**INDIVIDUAL ASSIGNMENT-1**

1. A)

Mod.Q1A = lm (Rating ~ Age +Income, data = Courtyard.df)

summary (Mod.Q1A)

**OUTPUT :**

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 63.5304 19.9096 3.191 0.00535 \*\*

Age 0.2501 0.5627 0.444 0.66236

Income -0.2470 0.1957 -1.262 0.22390

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 19.42 on 17 degrees of freedom

Multiple R-squared: 0.08967, Adjusted R-squared: -0.01742

F-statistic: 0.8373 on 2 and 17 DF, p-value: 0.45

As the regression model has a p-value greater than 0.05 we can say that the regression model has a good explanatory power.

**Scatter Plots for Ratings against each input variable:**

#SCATTER PLOT OF RATINGS VS AGE

plot(Rating ~ Age, data=Courtyard.df,

pch= 22, lwd=0.5 , bg= "Blue",

main = "Scatter plot of Ratings vs Age",

xlab= "Ratings on a scale of 1-100", ylab = "Age in years")

#SCATTER PLOT OF RATINGS VS INCOME

plot(Rating ~ Income , data=Courtyard.df,

pch= 22, lwd=0.5 , bg= "Green",

main = "Scatter plot of Ratings vs Age",

xlab= "Ratings on a scale of 1-100", ylab = "Income in $ 1000s")

#SCATTER PLOT FOR RESIDUALS

#PLOT 1 : RESIDUALS VS AGE

plot (y= Mod.Q1A$residuals, x= Courtyard.df$Age,

pch= 22, lwd=0.5 , bg= "Red",

main = "Scatter plot of Residuals",

xlab= "Age in years", ylab = "Residuals")

abline (h=0, lty = 2)

#PLOT 2 : RESIDUALS VS INCOME

plot (y= Mod.Q1A$residuals, x= Courtyard.df$Income,

pch= 22, lwd=0.5 , bg= "Violet",

main = "Scatter plot of Residuals",

xlab= "Income in $1000s", ylab = "Residuals")

abline (h=0, lty = 2)









#REGRESSION MODEL WITH JUST AGE AS INPUT

Mod.Q1a = lm (Rating ~ Age, data = Courtyard.df)

summary(Mod.Q1a)

**OUTPUT:**

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 56.7768 19.4907 2.913 0.00928 \*\*

Age -0.1349 0.4806 -0.281 0.78223

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 19.74 on 18 degrees of freedom

Multiple R-squared: 0.004355, Adjusted R-squared: -0.05096

F-statistic: 0.07873 on 1 and 18 DF, p-value: 0.7822

#REGRESSION MODEL WITH JUST INCOME AS INPUT

Mod.Q1a = lm (Rating ~ Income, data = Courtyard.df)

summary(Mod.Q1a)

**OUTPUT:**

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 69.2181 14.9071 4.643 0.000202 \*\*\*

Income -0.1999 0.1607 -1.243 0.229668

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 18.98 on 18 degrees of freedom

Multiple R-squared: 0.0791, Adjusted R-squared: 0.02794

F-statistic: 1.546 on 1 and 18 DF, p-value: 0.2297

#REGRESSION MODEL WITH JUST LOG(INCOME) AS INPUT

Mod.Q1a = lm (Rating ~ log(Income), data = Courtyard.df)

summary(Mod.Q1a)

**OUTPUT:**

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 99.60 65.24 1.527 0.144

log(Income) -10.84 14.65 -0.740 0.469

Residual standard error: 19.49 on 18 degrees of freedom

Multiple R-squared: 0.02951, Adjusted R-squared: -0.02441

F-statistic: 0.5473 on 1 and 18 DF, p-value: 0.469

When we take the correlation of the two input variables, we see that its 0.54 which shows a very high correlation between the input variables, thus we can eliminate the income factor, as both age and income give us the same information about ratings. But because the income and the log (income) give a non-linear plot, a simpler regression model would eliminate one of the input variables and keeping age which is capable of representing the same relationship with rating.

