**Q1)**

For this problem, load the College dataset from the ISLR package

**library**("ISLR")

**data**("College")

Please estimate a linear regression model (using the lm function) with Personal as the dependent variable and Room.Board as the independent variable. What are the model’s R-squared and adjusted R-squared values respectively?

1. 0.00549, 0.048
2. 0.0143, 0.022
3. 0.0398, 0.0385
4. 0.0325, 0.0336

**Answer: C (Week 1 Lesson 4)**

**library**("ISLR")

**data**("College")

**summary**(**lm**(College**$**Personal**~**College**$**Room.Board))

##

## Call:

## lm(formula = College$Personal ~ College$Room.Board)

##

## Residuals:

## Min 1Q Median 3Q Max

## -1153.1 -444.6 -92.3 316.0 5505.2

##

## Coefficients:

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

## (Intercept) 1877.14827 97.64374 19.224 < 2e-16 \*\*\*

## College$Room.Board -0.12312 0.02173 -5.666 2.06e-08 \*\*\*

## ---

## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

##

## Residual standard error: 663.9 on 775 degrees of freedom

## Multiple R-squared: 0.03977, Adjusted R-squared: 0.03853

## F-statistic: 32.1 on 1 and 775 DF, p-value: 2.065e-08

**Q2)**Based on the linear-linear regression model in the previous question (with Personal as the dependent variable and Room.Board as the independent variable), fit three nonlinear models using those two variables. Based on their adjusted R-squared values, which one of the four models is most appropriate to use?

