stsci4110hw6

Nick Gembs

11/29/2022

1.

a. 𝜋+1 - 𝜋1+ is the probability of Bryant making the second free throw minus the probability of Bryant making the first free throw. A larger absolute value of this number indicates dependance of the first on the second free throw.

b.

𝐻0: 𝜋12 = 𝜋21 ≡ 𝐻0: 𝜋+1 = 𝜋1+ 𝐻𝐴: 𝜋12 ≠ 𝜋21 ≡ 𝐻0: 𝜋+1 ≠ 𝜋1+

M = (33-37)^2/(33+37)  
1-pchisq(M,1,lower.tail = T)

## [1] 0.6325851

p-value > .05, fail to reject the null.

Conclude: Table margins are homogeneous. 𝜋12 = 𝜋21 ≡ 𝐻0: 𝜋+1 = 𝜋1+ The probability of Bryant making the second free throw does not change based on the outcome of the first free throw

c.

library(PropCIs)  
diffpropci.mp(33, 37, 230, conf.level = 0.95)

##   
##   
##   
## data:   
##   
## 95 percent confidence interval:  
## -0.05390917 0.08839193  
## sample estimates:  
## [1] 0.01724138

0 is included in the 95% CI.Therefore, there is not enough evidence to prove non-homogeneity.

2.

a.

df = read.csv("C:/Users/Nick/Downloads/afterlife2.csv")

df$belief = factor(df$belief, ordered = F)  
df$race = factor(df$race, ordered = F)  
df$gender = factor(df$gender, ordered = F)  
df$belief = relevel(df$belief, ref = "No")

library(tidyr)  
  
df = uncount(df, count, .remove = TRUE, .id = NULL)

library(nnet)

## Warning: package 'nnet' was built under R version 4.2.2

bcl = multinom(belief ~ race+gender+religiosity, data = df)

## # weights: 15 (8 variable)  
## initial value 1231.544376   
## iter 10 value 754.526566  
## final value 751.751592   
## converged

summary(bcl)

## Call:  
## multinom(formula = belief ~ race + gender + religiosity, data = df)  
##   
## Coefficients:  
## (Intercept) raceWhite genderMale religiosity  
## Undecided -0.6942144 0.1944430 -0.1445318 0.06299078  
## Yes -1.0998846 0.3690174 -0.4902732 0.89822127  
##   
## Std. Errors:  
## (Intercept) raceWhite genderMale religiosity  
## Undecided 0.3194276 0.2635508 0.2306095 0.10141679  
## Yes 0.2517191 0.2006170 0.1766465 0.07473111  
##   
## Residual Deviance: 1503.503   
## AIC: 1519.503

b.

# Undecided = -0.6942144 + 0.1944430\*raceWhite + -0.1445318\*genderMale + 0.06299078\*religiosity  
  
# Yes = -1.0998846 + 0.3690174\*raceWhite + -0.4902732\*genderMale + 0.89822127\*religiosity

# odds of “yes” versus “no” for the two genders  
  
odds = exp(-0.4902732)  
odds

## [1] 0.612459

cat("being male has a 0.612459 multiplicative effect on odds of being 'yes' belief over 'no'\n")

## being male has a 0.612459 multiplicative effect on odds of being 'yes' belief over 'no'

c.

# odds of “yes” versus “no” for the two races  
  
odds = exp(0.3690174)  
odds

## [1] 1.446313

cat("being white has a 1.446313 multiplicative effect on odds of being 'yes' belief over 'no'\n")

## being white has a 1.446313 multiplicative effect on odds of being 'yes' belief over 'no'

# change in odds of “yes” versus “no” for a one unit increase in religiosity  
  
odds = exp(0.89822127)  
odds

## [1] 2.455232

cat("A one unit increase in religiosity has a 2.455232 multiplicative effect on odds of being 'yes' belief over 'no'\n")

## A one unit increase in religiosity has a 2.455232 multiplicative effect on odds of being 'yes' belief over 'no'

d.

library(lmtest)

## Loading required package: zoo

##   
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':  
##   
## as.Date, as.Date.numeric

#HO : gender is not significant in predicting belief  
#H1 : gender is significant in predicting belief  
bclg = multinom(belief ~ race+religiosity, data = df)

## # weights: 12 (6 variable)  
## initial value 1231.544376   
## iter 10 value 756.011597  
## final value 756.008290   
## converged

lrtest(bcl,bclg)

## Likelihood ratio test  
##   
## Model 1: belief ~ race + gender + religiosity  
## Model 2: belief ~ race + religiosity  
## #Df LogLik Df Chisq Pr(>Chisq)   
## 1 8 -751.75   
## 2 6 -756.01 -2 8.5134 0.01417 \*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

# p-value<.05, reject null

e.

#HO : race is not significant in predicting belief  
#H1 : race is significant in predicting belief  
bclrace = multinom(belief ~ gender+religiosity, data = df)

## # weights: 12 (6 variable)  
## initial value 1231.544376   
## iter 10 value 753.442157  
## final value 753.432581   
## converged

lrtest(bcl,bclrace)

## Likelihood ratio test  
##   
## Model 1: belief ~ race + gender + religiosity  
## Model 2: belief ~ gender + religiosity  
## #Df LogLik Df Chisq Pr(>Chisq)  
## 1 8 -751.75   
## 2 6 -753.43 -2 3.362 0.1862

# p-value>.05, fail to reject null

f.

