TOPIC 5

GAMES

Name: Akash Mandal

Roll No. 16MA20007

Problem Statement

Draw the relevance contingency table to test the hypothesis “action video game is highly rated among teens”.

T=teens.(rating column in game data)

Professor's correction -

Professor has clarified that instead of action video game, we have to decide if video game Genre is independent of Rating.

Processing data

We check for missing values, and remove such rows from our data frame.

> #Remove rows with missing values of Rating or Genre

> table(is.na(GAMES$Genre))

FALSE TRUE

16717 2

> table(is.na(GAMES$Rating))

FALSE TRUE

9950 6769

>

> completeRows <- function(data, reqCols) {

+ completeVec <- complete.cases(data[, reqCols])

+ return(data[completeVec, ])

+ }

>

> newGames <- completeRows(GAMES, c("Rating","Genre"))

>

> table(is.na(newGames$Genre))

FALSE

9950

> table(is.na(newGames$Rating))

FALSE

9950

Chi - squared test-

H0 : rho = 0 (no correlation between genre and rating)

H1 : rho <> 0

We make a contingency table of Genre vs Rating for our data, and perform the chi square test. We obtain a chisquared value of 5255.7, and a p value < 2.2e-16. Thus we reject the null hypothesis and conclude that Genre must be related to Rating.

> tbl <- table(newGames$Genre, newGames$Rating)

> print(tbl)

AO E E10+ EC K-A M RP T

Action 1 416 481 1 0 608 0 681

Adventure 0 162 68 2 0 99 0 115

Fighting 0 8 19 0 0 49 0 362

Misc 0 457 167 5 1 13 0 239

Platform 0 358 144 0 0 3 0 64

Puzzle 0 289 43 0 0 0 0 10

Racing 0 585 96 0 0 18 1 172

Role-Playing 0 84 111 0 0 162 0 420

Shooter 0 48 58 0 0 565 0 348

Simulation 0 326 48 0 0 5 0 190

Sports 0 1188 107 0 0 16 0 198

Strategy 0 70 78 0 2 25 2 162

> chisq.test(tbl)

Pearson's Chi-squared test

data: tbl

X-squared = 5255.7, df = 77, p-value < 2.2e-16