1. Log-Linear
2. Log-Log
3. Linear-Linear (level-level)
4. Linear-Log

**Answer: B (Week 3 lesson 4)**

**summary**(**lm**(**log**(College**$**Personal)**~**College**$**Room.Board))

## Call:

## lm(formula = log(College$Personal) ~ College$Room.Board)

##

## Residuals:

## Min 1Q Median 3Q Max

## -1.61024 -0.31235 0.03383 0.31037 1.77383

##

## Coefficients:

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

## (Intercept) 7.485e+00 6.992e-02 107.057 < 2e-16 \*\*\*

## College$Room.Board -9.187e-05 1.556e-05 -5.904 5.3e-09 \*\*\*

## ---

## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

##

## Residual standard error: 0.4754 on 775 degrees of freedom

## Multiple R-squared: 0.04304, Adjusted R-squared: 0.04181

## F-statistic: 34.86 on 1 and 775 DF, p-value: 5.303e-09

**summary**(**lm**(**log**(College**$**Personal)**~log**(College**$**Room.Board)))

##

## Call:

## lm(formula = log(College$Personal) ~ log(College$Room.Board))

##

## Residuals:

## Min 1Q Median 3Q Max

## -1.60098 -0.31047 0.03916 0.30663 1.78574

##

## Coefficients:

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

## (Intercept) 10.47164 0.56140 18.653 < 2e-16 \*\*\*

## log(College$Room.Board) -0.40568 0.06722 -6.035 2.46e-09 \*\*\*

## ---

## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

##

## Residual standard error: 0.4749 on 775 degrees of freedom

## Multiple R-squared: 0.04489, Adjusted R-squared: 0.04366

## F-statistic: 36.42 on 1 and 775 DF, p-value: 2.46e-09

**summary**(**lm**(College**$**Personal**~**College**$**Room.Board))

##

## Call:

## lm(formula = College$Personal ~ College$Room.Board)

##

## Residuals:

## Min 1Q Median 3Q Max

## -1153.1 -444.6 -92.3 316.0 5505.2

##

## Coefficients:

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

## (Intercept) 1877.14827 97.64374 19.224 < 2e-16 \*\*\*

## College$Room.Board -0.12312 0.02173 -5.666 2.06e-08 \*\*\*

## ---

## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

##

## Residual standard error: 663.9 on 775 degrees of freedom

## Multiple R-squared: 0.03977, Adjusted R-squared: 0.03853

## F-statistic: 32.1 on 1 and 775 DF, p-value: 2.065e-08

**summary**(**lm**(College**$**Personal**~log**(College**$**Room.Board)))

##

## Call:

## lm(formula = College$Personal ~ log(College$Room.Board))

##

## Residuals:

## Min 1Q Median 3Q Max

## -1165.3 -442.5 -98.8 296.5 5520.4

##

## Coefficients:

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

## (Intercept) 5818.18 784.51 7.416 3.16e-13 \*\*\*

## log(College$Room.Board) -536.36 93.93 -5.710 1.61e-08 \*\*\*

## ---

## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

##

## Residual standard error: 663.7 on 775 degrees of freedom

## Multiple R-squared: 0.04037, Adjusted R-squared: 0.03913

## F-statistic: 32.61 on 1 and 775 DF, p-value: 1.609e-08

**Q3)** Interpret the coefficient of the independent variable for the Linear-Log model.

1. 1% increase in Room.Board leads to approximately 536.36 units decrease in Personal
2. 1 unit increase in Room.Board leads to approximately 536.36 units decrease in Personal
3. 1 unit increase in Room.Board leads to approximately 0.01\*536.36 units decrease in Personal
4. 1% increase in Room.Board leads to approximately 0.01\*536.36 units decrease in Personal

**Answer: D** Explanation: One percent increase in X increases log(X) by 0.01 and therefore changes the Y variable by 0.01 \* b1. **(week 3 lesson 2)**

**summary**(**lm**(College**$**Personal**~log**(College**$**Room.Board)))

##

## Call:

## lm(formula = College$Personal ~ log(College$Room.Board))

##

## Residuals:

## Min 1Q Median 3Q Max

## -1165.3 -442.5 -98.8 296.5 5520.4

##

## Coefficients:

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

## (Intercept) 5818.18 784.51 7.416 3.16e-13 \*\*\*

## log(College$Room.Board) -536.36 93.93 -5.710 1.61e-08 \*\*\*

## ---

## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

##

## Residual standard error: 663.7 on 775 degrees of freedom

## Multiple R-squared: 0.04037, Adjusted R-squared: 0.03913

## F-statistic: 32.61 on 1 and 775 DF, p-value: 1.609e-08

*#One percent increase in X increases log(X) by 0.01 and therefore changes the Y variable by 0.01 \* b1*

**Q4)**Interpret the coefficient of the independent variable for the Log-Linear model. (E-05=10^-5)

1. 1% increase in Room.Board leads to e^(9.187E-05) units decrease in Personal
2. 1 unit increase in Room.Board leads to ((e^(9.187E-05)-1) \* 100)% decrease in Personal
3. 1 unit increase in Room.Board leads to e^(9.187E-05) units decrease in personal
4. 1% increase in Room.Board leads to (e^(9.187E-05) \* 100)% decrease in Personal

**Answer: B** Explanation: The accurate percentage change in Y is (e^b1 - 1)\*100 for a one unit change in X. **(Week 3 Lesson 4)**

**summary**(**lm**(**log**(College**$**Personal)**~**College**$**Room.Board))

##

## Call:

## lm(formula = log(College$Personal) ~ College$Room.Board)

##

## Residuals:

## Min 1Q Median 3Q Max

## -1.61024 -0.31235 0.03383 0.31037 1.77383

##

## Coefficients:

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

## (Intercept) 7.485e+00 6.992e-02 107.057 < 2e-16 \*\*\*

## College$Room.Board -9.187e-05 1.556e-05 -5.904 5.3e-09 \*\*\*

## ---

## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

##

## Residual standard error: 0.4754 on 775 degrees of freedom

## Multiple R-squared: 0.04304, Adjusted R-squared: 0.04181

## F-statistic: 34.86 on 1 and 775 DF, p-value: 5.303e-09

*#The accurate percentage change in Y is (e^b1 - 1)\*100 for a one unit change in X*

**Q5)** Interpret the coefficient of the independent variable for the Log-Log model.

1. 1% increase in Room.Board leads to e^0.0040568 increase in Personal
2. 1 unit increase in Room.Board leads to 0.40568\*100% decrease in Personal
3. 1 unit increase in Room.Board leads to (e^0.40568)\*100% decrease in Personal
4. 1% increase in Room.Board leads to 0.40568% ~~increase~~ decrease in Personal

**Answer: ~~A~~ D** Explanation**: Increasing log(X) by 0.01 leads to increasing log(Y) by e^0.01b1 - 1, which implies increasing X by 1% changes Y by (e^0.01b1 - 1)\*100% (Week 4 Lesson 4)**

