

Homework 1, Biostatistics

90

Q1

```
binom.test(12,85,p=0.5,alt='two.sided',conf.level = 0.95)
```

```
##
##  Exact binomial test
##
## data:  12 and 85
## number of successes = 12, number of trials = 85, p-value = 8.091e-12
## alternative hypothesis: true probability of success is not equal to 0.5
## 95 percent confidence interval:
##  0.07513135 0.23362462
## sample estimates:
## probability of success
##           0.1411765
```

```
binom.test(12,85,p=0.5,alt='two.sided',conf.level = 0.99)
```

```
##
##  Exact binomial test
##
## data:  12 and 85
## number of successes = 12, number of trials = 85, p-value = 8.091e-12
## alternative hypothesis: true probability of success is not equal to 0.5
## 99 percent confidence interval:
##  0.06036021 0.26403358
## sample estimates:
## probability of success
##           0.1411765
```

1.the confidence interval for spleen rate at $\alpha = 0.05$ is (0.07513135,0.3362462)

2.the confidence interval for spleen rate at $\alpha = 0.01$ is (0.06036021,0.26403358)

Q2

(a)

we can create a table

table0

	dead	survival	Total
control	4	15	19
case	2	13	15
Total	6	28	34

if $H_0 : p_1 = p_2$ vs $H_1 : p_1 > p_2$

table1

	dead	survival	Total
control	5	14	19

	dead	survival	Total
case	1	14	15
Total	6	28	34

table2

	dead	survival	Total
control	6	13	19
case	0	15	15
Total	6	28	34

$p_value = \Pr(\text{table1}|H_0) + \Pr(\text{table2}|H_0) + \Pr(\text{table0}|H_0)$

```
p_value=0
for (i in c(4:6)){
  p_value = p_value + choose(6,i)*choose(28,19-i)/choose(34,19)
}
print(p_value)
```

```
## [1] 0.4524724
```

check

```
m=matrix(c(4,2,15,13),c(2,2),dimnames = list(c('control','streptokinase'),c('died','survival')))
fisher.test(m,alt='greater')$p.value
```

```
## [1] 0.4524724
```

if $H_0 : p1 = p2$ vs $H_1 : p1 < p2$

table3

	dead	survival	Total
control	3	16	19
case	3	12	15
Total	6	28	34

table4

	dead	survival	Total
control	2	17	19
case	4	11	15
Total	6	28	34

table5

	dead	survival	Total
control	1	18	19
case	5	10	15
Total	6	28	34

table6

	dead	survival	Total
control	0	19	19
case	6	9	15
Total	6	28	34

$p_value = \Pr(\text{table3}|H_0) + \Pr(\text{table4}|H_0) + \Pr(\text{table5}|H_0) + \Pr(\text{table6}|H_0) + \Pr(\text{table0}|H_0)$

```
p_value=0
for (i in c(4:0)){
  p_value = p_value + choose(6,i)*choose(28,19-i)/choose(34,19)
}
print(p_value)
```

```
## [1] 0.8501365
```

check

```
m=matrix(c(4,2,15,13),c(2,2),dimnames = list(c('control','streptokinase'),c('died','survival')))
fisher.test(m,alt='less')$p.value
```

```
## [1] 0.8501365
```

if $H_0 : p1 = p2$ vs $H_1 : p1 \neq p2$

p-value = take sum all table when $\Pr(\text{tablei}|H_0) < \Pr(\text{table0}|H_0)$

```
p_value=0
for (i in c(0:6)){
  r= choose(6,i)*choose(28,19-i)/choose(34,19)
  if (r<=choose(6,4)*choose(28,19-4)/choose(34,19)){
    p_value = p_value + r
  }
}
print(p_value)
```

```
## [1] 0.6721736
```

check

```
m=matrix(c(4,2,15,13),c(2,2),dimnames = list(c('control','streptokinase'),c('died','survival')))
fisher.test(m,$p.value
```

```
## [1] 0.6721736
```

(b)

```
m=matrix(c(4,2,15,13),c(2,2),dimnames = list(c('control','streptokinase'),c('died','survival')))
chisq.test(m)
```

```
## Warning in chisq.test(m): Chi-squared approximation may be incorrect
```

```
##
## Pearson's Chi-squared test with Yates' continuity correction
##
## data: m
## X-squared = 0.017753, df = 1, p-value = 0.894
```

Conclusion : whether fisher exact test or chi_square test always do not reject H_0 (i.e mortality rate is equal)

