|  |
| --- |
| > rm(list=ls())  > ###2  > ##d  >  > #Yes  >  > #Yes. There is an association between final grade and study amount.  > #For difference proportions, CI is [0.65946,0.92388] which is close to 1  > #suggesting that there is a strong association  >  > #For relative risk, CI is [7.09559,15.53782] which is quite larger than 1  > #suggesting that there is a strong association  >  > #For odds ratio, CI is [8.17106,215.88374] which is quite larger that 1  > #suggesting that there is a strong association  >  >  >  > ##e  >  >  > CI=matrix(0,nrow=100,ncol=2)  > t=0  > theta=(70\*12)/(2\*10)  > for(i in 1:100){  + data=runif(500,0,1)  + n11=0  + n12=0  + n21=0  + n22=0  + for(j in 1:500){  + if(data[j]<70/94){  + n11=n11+1  + } else if(data[j]<80/94){  + n12=n12+1  + } else if(data[j]<82/94){  + n21=n21+1  + } else{  + n22=n22+1  + }  + }  +  + theta\_hat=(n11\*n22)/(n12\*n21)  + CI[i,1]=exp(log(theta\_hat)-qnorm(1-(.05/2))\*(1/n11+1/n12+1/n21+1/n22)^0.5)  + CI[i,2]=exp(log(theta\_hat)+qnorm(1-(.05/2))\*(1/n11+1/n12+1/n21+1/n22)^0.5)  + if((theta>CI[i,1])&(theta<CI[i,2])){  + t=t+1  + }  + }  >  > t  [1] 96  >  > #change n = 100  >  > CI=matrix(0,nrow=100,ncol=2)  > t=0  > theta=(70\*12)/(2\*10)  > for(i in 1:100){  + data=runif(100,0,1)  + n11=0  + n12=0  + n21=0  + n22=0  + for(j in 1:100){  + if(data[j]<70/94){  + n11=n11+1  + } else if(data[j]<80/94){  + n12=n12+1  + } else if(data[j]<82/94){  + n21=n21+1  + } else{  + n22=n22+1  + }  + }  +  + theta\_hat=(n11\*n22)/(n12\*n21)  + CI[i,1]=exp(log(theta\_hat)-qnorm(1-(.05/2))\*(1/n11+1/n12+1/n21+1/n22)^0.5)  + CI[i,2]=exp(log(theta\_hat)+qnorm(1-(.05/2))\*(1/n11+1/n12+1/n21+1/n22)^0.5)  + if((theta>CI[i,1])&(theta<CI[i,2])){  + t=t+1  + }  + }  Error in if ((theta > CI[i, 1]) & (theta < CI[i, 2])) { :  missing value where TRUE/FALSE needed  >  > t  [1] 1  >  > ###4  > ##a  > rm(list=ls())  >  > x=matrix(c(192,75,8,459,586,471),nrow = 2,ncol = 3,byrow=TRUE)  > colnames(x)=c("PhD","Employed","Unemployed")  > rownames(x)=c("Yes","No")  > x  PhD Employed Unemployed  Yes 192 75 8  No 459 586 471  >  > pi\_hat=x/sum(x)  > pi\_hat  PhD Employed Unemployed  Yes 0.1072027 0.04187605 0.004466778  No 0.2562814 0.32719151 0.262981575  >  > x\_marg=apply(x,1,sum)  > y\_marg=apply(x,2,sum)  >  > mu\_hat=(x\_marg %\*% t(y\_marg))/sum(x)  > mu\_hat  PhD Employed Unemployed  [1,] 99.95812 101.4936 73.5483  [2,] 551.04188 559.5064 405.4517  >  > #X2 test  > X2=sum((x-mu\_hat)^2/mu\_hat)  > X2  [1] 177.3124  >  > qchisq(0.95,((dim(x)[1]-1)\*(dim(x)[2]-1)))  [1] 5.991465  >  > #177.3124>5.991465,reject H0,which means that they are not independent  >  > chi2=chisq.test(x)  > chi2  Pearson's Chi-squared test  data: x  X-squared = 177.31, df = 2, p-value < 2.2e-16  >  > #p-value is 2.2e-16, which means that assuming they are independent,  > #we have extremely small chance to get this sample of data.  >  > #G2 test  > G2=2\*sum(x\*log(x/mu\_hat))  > G2  [1] 197.3944  >  > qchisq(0.95,((dim(x)[1]-1)\*(dim(x)[2]-1)))  [1] 5.991465  >  >  >  > #197.3944>5.991465,reject H0,which means that they are not independent  >  > library(DescTools)  > LRT=GTest(x)  > LRT  Log likelihood ratio (G-test) test of independence without correction  data: x  G = 197.39, X-squared df = 2, p-value < 2.2e-16  >  > #p-value is 2.2e-16, which means that assuming they are independent,  > #we have extremely small chance to get this sample of data.  >  >  > ##b  >  > chisq.test(x)$stdres  PhD Employed Unemployed  Yes 12.54209 -3.598555 -9.70631  No -12.54209 3.598555 9.70631  >  > ##c  > #PhD and employed  > GTest(matrix(c(192,75,459,586),nrow=2,byrow=TRUE))  Log likelihood ratio (G-test) test of independence without correction  data: matrix(c(192, 75, 459, 586), nrow = 2, byrow = TRUE)  G = 68.448, X-squared df = 1, p-value < 2.2e-16  >  > #combined and unemployed  > GTest(matrix(c(192+75,8,459+586,471),nrow=2,byrow=TRUE))  Log likelihood ratio (G-test) test of independence without correction  data: matrix(c(192 + 75, 8, 459 + 586, 471), nrow = 2, byrow = TRUE)  G = 128.95, X-squared df = 1, p-value < 2.2e-16  >  > #Results shows that being PhD or employed is quite dependent to whether taking  > #6231 or not and being umemployed or not is also quite dependent to  > #whether taking 6231 or not.  >  >  > ###5  > ##a b  > rm(list=ls())  > data = as.data.frame(UCBAdmissions)  >  >  >  > G=matrix(0,nrow = 9,ncol = 3)  > rownames(G)=(c("AGD","AG,AD,GD","AD","GD","AG",  + "AD,GD","AG,GD","AG,AD","A,G,D"))  > colnames(G)=(c("G2","df","p-value"))  > #model1: Most general (AGD)  > #  >  > model1=glm(Freq~Admit\*Gender\*Dept,data=data,family = poisson())  > G[1,]=c(model1$deviance,model1$df.residual,  + "NULL")  >  >  > #model2: Homogeneous association (AG,AD,GD)  > #  >  > model2=glm(Freq~Admit+Gender+Dept+Admit\*Gender+Admit\*Dept+  + Gender\*Dept,data=data,family = poisson())  > G[2,]=c(model2$deviance,model2$df.residual,  + 1-pchisq(model2$deviance,model2$df.residual))  >  > #model3: Joint y-x,z (AD)  > #  > model3=glm(Freq~Admit+Gender+Dept+Admit\*Dept,data=data,family = poisson())  > G[3,]=c(model3$deviance,model3$df.residual,  + 1-pchisq(model3$deviance,model3$df.residual))  >  > #model4: Joint x-y,z (GD)  > #  > model4=glm(Freq~Admit+Gender+Dept+Gender\*Dept,data=data,family = poisson())  > G[4,]=c(model4$deviance,model4$df.residual,  + 1-pchisq(model4$deviance,model4$df.residual))  >  > #model5: Joint z-x,y (AG)  > #  > model5=glm(Freq~Admit+Gender+Dept+Admit\*Gender,data=data,family = poisson())  > G[5,]=c(model5$deviance,model5$df.residual,  + 1-pchisq(model5$deviance,model5$df.residual))  >  > #model6: Conditional x,y-z (AD,GD)  > #  > model6=glm(Freq~Admit+Gender+Dept+Admit\*Dept+  + Gender\*Dept,data=data,family = poisson())  > G[6,]=c(model6$deviance,model6$df.residual,  + 1-pchisq(model6$deviance,model6$df.residual))  >  > #model7: Conditional x,z-y (AG,GD)  > #  > model7=glm(Freq~Admit+Gender+Dept+Admit\*Gender+  + Gender\*Dept,data=data,family = poisson())  > G[7,]=c(model7$deviance,model7$df.residual,  + 1-pchisq(model7$deviance,model7$df.residual))  >  > #model8: Conditional y,z-x (AG,AD)  > #  > model8=glm(Freq~Admit+Gender+Dept+Admit\*Gender+Admit\*Dept  + ,data=data,family = poisson())  > G[8,]=c(model8$deviance,model8$df.residual,  + 1-pchisq(model8$deviance,model8$df.residual))  >  > #model9: Mutual Independent (A,G,D)  > #  > model9=glm(Freq~Admit+Gender+Dept,data=data,family = poisson())  > G[9,]=c(model9$deviance,model9$df.residual,  + 1-pchisq(model9$deviance,model9$df.residual))  > G  G2 df p-value  AGD "1.11910480882207e-13" "0" "NULL"  AG,AD,GD "20.2042753272416" "5" "0.00114407845139408"  AD "1242.35030597543" "11" "0"  GD "877.056413219775" "11" "0"  AG "2004.22180522134" "15" "0"  AD,GD "21.7355067781329" "6" "0.00135199265317087"  AG,GD "783.607006024044" "10" "0"  AG,AD "1148.9008987797" "10" "0"  A,G,D "2097.67121241707" "16" "0"  >  > ##c  > #I would choose saturated model(model1 (AGD)), because in the table  > #all p-value is less than 0.01, which means we reject every reduced  > #model.  >  > ##d  >  > GG=matrix(0,nrow = 3,ncol = 3)  > rownames(GG)=c("AD,GD|AG,AD,GD","AG,GD|AG,AD,GD","AG,AD|AG,AD,GD")  > colnames(GG)=(c("G2","df","p-value"))  >  > GG[1,]=c(model6$deviance-model2$deviance,  + model6$df.residual-model2$df.residual,  + 1-pchisq(model6$deviance-model2$deviance  + ,model6$df.residual-model2$df.residual))  >  > GG[2,]=c(model7$deviance-model2$deviance,  + model7$df.residual-model2$df.residual,  + 1-pchisq(model7$deviance-model2$deviance  + ,model7$df.residual-model2$df.residual))  >  > GG[3,]=c(model8$deviance-model2$deviance,  + model8$df.residual-model2$df.residual,  + 1-pchisq(model8$deviance-model2$deviance  + ,model8$df.residual-model2$df.residual))  >  > GG  G2 df p-value  AD,GD|AG,AD,GD 1.531231 1 0.2159277  AG,GD|AG,AD,GD 763.402731 5 0.0000000  AG,AD|AG,AD,GD 1128.696623 5 0.0000000 |
|  |
| |  | | --- | |  | |