**summary**(**lm**(**log**(College**$**Personal)**~log**(College**$**Room.Board)))

##

## Call:

## lm(formula = log(College$Personal) ~ log(College$Room.Board))

##

## Residuals:

## Min 1Q Median 3Q Max

## -1.60098 -0.31047 0.03916 0.30663 1.78574

##

## Coefficients:

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

## (Intercept) 10.47164 0.56140 18.653 < 2e-16 \*\*\*

## log(College$Room.Board) -0.40568 0.06722 -6.035 2.46e-09 \*\*\*

## ---

## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

##

## Residual standard error: 0.4749 on 775 degrees of freedom

## Multiple R-squared: 0.04489, Adjusted R-squared: 0.04366

## F-statistic: 36.42 on 1 and 775 DF, p-value: 2.46e-09

*#Increasing log(X) by 0.01 leads to increasing log(Y) by b1 \* 0.01 units which implies increasing X by 1% changes Y by b1 %*

**Instructions for Q6 – Q10**

Imagine you are interested in knowing how variables like GRE (Graduate Record Exam scores), GPA (Grade Point Average) etc affect admission into graduate school. The response variable, **"admit”** (admit/don’t admit), is a binary variable (admit = 1 / don't admit = 0). Create a logistic regression model using the dataset **binary.csv**. Use the information from the model to answer the following five questions. Select the closest answer.

You might want to use this code to read the file:  
read.csv(... , fileEncoding="UTF-8-BOM",..)

**Q6)** How to interpret the coefficient of gre?

1. If gre increases by 1 unit, the natural log of the odds of admission increases by 0.003.
2. If gre increases by 1 unit, the odds of admission increase by a factor of exp(0.003).
3. If gre increases by 1 unit, the odds of admission increase by roughly 100\*0.003 percent.
4. All of the above.

**Answer: D** The coefficient of gre is 0.003, hence when the value of gre changes by 1, the log(odds) change by 0.003, and so the off change by exp(0.003), and since 0.003 is very small, this can be approximated to 0.003\*100%. Hence the answer is **(D) All of the above. (Week 4 Lesson 3. Page 11, Slide 2)**

# Load dataset

mydata <- read.csv("binary.csv")

# Create Logistic Model

mylogit <- glm(admit ~ gre + gpa, data = mydata, family = "binomial")

# Model summary

summary(mylogit)



**Q7)** How to interpret the coefficient of gpa?  
A. If gpa increases by 1 unit, the natural log of the odds of admission increases by 0.755.  
B. If gpa increases by 1 unit, the odds of admission increase by 0.755.  
C. If gpa increases by 1 unit, the odds of admission increase by 10^(0.755).  
D. All of the above.

**Answer: A** The coefficient of gpa is 0.755, hence when the value of gpa changes by 1, the log(odds) change by 0.755. Answer is **(A) (Week 4 Lesson 3 page 11)**

**Q8)**

A student has the GPA of 3.5 and GRE score of 330. What is the predicted probability of this student getting admitted into graduate school?

1. exp(-4.949 + 0.003\*3.5 + 0.755\*330)/[1 + exp(-4.949 + 0.003\*3.5 + 0.755\*330)]
2. exp(-4.949 + 0.003\*330 + 0.755\*3.5)/[1 + exp(-4.949 + 0.003\*330 + 0.755\*3.5)]
3. [1 - exp(-4.949 + 0.003\*330 + 0.755\*3.5)]/[1 + exp(-4.949 + 0.003\*330 + 0.755\*3.5)]
4. [1 - exp(-4.949 + 0.003\*330 + 0.755\*3.5)]/exp(-4.949 + 0.003\*330 + 0.755\*3.5)

**Answer: B. (Week 4 Lesson 3. Page 10 S1)**

Probability of being admitted is given by



**Q9)**

If a student has a GRE score of 330, with 0.1 unit increase in GPA, what is the change of the natural log of predicted odds of this student getting admitted into graduate school?

