Bioinformatics 525: Module 2

Introduction to Statistics

Lab #3

Read TROPHY.csv data in RStudio using "Import Dataset" on the Workspace Window.
IMPORTATNT: type attach(TROPHY) to have the variables accessible for analysis.

1. **One-sample test:** Compare the proportion for a binary outcome to a given proportion value p_0
   H_0: $p = p_0$ vs. H_A: $p \neq p_0$.

   a. Create the following binary variables indicating a risk factor using a cutoff point. The "risk" category is coded as 1.

      HDL01=1 if HDL ≤ 40
      trig01=1 if Triglyceride ≥ 150
      obese=1 if BMI24 ≥ 30

      HDL01=1*(HDL ≤ 40);trig01=1*(Triglyceride >=150); obese=1*(BMI24 ≥ 30)

   b. What are the number and the proportion of subjects with risk for the above variables?

      table(HDL01)
      prop.table(table(HDL01)

   c. Visually display the number and the proportion of subjects with and without risk.

      barplot(table(HDL01),names=c("HDL > 40", "HDL <=40"), main="Number of Subjects with HDL<=40 and HDL > 40")

      barplot(prop.table(table(HDL01)),names=c("HDL > 40", "HDL <=40"),main="Proportion with HDL<=40 and HDL > 40")

   d. Test if the proportion of subjects with HDL ≤ 40 is different from 40%: H_0: $p = .4$ vs. H_A: $p \neq .4$ using the chi-square test.

      chisq.test(table(HDL01),p=c(.6,.4))

   e. Repeat d), but test H_0: $p = .4$ vs. H_A: $p \neq .4$ separately for males and females. Are the results the same?

      chisq.test(table(HDL01[Sex=="Female"]),p=c(.6,.40))
      chisq.test(table(HDL01[Sex=='Male']),p=c(.6,.40))

2. **<u>Two-sample test:</u>** Compare the proportions of a binary outcome between two groups. H_0: $p_1 = p_2$ vs. H_A: $p_1 \neq p_2$.

   a. Calculate the number and the % of males/females with HDL <=40 and HDL > 40

	Female	Male	Total
HDL > 40	95(.9))	83(.55)	178
HDL ≤40	10(p1=.1)	67(p2=.45)	77(p=.30)
Total	105	150	255

```
table(HDL01,Sex)
prop.table(table(HDL01,Sex),margin=2)
```

   b. Visually display the number and the proportion of subjects with HDL <=40 and HDL > 40 for males and females. Is there any evidence of a difference by sex?

```
barplot(table(HDL01,Sex),main="Number of Subjects with HDL <= 40")
barplot(prop.table(table(HDL01,Sex),margin=2),main="Proportion with HDL <= 40")
```

   c. Test whether the proportions of subjects with HDL <= 40 are different between males and females: H_0: $p_1 = p_2$ vs. H_A: $p_1 \neq p_2$ using one of the following tests:

      i. Chi-square test

```
chisq.test(HDL01,Sex)
```

      or

```
chisq.test(matrix(c(95,83,10,67),2,2))
```

      ii. Fisher's exact test

```
fisher.test(HDL01,Sex)
fisher.test(matrix(c(95,83,10,67),2,2))
```

3. **<u>Logistic Regression</u>** for predicting obesity (in each of the treatment group).

   a. Test if trig01 variable predicts obesity for subjects in the <u>placebo group</u>. Calculate the ARR, RR, and OR for obesity outcome by trig01 (< 150 vs. ≥ 150).
      i. Use Chi-square test

```
chisq.test(obese[Trt==2],trig01[Trt==2])
```

ii. Use Fisher's exact test

fisher.test(obese[Trt==2],trig01[Trt==2])

prop.table(table(obese[Trt==2],trig01[Trt==2],marging=2)
ARR=.489-.243
RR=.489/.243
OR=(.489/(1-.489))/(.243/(1-.243))

OR here is Trig < 150 vs >=150
OR from Fisher's exact test is Trig >=150 vs. <150

b. Repeat the same analysis as in a), but for subjects in the candesartan group. Are the findings from a) and b) the same? Does candesartan changes/modifies the effect of triglyceride on predicting obesity?

chisq.test(obese[Trt==1],trig01[Trt==1])
fisher.test(obese[Trt==1],trig01[Trt==1])

prop.table(table(obese[Trt==1],trig01[Trt==1]),margin=2)
ARR=.308-.312
RR=.308/.312
OR=(.308/(1-.308))/(.312/(1-.312))

c. For the placebo group, use logistic regression model to predict obesity by trig01. Calculate the OR and compare it with the OR obtained from Fisher's exact test.

$$\log\left(\frac{\Pr(Obese=1)}{1-Pt(Obese=1)}\right)= \beta_0+ \beta_1*\text{trig01}$$

OR(trig01)=exp(β_1)

glm(obese[Trt==2]~trig01[Trt==2],binomial)
summary(glm(obese[Trt==2]~trig01[Trt==2],binomial))
OR(trig01)=exp(1.09)=2.98

d. Repeat c), but use the following variables, trig01, insulin, and SBP0 to predict obesity. Calculate the OR for trig01 (i.e. adjusted OR). Compare the unadjusted OR from c) with the adjusted OR from d), are they the same?

$$\log\left(\frac{\Pr(Obese=1)}{1-Pt(Obese=1)}\right) = \beta_0 + \beta_1 * \text{trig01} + \beta_2 * \text{Insulin} + \beta_3 * \text{SBP0}$$

Adjusted OR for trig01: OR(trig01)=exp(β_1)

summary(glm(obese[Trt==2]~trig01[Trt==2]+Insulin[Trt==2]+SBP0[Trt==2],binomial))
OR(trig01)=exp(0.77)=2.16