MATH 7343 : Applied Statistics

Homework 8

Sai Nikhil

NUID: 001564864

T.Sai

① 16.4.5

Zero or One partner

Cancer	Smoker		Total
	Yes	No	
Yes	12	25	37
No	21	118	139
Total	33	143	176

Two or More Partners

Cancer	Smoker		Total
	Yes	No	
Yes	96	92	188
No	142	150	292
Total	238	242	480

(a) Odds ratio for smoker to non-smoker (zero or one partner) $= \dfrac{12 \times 118}{21 \times 25}$

≈ 2.69

(b) Odds ratio (two or more partners) $= \dfrac{96 \times 150}{92 \times 142}$

≈ 1.102

(c) odds ratio > 1 in both cases

$\Rightarrow$ higher for women who do smoke.

(d)

Cancer	Smoker: Yes	Smoker: No	Total
Yes	108	117	225
No	163	268	431
Total	271	385	656

Odds ratio (overall) $= \dfrac{108 \times 268}{117 \times 163}$

≈ 1.51

While combining the data, we didn't take confounders into consideration. Hence, the overall odds-ratio can over/under estimate reel likelihood.

(e) H_0 : $OR_1 = OR_2$

H_A : $OR_1 \neq OR_2$

$$\chi^2 = \sum_{i=1}^{n} \omega_i (y_i - y)^2$$

$$= \omega_1 (y_1 - y)^2 + \omega_2 (y_2 - y)^2$$

$y_1 = \log(OR_1)$	$y_2 = \log(OR_2)$
$= \log(2.69)$	$= \log(1.102)$
≈ 0.9895	≈ 0.0971

$$\omega_1 = \frac{1}{\frac{1}{12} + \frac{1}{25} + \frac{1}{21} + \frac{1}{118}}$$

$$\boxed{\omega_1 \approx 5.57}$$

$$w_2 = \frac{1}{\frac{1}{96} + \frac{1}{92} + \frac{1}{142} + \frac{1}{250}} \simeq 28.57$$

$$y = \frac{w_1 y_1 + w_2 y_2}{w_1 + w_2}$$

$$y = \frac{5.57 \times 0.9805 + 28.57 \times 0.0971}{5.57 + 28.57}$$

$$y \simeq 0.2427$$

$$\chi^2 = 5.57\,(0.9895 - 0.2427)^2 + 28.57\,(0.071 - 0.2372)^2$$

$$\simeq 3.89$$

$$df = 1$$

$$\Rightarrow \quad p > 0.10$$

$\Rightarrow$ Accept H_0

It looks appropriate to use Mantel-Haenszel method.

(f) $$\text{Odds ratio} = \frac{\sum_{i=1}^{2} \frac{a_i d_i}{T_i}}{\sum_{i=1}^{2} \frac{b_i c_i}{T_i}}$$

$$= \frac{\frac{12 \times 118}{176} + \frac{98 \times 180}{480}}{\frac{25 \times 21}{176} + \frac{92 \times 142}{480}}$$

$$\simeq 1.259$$

(g) $$se(Y) = \frac{1}{\sqrt{\omega_1 + \omega_2}}$$

$$= \frac{1}{\sqrt{5.57 + 28.57}}$$

$$\simeq 0.171$$

$\therefore$ 99% CI for $\ln(OR) = (0.2427 - 2.57 \times 0.1711,\ 0.2427 + 2.57 \times 0.1711)$

$\simeq (-0.197,\ 0.6824)$

99% CI for OR $= (e^{-0.197},\ e^{0.6824}) \simeq (0.8211,\ 1.9786)$

Yes. It contains "1".

$\Rightarrow$ It is possible that smoking is not associated with increased likelihood of cancer.

(h)

$$\chi^2 = \frac{\left[\sum_{i=1}^{2} a_i - \sum_{i=1}^{2} m_i\right]^2}{\sum_{i=1}^{2} \sigma_i^2}$$

$$= \frac{12 + 96 - \left(\frac{33 \times 37}{176} + \frac{238 \times 188}{480}\right)^2}{\frac{33 \times 143 \times 37 \times 139}{176^2 \times 175} + \frac{289 \times 242 \times 188 \times 292}{480^2 \times 479}}$$