1. exp(-4.949 + 0.003\*0.2 + 0.755\*330)/[1 + exp(-4.949 + 0.003\*0.2 + 0.755\*330)]
2. exp(-4.949 + 0.003\*0.2 + 0.755\*330)
3. 0.1509
4. None of the above

**Answer: D**

**(Week 4 Lesson 3) 0.754687\*0.1 = 0.0754687**

**Q10)**

What is the value of the area under the curve (AUC) for the model created? To calculate the AUC, use a threshold of 0.5. Please select the closest answer.

1. 0.804
2. 0.935
3. 0.635
4. 0.832

**Answer: C (week 4 Lesson 5,6, page 29 of slides)**

**Code:** The code is



Hence are under the curve is **(C) 0.635**

**Instructions for Q11 to Q16**

Use the dataset **Berkshire.csv** with the following variables. All returns reported as a decimal (as opposed to a percentage).

**Sample Period: From** 30th Nov 1976 **to** 31st Dec 2005 (Inclusive)

* Column (1):  *Date*, Calendar Date
* Column (2):  *BRKret*, Berkshire Hathaway’s monthly return
* Column (3):  *MKT*, the return on the aggregate stock market
* Column (4):  *RF*, the risk free rate of return

**Q11)**

What is the standard deviation for Berkshire Hathaway’s returns over the sample period?

Note: Remember to filter the data for the sample period, before solving the problem.

1. 7.55%
2. 7.61%
3. 7.42%
4. 7.37%

**Ans: C (code given below) (Week 6 Lesson 2)**

**Q12)**

What is Berkshire Hathaway’s arithmetic average return over the sample period? (Select the closest answer)

Note: Remember to filter the data for the sample period, before solving the problem.

1. 1.5%
2. 1.9%
3. 2.3%
4. 2.7%

**Ans: C ( Week 6 Lesson 1)**

data <- read.csv("Berkshire.csv")

sd1 <-sd(data$BRKRet, na.rm = TRUE) #0.07422054

mean<- mean(data$BRKRet, na.rm = TRUE) #0.02333686

**Q13)** Over the sample period, relative to the aggregate market, Berkshire Hathaway has:

Note: Remember to filter the data for the sample period, before solving the problem.

1. Underperformed the market
2. Underperformed the market by 0.5% to 1.50% per month on average
3. Outperformed the market by 0.5% to 1.0% per month on average
4. Outperformed the market by greater than 1.00% per month on average

**Answer: D (week 6 Lesson 3)**

**df <- read.csv("Berkshire.csv", header = TRUE)**

**df$ExcessPerformance<- df$BrkRet-df$MKT**

**mean(df$ExcessPerformance) #**0.01230057

Therefore, Outperformed the market by 1.23% (> 1.00%) on average.

**Q14)**

What is the compound return of a $10,000 investment in Berkshire Hathaway at the end of the sample period?

Note: Remember to filter the data for the sample period, before solving the problem.

a. $977,214

b. $12,207,145

c. $13,013,642

d. Over $20,000,000

**Answer: C (Week 6 Lesson 1)**

**(prod(df$BrkRet+1)-1)\*10000 #** 13,013,642

**Q15)** What is Berkshire Hathaway’s monthly Sharpe ratio for the sample period? (Please indicate the std Sharpe ratio)

Note: Remember to filter the data for the sample period, before solving the problem.

a. 0.432

b. 0.324

c. 0.179

d. 0.248

**Answer: D (Week 7 Lesson 1)**

**row.names(df)<-as.Date(df$Date, format = c("%m/%d/%Y"))**

**SharpeRatio(df[,2,drop=FALSE], Rf = df[,4,drop=FALSE])**

or  
data\_edited = xts(Berkshire[,-1], order.by = as.POSIXct(Berkshire$Date))  
SharpeRatio(data\_edited$BrkRet, data\_edited$RF) **#**0.2479675

**#** Note that we are using std Sharpe ratio

**Q16)**

Berkshire Hathaway’s Sharpe Ratio is \_\_\_\_\_\_\_\_ than the aggregate stock market for the sample period? (Please use the std Sharpe ratios)

Note: Remember to filter the data for the sample period, before solving the problem.

a. Higher

b. Lower

**Answer: A (Week 7 Lesson 1)**

**library(PerformanceAnalytics)**

**SharpeRatio(df[,3,drop=FALSE], Rf = df[,4,drop=FALSE])**

Or SharpeRatio(data\_edited$MKT, data\_edited$RF)  #0.13713315

Market sharpe ratio is 0.13713315 while Berkshire Hathaway’s Sharpe Ratio is 0.2479675

**\*\*Instructions for Q17 to Q18**

In this question, we will determine the factors explaining the returns for

the HiTec industry portfolio. We will build a factor regression model using

the data in the Factor\_HiTec.csv file to answer the questions below.