#HO : religiosity is not significant in predicting belief  
#H1 : religiosity is significant in predicting belief  
bclr = multinom(belief ~ race+gender, data = df)

## # weights: 12 (6 variable)  
## initial value 1231.544376   
## iter 10 value 882.187101  
## final value 882.187030   
## converged

lrtest(bcl,bclr)

## Likelihood ratio test  
##   
## Model 1: belief ~ race + gender + religiosity  
## Model 2: belief ~ race + gender  
## #Df LogLik Df Chisq Pr(>Chisq)   
## 1 8 -751.75   
## 2 6 -882.19 -2 260.87 < 2.2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

# p-value<.05, reject null

g.

bclrace = multinom(belief ~ gender+religiosity, data = df)

## # weights: 12 (6 variable)  
## initial value 1231.544376   
## iter 10 value 753.442157  
## final value 753.432581   
## converged

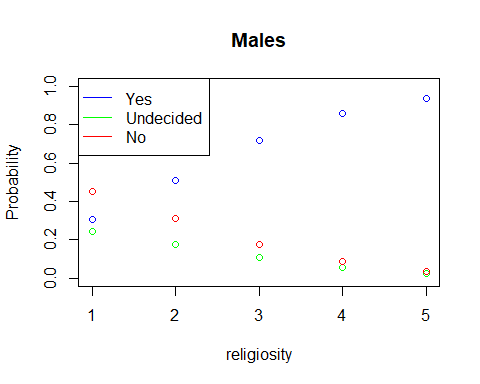
summary(bclrace)

## Call:  
## multinom(formula = belief ~ gender + religiosity, data = df)  
##   
## Coefficients:  
## (Intercept) genderMale religiosity  
## Undecided -0.5578253 -0.1295137 0.06309689  
## Yes -0.8341452 -0.4617498 0.89802444  
##   
## Std. Errors:  
## (Intercept) genderMale religiosity  
## Undecided 0.2591456 0.2296061 0.10138932  
## Yes 0.2049019 0.1754738 0.07460479  
##   
## Residual Deviance: 1506.865   
## AIC: 1518.865

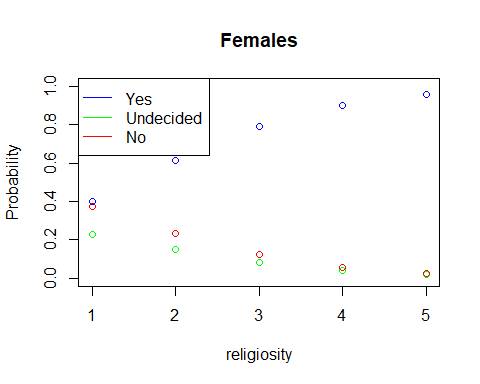
# Undecided = -0.5578253 -0.1295137\*genderMale + 0.06309689\*religiosity  
# Yes = -0.8341452 -0.4617498\*genderMale + 0.89802444\* religiosity

h.

newdf = data.frame(df$religiosity, fitted(bclrace), df$gender)  
  
plot(newdf$df.religiosity[newdf$df.gender=="Male"], newdf$Yes[newdf$df.gender=="Male"], type = "p", col="blue", xlab = "religiosity", ylab = "Probability", xlim = c(1,5), ylim = c(0,1), main = "Males")  
  
legend("topleft", lty = c(1,1,1), pch = c(-1,-1,-1), col = c("blue", "green", "red"), legend = c("Yes","Undecided", "No"))  
  
points(newdf$df.religiosity[newdf$df.gender=="Male"], newdf$Undecided[newdf$df.gender=="Male"], col="green")  
  
points(newdf$df.religiosity[newdf$df.gender=="Male"], newdf$No[newdf$df.gender=="Male"], col="red")



plot(newdf$df.religiosity[newdf$df.gender=="Female"], newdf$Yes[newdf$df.gender=="Female"], type = "p", col="blue", xlab = "religiosity", ylab = "Probability", xlim = c(1,5), ylim = c(0,1), main = "Females")  
  
legend("topleft", lty = c(1,1,1), pch = c(-1,-1,-1), col = c("blue", "green", "red"), legend = c("Yes","Undecided", "No"))  
  
points(newdf$df.religiosity[newdf$df.gender=="Female"], newdf$Undecided[newdf$df.gender=="Female"], col="green")  
  
points(newdf$df.religiosity[newdf$df.gender=="Female"], newdf$No[newdf$df.gender=="Female"], col="red")

 Non reiligious males are less likely to believe than non religious females. At levels of high religiosity, both males and females are very likely to believe.

i.

#sample(newdf[newdf$df.religiosity == 3,])

Table

Description automatically generated with medium confidence

3.

