---
title: "STAT2170-MERYN_JEROME_EMILIANUS_47083948"
output: pdf_document
date: '2022-05-19'

#a)

table(movie[, 1:2]) # This study is unbalanced as the number of observations is unequal. # The different levels of each factor combination have different size groups.

#b)

with(movie, interaction.plot(Gender, Genre, Score)) with(movie, interaction.plot(Genre, Gender, Score)) # By looking at the plots there is indication that there is some interaction between the factors and the response, as there are different slopes between the levels of each factor. # The interaction is not significant as the sample size is too low , but there is a slight change in the slope of the variable lines.

boxplot(Score ~ as.factor(Gender) + as.factor(Genre), data = movie) # box plot shows constant variability between the levels of each factor.

#c)

Mathematical Model movie.lm = lm(Score ~ as.factor(Gender) + as.factor(Genre), data = movie) anova(movie.lm) summary(movie.lm)$coefficients

$$Yi = \mu + \alpha i + \beta j + \gamma ij + \epsilon i$$

Yi = score response

μ = overall mean

α = Gender effect for i = 1,2

β = Genre effect for j = 1,2,3

γij = interaction between Gender and Genre

#d)

Test: Effect of Gender on Score

Null hypothesis - H0: There is interaction between Gender and Score

Alternative Hypothesis - H1: There is no interaction between Gender and Score

movie.1 = lm(Score ~ as.factor(Gender), data = movie) anova(movie.1) # In this test we reject the null hypothesis # Therefore there is no interaction between Gender and Score

plot(movie.1, which = 1:2) # The diagnostic plots validates the model. # On the residual vs fitted plot each cell has a similar variance. # Quantile plot of residuals shows a close linear relationship, suggesting they are normally distributed.

Test: Effect of Genre on Score

Null hypothesis - H0: There is interaction between Genre and Score

Alternative Hypothesis - H1: There is no interaction between Genre and Score

movie.2 = lm(Score ~ as.factor(Genre), data = movie) anova(movie.2) # In this test we reject the null hypothesis # Therefore, there is no interaction between Genre and Score

plot(movie.2, which = 1:2) # The diagnostic plots validates the model. # On the residual vs fitted plot each cell has again a similar variance. # Quantile plot of residuals shows a close linear relationship, suggesting they are normally distributed.