In the file,

· Mkt\_rf: Monthly excess return on the aggregate stock market

·SMB: Size Factor

·HML: Value Factor

·QMJ: Quality Factor

·BAB: Betting against beta factor

·Mom: Momentum factor

· HiTec\_rf: Monthly excess return on the HiTec industry portfolio

**Q17)** Which factors have the highest positive and highest negative exposure on the portfolio respectively?

A. Mom and SMB

B. HML and Mkt\_rf

C. SMB and BAB

D. Mkt\_rf and HML

E. BAB and QMJ

**Solution: D (Week 8 Lesson 2)**

**dat = read.csv("Factor\_HiTec.csv")**

**linearMod <-lm(HiTec\_rf~dat$Mkt\_rf+dat$SMB+dat$HML+dat$QMJ+dat$BAB+dat$Mom, data=dat)**

**print(linearMod)**

**Q18)** Given a significance level of 0.001, which factor could be removed from this model if we have to limit the number of features less than 6?

A. QMJ

B. SMB

C. HML

D. BAB

**Solution: A (Week 8 Lesson 1)**

Because it is not statistically significant.

Summary(linearMod)

**Explanation for Q17 and Q18:**

dat = **read.csv**("Factor\_HiTec.csv")  
linearMod <-**lm**(HiTec\_rf**~** Mkt\_rf **+** SMB **+** HML **+** QMJ **+** BAB **+** Mom, data=dat)   
summary(linearMod)

##   
## Call:  
## lm(formula = HiTec\_rf ~ Mkt\_rf + SMB + HML + QMJ + BAB + Mom,   
## data = dat)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.073606 -0.015719 -0.001154 0.015072 0.085923   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 0.004567 0.001362 3.354 0.000863 \*\*\*  
## Mkt\_rf 1.113453 0.035396 31.457 < 2e-16 \*\*\*  
## SMB 0.222049 0.048208 4.606 5.33e-06 \*\*\*  
## HML -0.543901 0.051113 -10.641 < 2e-16 \*\*\*  
## QMJ -0.038604 0.073603 -0.524 0.600193   
## BAB -0.286339 0.040218 -7.120 4.25e-12 \*\*\*  
## Mom -0.093566 0.030698 -3.048 0.002438 \*\*   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.02614 on 455 degrees of freedom  
## Multiple R-squared: 0.8507, Adjusted R-squared: 0.8487   
## F-statistic: 432.1 on 6 and 455 DF, p-value: < 2.2e-16

**Instructions for Q19 & Q20**

Use the data set UPS\_KO.csv to answer the following questions:

Date: This column represents date from 09/2014 to 08/2019.

Mkt\_RF: This column represents market premium (i.e., Market return –

risk\_free rate).

SMB: This column represents the value of the size factor.

HML: This column represents the value of the value factor.

RF: This column represents risk free rate.

UPS: This column represents the return of UPS.

KO: This column represents the return of KO.

**Sample Period : 04/2015 to 11/2018 (Inclusive)**

Estimate a three-factor model by regressing return in excess of the risk free rate on Mkt\_rf; SMB; and HML for both UPS and KO for the sample period.

**Q19)** The coefficient of HML for the three factor model for UPS suggests that ~~is most~~:

Note: Remember to filter the data for the sample period, before solving the problem.

A. UPS is tilted towards small cap stocks

B. UPS is tilted towards large cap stocks

C. UPS is tilted towards value stocks

D. UPS is tilted towards growth stocks

**Answer: C (Week 8 Lesson 2 Slide 11)**

**Q20)** Based on the three factor model, which firm has a higher level of performance? What is this firm's return (performance level)?

Note: Remember to filter the data for the sample period, before solving the problem.

A. UPS, 0.06% per month

B. UPS, 0.6% per month

C. KO, 0.3 % per month

D. KO, 0.3% per year

**Answer: C (Week 8 Lesson 2)**

**Explanation for Q19 and Q20:**