(c)

power = 90% , $P_1 = 0.2$

$H_0 : P_2 = 0.2$ vs $H_1 : P_2 \neq 0.2$

at $\alpha=0.05$

$\hat{P}_1 \sim N(0.2 \times n, 0.16/n)$

$\hat{P}_2 \sim N(P_2 \times n, P_2(1 - P_2)/n)$

$\hat{P}_2 - \hat{P}_1 \sim N((P_2 - 0.2) \times n, (0.16 + P_2(1 - P_2))/n)$

critical region = $\frac{\hat{P}_2 - \hat{P}_1}{\sqrt{0.32/n}} > z_{\alpha/2}$

= $\hat{P}_2 - \hat{P}_1 > z_{\alpha/2} \times \sqrt{\frac{0.32}{n}}$

$0.45 = \Pr(\text{critical region} \mid H_1 \text{ is true}) = \Pr(\hat{P}_2 - \hat{P}_1 > z_{\alpha/2} \times \sqrt{\frac{0.32}{n}} \mid P_2)$

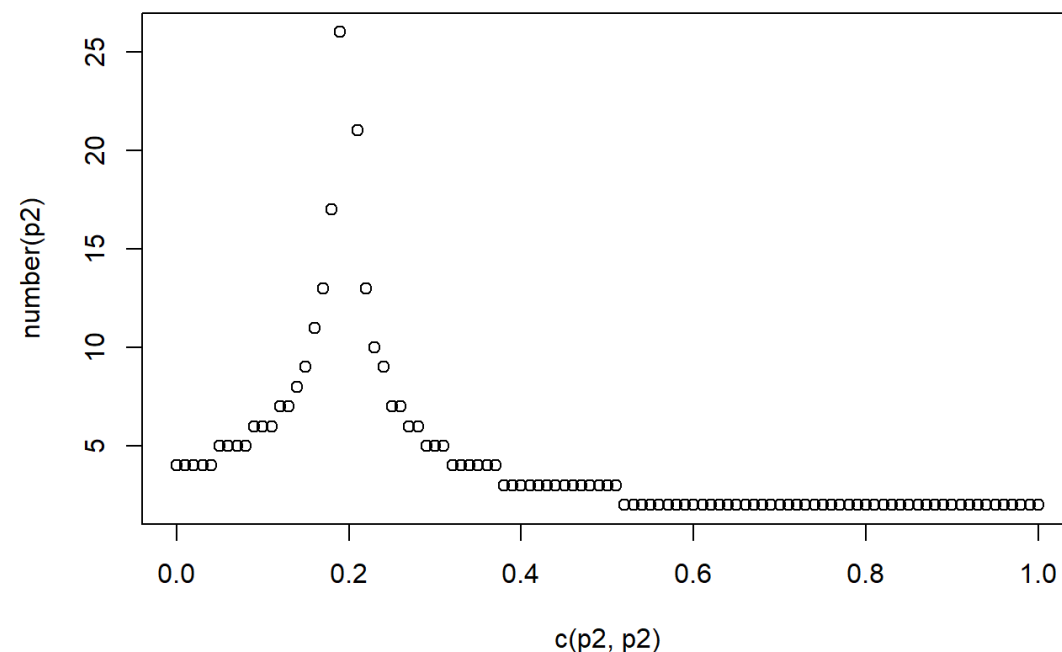
$0.45 = \Pr(Z > (z_{\alpha/2} \times \sqrt{\frac{0.32}{n}} - (P_2 - 0.2) \times n) / \sqrt{\frac{0.16 + P_2(1 - P_2)}{n}})$

and $z_{0.45} = 0.3264$

$n^{3/2} = (-0.3264 \times \sqrt{0.16 + P_2(1 - P_2)} + 1.96 \times \sqrt{0.32}) / (P_2 - 0.2)$

```
ceiling_dec <- function(x, level=1) round(x + 5*10^(-level-1), level)

number=function(p2){
  ans1<-((-0.3264*(0.16+p2*(1-p2))**0.5 - 1.96*(0.32)**0.5) / (p2-0.2))**(2/3)
  ans2<-((-0.3264*(0.16+p2*(1-p2))**0.5 + 1.96*(0.32)**0.5) / (p2-0.2))**(2/3)
  ans = ceiling_dec(c(ans1,ans2),0)
  return (ans)
}
p2=seq(0,1,0.01)
plot(c(p2,p2),number(p2))
```