A regression model of rating and age has a higher p-value as well. The p-value of this regression model is 0.78 which is much higher than the previous regression model that takes both the input variables. It is closer to 1 and has a better explanatory power.

I propose a regression model of Rating as a function of Age.

1)B)

#REGRESSION MODELS :

Mod.Q1B = lm (Rating ~ Minc + Hinc, data = Courtyard.df)

summary(Mod.Q1B)

**OUTPUT:**

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 45.333 5.515 8.219 2.52e-07 \*\*\*

Minc 22.667 7.296 3.107 0.00641 \*\*

Hinc -9.833 7.800 -1.261 0.22445

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 13.51 on 17 degrees of freedom

Multiple R-squared: 0.5596, Adjusted R-squared: 0.5078

F-statistic: 10.8 on 2 and 17 DF, p-value: 0.0009397

We can see from the p-value that the regression result is consistent with the scatterplot of Rating vs Income that we did in the previous question.

The coefficient of the high-income dummy is significantly different from zero. It is -9.833 which tells us that for every increase in 1000$ rating will decrease by 9.833. Thus, it is non-zero relationship.

1)C)

Mod.Q1b = lm (Rating ~ Hlvl + Mlvl, data = Courtyard.df)

summary(Mod.Q1b)

**OUTPUT:**

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 42.400 5.346 7.932 4.1e-07 \*\*\*

Hlvl -5.543 6.999 -0.792 0.439306

Mlvl 27.475 6.814 4.032 0.000865 \*\*\*

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 11.95 on 17 degrees of freedom

Multiple R-squared: 0.6552, Adjusted R-squared: 0.6147

F-statistic: 16.15 on 2 and 17 DF, p-value: 0.0001172

The slope coefficients of Mid-Level Managers are more significant than the High-Level ones.

The R- square values in 1)B) are 55.9% and 50.78% while for this question they are 65.52% and 61.47% which means that the regression model of occupation and rating is a better fit for the data than income and rating.

1)D)

Mod.Q1D = lm (Rating ~ (Hlvl + Mlvl) + (Minc + Hinc), data = Courtyard.df)

summary(Mod.Q1D)

**OUTPUT:**

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 42.400 5.465 7.758 1.25e-06 \*\*\*

Hlvl -8.686 18.705 -0.464 0.649

Mlvl 17.600 13.387 1.315 0.208

Minc 11.286 13.064 0.864 0.401

Hinc 1.786 18.572 0.096 0.925

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 12.22 on 15 degrees of freedom

Multiple R-squared: 0.682, Adjusted R-squared: 0.5972

F-statistic: 8.043 on 4 and 15 DF, p-value: 0.001133

Mid-level income and Mid-level managers have significant slope coefficients.

We can see that while the value of the R-square has gone up marginally that tells us that the explain ability of the model has increased marginally by approximately 2% but the Adjusted R-square value which depicts the quality has decreased. Thus we can conclude that for such a marginal increase in explainabilty, compromising on the quality is not the kind of tradeoff we would prefer for our model, thus we prefer a regression model with a single input as occupation as it is the best fit among the three cases for our data.

2)A) Part Worth for Companies



**Output:**

Xerox HP Cannon Epson

0.1097635 0.5773134 -0.2890619 -0.3980150

Part Worth for Price



**OUTPUT:**

Price149 Price199 Price249 Price349

0.4994264 0.1482798 -0.1471325 -0.5005736

2)B)

2)C) Based on the share of preference method, the following are the market shares of P1,P2 and P3.

**OUTPUT :**

P1 P2 P3

0.4461008 0.3770772 0.1768220

3)

Market Share of P1 = 0.1640650

Market Share of P2 = 1694678

Market Share of P3 = 1462418

Thus, we can see that P2 has the largest market share and thus it should be launched.