108年台北醫學大學生物統計學分班第八章作業

Bo-Jiang Lin

2019-09-28

#### 1. 下表為一探討心血管疾病與飲酒關係之研究，請問飲酒得心血管疾病與不飲酒得心血管疾病的相對風險為何？當顯著水準設定為0.05時，試利用檢定方法探討飲酒習慣是否與心血疾病的風險有關，其p值為何？兩者是否有顯著的相關？

drink <- matrix(c(16, 12, 784, 188), nrow = 2, ncol = 2, dimnames = list(c("non-Alcohol", "Alcohol"),c("CVD","non-CVD")))  
#drink

riskratio(drink)

## $data  
## CVD non-CVD Total  
## non-Alcohol 16 784 800  
## Alcohol 12 188 200  
## Total 28 972 1000  
##   
## $measure  
## NA  
## risk ratio with 95% C.I. estimate lower upper  
## non-Alcohol 1.0000000 NA NA  
## Alcohol 0.9591837 0.9249095 0.994728  
##   
## $p.value  
## NA  
## two-sided midp.exact fisher.exact chi.square  
## non-Alcohol NA NA NA  
## Alcohol 0.005616838 0.006142125 0.002162492  
##   
## $correction  
## [1] FALSE  
##   
## attr(,"method")  
## [1] "Unconditional MLE & normal approximation (Wald) CI"

飲酒相較於未飲酒者，其心血管疾病發生率0.959倍(p-value = 0.002 < 0.05)。

#### 2. 承上題，請問飲酒習慣得到心血管疾病相對於沒有飲酒習慣的勝算比為何？勝算比的95%信賴區間為何？當顯著水準為0.05時，心血管疾病是否與飲酒有顯著的相關？

## $data  
## CVD non-CVD Total  
## non-Alcohol 16 784 800  
## Alcohol 12 188 200  
## Total 28 972 1000  
##   
## $measure  
## NA  
## odds ratio with 95% C.I. estimate lower upper  
## non-Alcohol 1.0000000 NA NA  
## Alcohol 0.3191945 0.1481929 0.7060634  
##   
## $p.value  
## NA  
## two-sided midp.exact fisher.exact chi.square  
## non-Alcohol NA NA NA  
## Alcohol 0.005616838 0.006142125 0.002162492  
##   
## $correction  
## [1] FALSE  
##   
## attr(,"method")  
## [1] "median-unbiased estimate & mid-p exact CI"

飲酒相較未飲酒，其發生心血管疾病的勝算是0.31倍(P-value = 0.002 < 0.05)。

#### 3. 試利用肺癌資料(附件 lung\_cancer\_study.csv)分析回答以下問題

#names(lung)

###### 1. 利用簡單邏輯斯迴歸分析復發情形與性別的關係，請問男性相對於女性復發的勝算比為何？95%信賴區間為何？若顯著水準為0.05，復發與性別是否有顯著的關係？

logit <- glm(lung$FIRST\_PROGRESSION\_OR\_RELAPSE ~ lung$gender, na.action = na.exclude, family = binomial(link = "logit"))  
table(lung$FIRST\_PROGRESSION\_OR\_RELAPSE,lung$gender,dnn = c("FIRST\_PROGRESSION\_OR\_RELAPSE","gender"))

## gender  
## FIRST\_PROGRESSION\_OR\_RELAPSE 0 1  
## No 98 73  
## Yes 105 114

summary(logit)

##   
## Call:  
## glm(formula = lung$FIRST\_PROGRESSION\_OR\_RELAPSE ~ lung$gender,   
## family = binomial(link = "logit"), na.action = na.exclude)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.3716 -1.2068 0.9949 1.1483 1.1483   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)   
## (Intercept) 0.06899 0.14046 0.491 0.6233   
## lung$gender 0.37675 0.20542 1.834 0.0667 .  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 534.73 on 389 degrees of freedom  
## Residual deviance: 531.35 on 388 degrees of freedom  
## (88 observations deleted due to missingness)  
## AIC: 535.35  
##   
## Number of Fisher Scoring iterations: 4

男性相對於女性的復發勝算比為：

exp(logit$coef[2])

## lung$gender   
## 1.457534

男性相對於女性的復發勝算比之95% CI：

confint(logit)

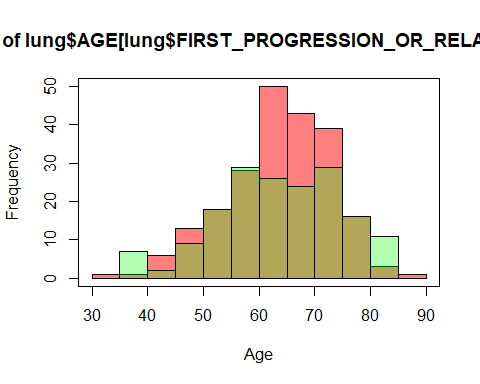
## Waiting for profiling to be done...

## 2.5 % 97.5 %  
## (Intercept) -0.20629587 0.3451557  
## lung$gender -0.02475887 0.7812139

復發與性別未有有顯著的關係(p-value = 0.06 > 0.05)。

###### 2. 利用簡單邏輯斯迴歸分析復發情形與年齡的關係，若顯著水準為0.05，復發與年齡是否有顯著的關係？

hist(lung$AGE[lung$FIRST\_PROGRESSION\_OR\_RELAPSE == "Yes"], col = rgb(1, 0, 0, 0.5), xlab = "Age")  
hist(lung$AGE[lung$FIRST\_PROGRESSION\_OR\_RELAPSE == "No"], col = rgb(0, 1, 0, 0.3), xlab = "Age", add = TRUE)  
box()  
legend(30, 90, legend = c("Recurence", "Non-recurence"), text.col = c(rgb(1, 0, 0, 0.5), rgb(0, 1, 0, 0.3)), pch = c(22, 22), pt.bg = c(rgb(1, 0, 0, 0.5), rgb(0, 1, 0, 0.3)))



logit\_age <- glm(lung$FIRST\_PROGRESSION\_OR\_RELAPSE ~ lung$AGE, family = binomial(link = "logit"), na.action = na.exclude)  
summary(logit\_age)

##   
## Call:  
## glm(formula = lung$FIRST\_PROGRESSION\_OR\_RELAPSE ~ lung$AGE, family = binomial(link = "logit"),   
## na.action = na.exclude)  
##   
## Deviance Residuals:   
## Min 1Q Median 3Q Max   
## -1.290 -1.284 1.072 1.075 1.078   
##   
## Coefficients:  
## Estimate Std. Error z value Pr(>|z|)  
## (Intercept) 0.2758336 0.6561073 0.420 0.674  
## lung$AGE -0.0004435 0.0101106 -0.044 0.965  
##   
## (Dispersion parameter for binomial family taken to be 1)  
##   
## Null deviance: 534.73 on 389 degrees of freedom  
## Residual deviance: 534.73 on 388 degrees of freedom  
## (88 observations deleted due to missingness)  
## AIC: 538.73  
##   
## Number of Fisher Scoring iterations: 3

復發跟年齡未有顯著關係(P-value = 0.965 > 0.05)。