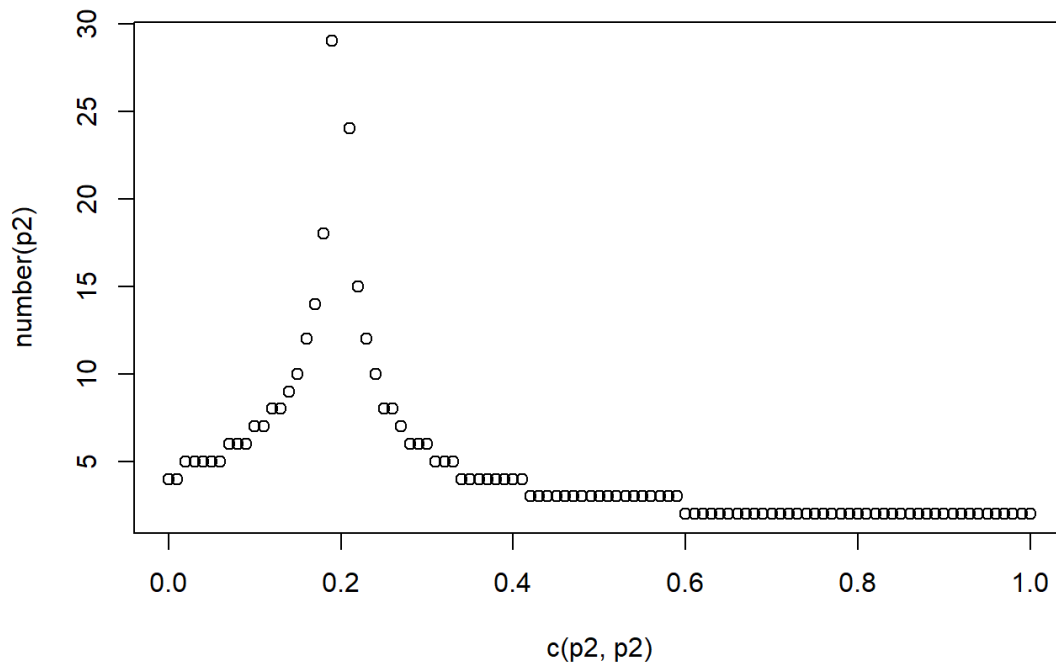


at $\alpha=0.01$

```

number=function(p2){
  ans1<-((-0.3264*(0.16+p2*(1-p2))**0.5 - 2.33*(0.32)**0.5)/(p2-0.2))**(2/3)
  ans2<-((-0.3264*(0.16+p2*(1-p2))**0.5 + 2.33*(0.32)**0.5)/(p2-0.2))**(2/3)
  ans = ceiling_dec(c(ans1,ans2),0)
  return (ans)
}
p2=seq(0,1,0.01)
plot(c(p2,p2),number(p2))

```



Q3

(a)

$H_0 : p = 0.25$ vs $H_1 : p \neq 0.25$

```

Variable = c('ID number','Age','FEV','Height','Sex','Smoking status')
data=read.table("C:\\Users\\xx958\\OneDrive\\桌面\\nycu 碩一_上\\生物統計\\HW01\\FEV.TXT",col.names=Variable)
binom.test(sum(data['FEV']>3),length(data[[1]]),p=0.25,alt='two.sided')

```

```

##
## Exact binomial test
##
## data: sum(data["FEV"] > 3) and length(data[[1]])
## number of successes = 203, number of trials = 654, p-value = 0.0004967
## alternative hypothesis: true probability of success is not equal to 0.25
## 95 percent confidence interval:
##  0.2750987 0.3474182
## sample estimates:
## probability of success
##           0.3103976

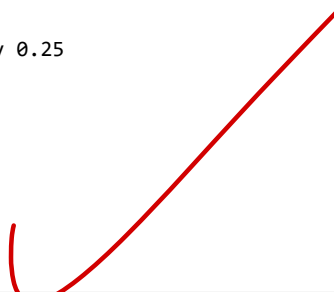
```

```

prop.test(sum(data['FEV']>3),length(data[[1]]),p=0.25,alt='two.sided')

```

```
##
## 1-sample proportions test with continuity correction
##
## data: sum(data["FEV"] > 3) out of length(data[[1]]), null probability 0.25
## X-squared = 12.404, df = 1, p-value = 0.0004285
## alternative hypothesis: true p is not equal to 0.25
## 95 percent confidence interval:
##  0.2753966 0.3476605
## sample estimates:
##           p
## 0.3103976
```



Conclusion : since p-value = 0.0004967 or 0.0004285 < 0.05 = α Thus reject H_0 (i.e the proportion of having perfect pulmonary function is not equal to 0.25)

