Fisher's Exact Test for Count Data

data:  CancerControl
p-value = 0.003444
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
   1.755377 126.314516
sample estimates:
odds ratio
  10.98843
```

The H_0 distribution of n_{11} is the hypergeometric distribution. Thus the probability of a particular value n_{11} is $P(n_{11})$ = $n_{1+}Cn_{11}$ * $n_{2+}Cn_{+1}$-n_{11} / nCn_{+1} = $23Cn_{11}$ * 17C29-n_{11} / 40C29 under H_0. Since H_1 is $\theta \neq 1$, p-value is sum of both right and left tail probability that makes $P(n_{11}) \leq P(21)$. We have the constraint $12 \leq n_{11} \leq 23$ and since it is hypergeometric, it will not be symmetric. By calculation, we can get the values of each partial value n_{11}.

n_{11}	12	13	...	20	**21**	22	23
$P(n_{11})$	0.000585	0.008513	...	0.0186231	**0.002660**	0.000193	0.000005

Thus, p-value = P(21)+P(22)+P(23)+P(12) = 0.003444. Since p-value is significantly small, one can conclude that cancer control and type of treatment are dependent.

2.

(a)

```
Standardized Pearson Residuals
           [,1]       [,2]       [,3]
[1,]  0.4061328 -0.1898118 -0.1903291
[2,]  1.5828205 -0.5440627 -0.9459053
[3,] -0.1286367  1.3041565 -1.2374420
[4,] -2.1078423 -0.4031584  2.4360173
```

(c)

$$OR = \frac{n11n}{n41n13} = \frac{9*27}{10*9} = 2.7$$

The estimated odds of study aspiration up to some high school from low income family was 2.7 times of the estimated odds of study aspiration up to college graduate from low income family.

3.

(a) H_0: $\rho = 0$, H_1: $\rho \neq 0$

The sample correlation $\hat{\rho} = 0.1321$.

```
> CMHtest(StudyAsp, rscores=c(1,2,3,4), cscores=c(1,2,3))
Cochran-Mantel-Haenszel Statistics

                 AltHypothesis  Chisq Df     Prob
cor        Nonzero correlation 4.7489  1 0.029317
rmeans  Row mean scores differ 7.2240  3 0.065090
cmeans  Col mean scores differ 4.8673  2 0.087717
general    General association 8.8384  6 0.182870
```

The test statistic for testing H_0 is $M^2 = (273-1) * 0.1321^2 = 4.7489$ with df=1 and p-value 0.0293.

Assuming α=0.05, one can reject H_0 since p-value is smaller than α. That is, there is ~~linear relationship~~ an increasing trend between study aspiration and family income.

(b)

```
> CMHtest(StudyAsp, rscores=c(10,20,30,40), cscores=c(-1,0,1))
Cochran-Mantel-Haenszel Statistics

                 AltHypothesis  Chisq Df     Prob
cor        Nonzero correlation 4.7489  1 0.029317
rmeans  Row mean scores differ 7.2240  3 0.065090
cmeans  Col mean scores differ 4.8673  2 0.087717
general    General association 8.8384  6 0.182870
```

The result is the same if the scores maintain the same relative spacings between categories.

admitted.

(b)

Marginal odds ratio: $\theta_{AG} = \frac{1198*1278}{1493*557} = 1.8411$

The sample odds for male getting admitted is about 1.84 times higher than the sample odds for female getting admitted.

(c)

Dpt	M-Y	M-N	M-total	F-Y	F-N	F-total	Odds
1	512	313	825	89	19	108	0.3492
2	353	207	560	17	8	25	0.8025
3	120	205	325	202	391	593	1.1331
4	138	279	417	131	244	375	0.9213
5	53	138	191	94	299	393	1.2216
6	22	351	373	24	317	341	0.8279
Total	1198	1493	2691	557	1278	1835	1.8411

Unlike the marginal odds (>100%), conditional odds of majority of departments are less than 100%. That is, male are more likely to be admitted in terms of university as a whole, but in terms of each department, female are more likely to enter just except for department 3 and 5. This is called Simpson's paradox, and it came from the association between A(whether admitted) and D(department). Marginal odds ratio ignores D, but in reality it plays an important role to see the bigger picture because of the association.