$\simeq 1.5677$

$df = 1$

$\Rightarrow p > 0.1$

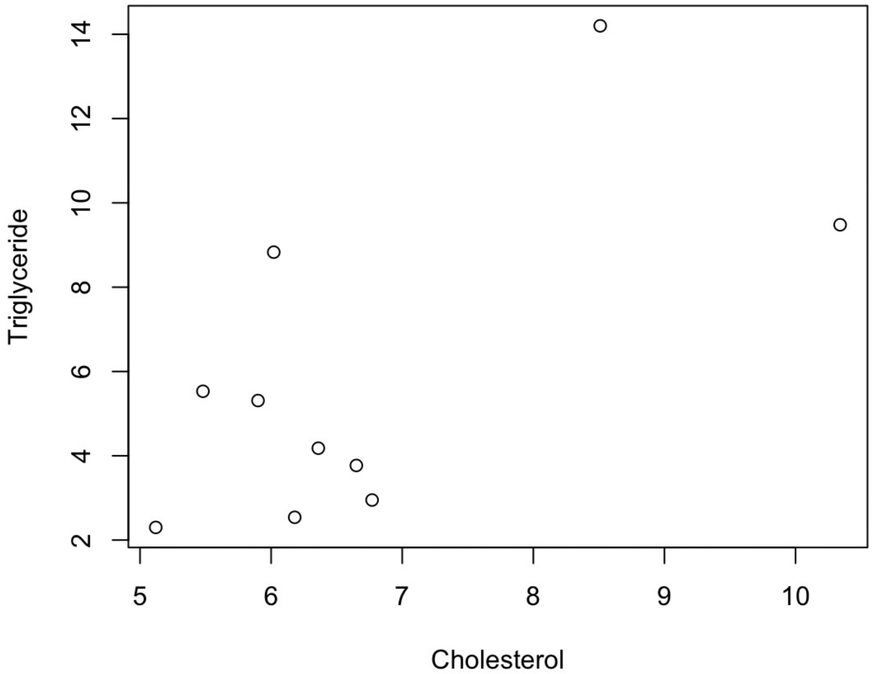
$\Rightarrow$ Accept H_0

$\therefore$ Smoking is not associated with presence of invasive cervical cancer.

② 17.5.5

(a)

```
> plot(data$Cholesterol, data$Triglyceride, xlab = "Cholesterol", ylab = "Triglyceride")
```



(b) From the above graph, it appears like the two variables are linear correlated with little noise distributed around the line of fit.

(c)

```
> (cor(data$Cholesterol, data$Triglyceride, method = "pearson"))
[1] 0.6496543
```

r

(d)

```
> (cor.test(data$Cholesterol, data$Triglyceride, result = "pearson"))

	Pearson's product-moment correlation

data:  data$Cholesterol and data$Triglyceride
t = 2.417, df = 8, p-value = 0.04204
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 0.03389066 0.90790974
sample estimates:
      cor
0.6496543
```

$p = 0.04204 < \alpha = 0.05$

$\Rightarrow$ reject H_0

$\therefore$ we conclude that the population correlation "ρ" is not equal to "0".

$\Rightarrow$ cholesterol and triglyceride are correlated.

(e)

```
> (cor(data$Cholesterol, data$Triglyceride, method = "spearman"))
[1] 0.4181818
```

r_s

(f)

$r_s < r$

$\Rightarrow$ positive relationship between cholesterol and triglyceride

(g)

```
> (cor.test(data$Cholesterol, data$Patient, method = "spearman"))

   Spearman's rank correlation rho

data:  data$Cholesterol and data$Patient
S = 96, p-value = 0.2324
alternative hypothesis: true rho is not equal to 0
sample estimates:
      rho
0.4181818
```

p-value $= 0.2324 > \alpha = 0.05$

$\Rightarrow$ accept H_0 at 5% significance level

$\therefore$ population correlation $= 0$

$\Rightarrow$ They are linearly correlated.

(h)

```
> perm.cor.test<-function(x,y, n.perm=100000, cor.method="spearman")
+   {
+ n<-length(x)
+ T.obs<- cor(x,y, method=cor.method)
+ T.perm = rep(NA, n.perm)
+ for(i in 1:n.perm)
+   {y.perm = sample(y, n, replace=F)
+ T.perm[i] = cor(x, y.perm, method=cor.method)}
+ mean(abs(T.perm)>=abs(T.obs))
+ }
> perm.cor.test(data$Cholesterol, data$Triglyceride, cor.method = "pearson")
[1] 0.04744
> perm.cor.test(data$Cholesterol, data$Triglyceride, cor.method = "spearman")
[1] 0.23605
```

approximately equal to p-value in part (g)

approximately equal to p-value in part (d)