1. 𝑙𝑜𝑔𝑖𝑡 (𝑃 𝑌 ≤ 1) = 𝛼1 + 𝛽1\*therapyseq +  𝛽2\*genderm

𝑙𝑜𝑔𝑖𝑡 (𝑃 𝑌 ≤ 2) = 𝛼2 + 𝛽1\*therapyseq +  𝛽2\*genderm

𝑙𝑜𝑔𝑖𝑡 (𝑃 𝑌 ≤ 3) = 𝛼3 + 𝛽1\*therapyseq +  𝛽2\*genderm

𝑙𝑜𝑔𝑖𝑡 (𝑃 𝑌 ≤ 1) = -2.3652 + 0.5587\*therapyseq +  0.5299 \*genderm

𝑙𝑜𝑔𝑖𝑡 (𝑃 𝑌 ≤ 2) = -1.3041 + 0.5587\*therapyseq +  0.5299 \*genderm

𝑙𝑜𝑔𝑖𝑡 (𝑃 𝑌 ≤ 3) = 0.1582 + 0.5587\*therapyseq +  0.5299 \*genderm

1. e^.5587 = 1.748

Using sequential treatment over alternating will have a 1.748 multiplicative effect on the severity outcome.

1. In this model, therapy type is a significant predictor in severity outcome, while gender is not. There is lower severity among those who use the alternating treatment method.
2. Graphical user interface, email

   Description automatically generated

4.

a.

Northeast = c(266,10,8,7)  
Midwest = c(15,414,22,6)  
South = c(61,50,578,27)  
West = c(28,40,22,301)  
  
df = data.frame(Northeast,Midwest,South,West, row.names = c("Northeast", "Midwest", "South", "West"))  
  
mat = data.matrix(df)  
mat

## Northeast Midwest South West  
## Northeast 266 15 61 28  
## Midwest 10 414 50 40  
## South 8 22 578 22  
## West 7 6 27 301

sixteen = c()  
for (i in 1:4){  
 sixteen[i] = (sum(mat[i,]))/(sum(mat))  
}  
  
twentyten = c()  
for (i in 1:4){  
 twentyten[i] = (sum(mat[,i]))/(sum(mat))  
}  
  
sixteen

## [1] 0.1994609 0.2770889 0.3396226 0.1838275

twentyten

## [1] 0.1568733 0.2463612 0.3859838 0.2107817

b.

#H0: 𝜋𝑖+ = 𝜋+𝑖  
#for all 𝑖 = 1, … ,4.  
#𝐻𝐴: For at least one pair, 𝜋𝑖+ ≠ 𝜋+𝑖  
  
library(coin)

## Warning: package 'coin' was built under R version 4.2.2

## Loading required package: survival

library(mvtnorm)  
library(modeltools)

## Loading required package: stats4

mh\_test(as.table(mat))

##   
## Asymptotic Marginal Homogeneity Test  
##   
## data: response by  
## conditions (Var1, Var2)   
## stratified by block  
## chi-squared = 86.236, df = 3, p-value < 2.2e-16

# p - value is less than .05, reject the null. The marginal probabilities are not all the same. That is, region  
# preferences have shifted from age 16 to 2010.

5.

a.

ten2 = read.csv("C:/Users/Nick/Downloads/tennis-18-20v2.csv")

library(BradleyTerry2)

## Warning: package 'BradleyTerry2' was built under R version 4.2.2

tennisModel <- BTm(outcome = cbind(win1,win2), as.factor(first.player), as.factor(second.player),  
formula = ~ player, id = "player", data=ten2, refcat = "Kenin")  
tennisModel

## Bradley Terry model fit by glm.fit   
##   
## Call: BTm(outcome = cbind(win1, win2), player1 = as.factor(first.player),   
## player2 = as.factor(second.player), formula = ~player, id = "player",   
## refcat = "Kenin", data = ten2)  
##   
## Coefficients:  
## playerBarty playerHalep playerOsaka playerSWilliams   
## 0.3771 1.2245 0.6105 0.9643   
##   
## Degrees of Freedom: 10 Total (i.e. Null); 6 Residual  
## Null Deviance: 6.565   
## Residual Deviance: 4.491 AIC: 23.84

cat("\nPlayers ranked:\n 1. Halep - 1.2245\n 2. Williams - .9643\n 3. Osaka - .6105\n 4. Barty - .3771\n 5. Kenin - 0(ref)")

##   
## Players ranked:  
## 1. Halep - 1.2245  
## 2. Williams - .9643  
## 3. Osaka - .6105  
## 4. Barty - .3771  
## 5. Kenin - 0(ref)

b.

# P(Halep beats Barty)  
  
p = (exp(1.2245-.3771))/(1+exp(1.2245-.3771))  
p

## [1] 0.7000214

cat("There is a 70% chance that Halep beats Barty")

## There is a 70% chance that Halep beats Barty

c.

# P(Kenin beats Williams)  
  
p = (exp(0-.9643))/(1+exp(0-.9643))  
p

## [1] 0.2760181

cat("There is a 27.6% chance that Kenin beats Williams")

## There is a 27.6% chance that Kenin beats Williams