(b)

H_0 : odds ratio = 1 vs H_1 : odds ratio \neq 1

```
data['perfect_pulmonary'] = as.factor(data['FEV']>3)

table_Sex_and_perfect_pulmonary <- data%>%
  group_by(Sex,perfect_pulmonary)%>%
  summarise(n=n())
```

```
## `summarise()` has grouped output by 'Sex'. You can override using the `.groups` argument.
```

```
m = matrix(table_Sex_and_perfect_pulmonary$n,c(2,2),dimnames = list(c('Female','male'),c('no','yes')),byrow = T)
print(m)
```

```
##           no yes
## Female 240  78
## male   211 125
```

```
R1=m[2,2]/(m[2,1]+m[2,2])
R0=m[1,2]/(m[1,1]+m[1,2])
odds1=R1/(1-R1)
odds0=R0/(1-R0)
OR = odds1/odds0
cat('odds ratio is ',OR)
```

```
## odds ratio is  1.822822
```

check

```
fisher.test(m)$estimate
```

```
## odds ratio
##  1.821169
```

```
chisq.test(m,correct = T)
```

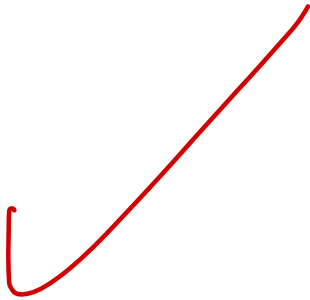
```
##
## Pearson's Chi-squared test with Yates' continuity correction
##
## data:  m
## X-squared = 11.675, df = 1, p-value = 0.0006333
```

```
prop.test(m[1:2,2],rowSums(m))
```

```
##
## 2-sample test for equality of proportions with continuity correction
##
## data:  m[1:2, 2] out of rowSums(m)
## X-squared = 11.675, df = 1, p-value = 0.0006333
## alternative hypothesis: two.sided
## 95 percent confidence interval:
## -0.19985286 -0.05362873
## sample estimates:
##      prop 1      prop 2
## 0.2452830 0.3720238
```

```
prop.test(m[1:2,2],rowSums(m),alt='less')
```

```
##
## 2-sample test for equality of proportions with continuity correction
##
## data:  m[1:2, 2] out of rowSums(m)
## X-squared = 11.675, df = 1, p-value = 0.0003167
## alternative hypothesis: less
## 95 percent confidence interval:
## -1.00000000 -0.06489118
## sample estimates:
##      prop 1      prop 2
## 0.2452830 0.3720238
```



Conclusion : since $p_value = 0.000633 < 0.05 = \alpha$, Thus reject H_0 (i.e odds ratio $\neq 1$) , furthermore male is superior than female at perfect_pulmonary

(c)

```
data['age_class']=1
data['age_class'][4<data['Age'] & data['Age']<10] = 1
data['age_class'][9<data['Age'] & data['Age']<15] = 2
data['age_class'][14<data['Age'] & data['Age']<20] = 3
```

```
table_age_class_and_perfect_pulmonary <- data%>%
  group_by(age_class,perfect_pulmonary)%>%
  summarise(n=n())
```

```
## `summarise()` has grouped output by 'age_class'. You can override using the `.groups` argument.
```

```
m = matrix(table_age_class_and_perfect_pulmonary$n,c(3,2),dimnames = list(c('5~9','10~14','15~20'),c('no','yes')),byrow = T)
print(m)
```

```
##           no yes
## 5~9      298  11
## 10~14    141 155
## 15~20     12  37
```

$H_0 : P1 = P2 = P3$ vs $H_1 : not H_0$

the number of perfect_pulmonary in 5~9 is binomial distribution (309,P1) P1 代表 5~9 中 perfect_pulmonary 的比例

the number of perfect_pulmonary in 10~14 is binomial distribution (296,P2) P2 代表 10~14 中 perfect_pulmonary 的比例

the number of perfect_pulmonary in 15~20 is binomial distribution (49,P3) P3 代表 15~20 中 perfect_pulmonary 的比例

```
chisq.test(m,correct = T)
```

```
##
## Pearson's Chi-squared test
##
## data: m
## X-squared = 217.17, df = 2, p-value < 2.2e-16
```

Conclusion : $p_value \ll 0.05 = \alpha$, reject H_0 (i.e 比例不同)

$H_0 : P1 = P2$ vs $H_1 : P1 < P2$

```
prop.test(m[1:2,2],rowSums(m)[1:2],correct = T,alt='less')
```

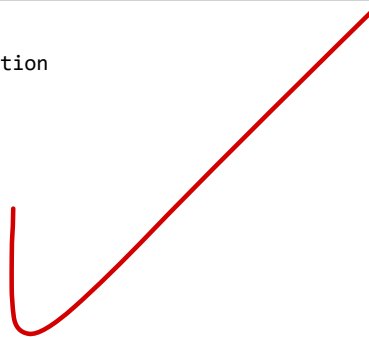
```
##
## 2-sample test for equality of proportions with continuity correction
##
## data: m[1:2, 2] out of rowSums(m)[1:2]
## X-squared = 178.42, df = 1, p-value < 2.2e-16
## alternative hypothesis: less
## 95 percent confidence interval:
## -1.0000000 -0.4339433
## sample estimates:
## prop 1 prop 2
## 0.03559871 0.52364865
```

Conclusion : $p_value \ll 0.05 = \alpha$, reject H_0 (i.e 10~14 比 5~9 have a higher proportion of having perfect pulmonary function)

$H_0 : P2 = P3$ vs $H_1 : P2 < P3$

```
prop.test(m[2:3,2],rowSums(m)[2:3],correct = T,alt='less')
```

```
##
## 2-sample test for equality of proportions with continuity correction
##
## data: m[2:3, 2] out of rowSums(m)[2:3]
## X-squared = 8.2115, df = 1, p-value = 0.002081
## alternative hypothesis: less
## 95 percent confidence interval:
## -1.0000000 -0.1077991
## sample estimates:
## prop 1 prop 2
## 0.5236486 0.7551020
```



Conclusion : $p_value=0.00208 < 0.05 = \alpha$, reject H_0 (i.e 15~20 比 10~14 have a higher proportion of having perfect pulmonary function)

Q4

(a)

P_i 代表各城市中 Number children with caries free teeth 的比例

```
m=matrix(c(243,16,83,36,60,32,31,31,39,12),c(2,5),dimnames = list(c('Number_children_with_caries','Number_children_with_caries_free_teeth'),c('Essex','Slough','Harwick','Burnham','Meres'))))
```

$H_0 : P1 = P2 = P3 = P4 = P5$ vs $H_1 : not H_0$

```
chisq.test(m)
```

```
##
## Pearson's Chi-squared test
##
## data: m
## X-squared = 80.216, df = 4, p-value < 2.2e-16
```



```
prop.test(m[2,],colSums(m))
```

```
##  
## 5-sample test for equality of proportions without continuity  
## correction  
##  
## data: m[2, ] out of colSums(m)  
## X-squared = 80.216, df = 4, p-value < 2.2e-16  
## alternative hypothesis: two.sided  
## sample estimates:  
##      prop 1      prop 2      prop 3      prop 4      prop 5  
## 0.06177606 0.30252101 0.34782609 0.50000000 0.23529412
```

Conclusion : since $p\text{-value} \ll 0.05 = \alpha$, Thus reject H_0

(b)

```
prop.test(m[2,4:5],colSums(m[1:2,4:5]),alt='greater')
```

```
##  
## 2-sample test for equality of proportions with continuity correction  
##  
## data: m[2, 4:5] out of colSums(m[1:2, 4:5])  
## X-squared = 7.2326, df = 1, p-value = 0.00358  
## alternative hypothesis: greater  
## 95 percent confidence interval:  
##  0.1038173 1.0000000  
## sample estimates:  
##      prop 1      prop 2  
## 0.5000000 0.2352941
```

Express $P_4 > P_5$

```
p = c(0.06177606 , 0.30252101 , 0.34782609 , 0.50000000 , 0.23529412 )  
N = length(p)  
value = critical.range = c()  
  
## Compute critical values.  
for (i in 1:(N-1))  
{ for (j in (i+1):N)  
  {  
    value = c(value,(abs(p[i]-p[j])))  
    critical.range = c(critical.range,  
      sqrt(qchisq(.95,4))*sqrt(p[i]*(1-p[i])/300 + p[j]*(1-p[j])/300))  
  }  
}  
  
m=round(cbind(value,critical.range),3)  
m[,1] > m[,2]
```

```
## [1] TRUE TRUE TRUE TRUE FALSE TRUE FALSE TRUE FALSE TRUE
```

Can see that

(P1 , P2),(P1 , P3),(P1 , P4),(P1 , P5)

(P2 , P4),(P3 , P4),(P4 , P5) there are different ,but not express proportions of children caries free teeth high with fluoride high