**MIDTERM – SOLUTION KEY**

**THEORY QNS**

**Q1)**

False negative is a Type II error, i.e., you incorrectly retain a false null hypothesis.

1. True
2. False

**Answer: A (True)**

**Explanation:**

A picture containing indoor, table, person

Description automatically generated

**Q2)**

Adding interaction terms:

A. Decreases model complexity

B. Increases model complexity

C. Always makes the independent variable insignificant

D. Always increases the Goodness of Fit of the model

**Answer: B**

**Explanation: Adding more interaction terms increases the complexity of interpreting the coefficient there by increasing the model complexity.**

**Q3)**

In a linear regression model, you add a categorical variable of City which takes values of 60 different cities. This leads to:

(i) Overfitting of your model

(ii) Underfitting of your model

(iii) Reduction in the Degrees of freedom of your model

A. (i) only

B. (ii) only

C. (iii) only

D. (i) and (ii)

E. (i) and (iii)

**Answer: E & C**

**Explanation: degrees of freedom = # of samples - # of parameters. As the parameters increase, degrees of freedom decreases. Model overfitting is a possibility if the data set is small. So, awarding credit for both E & C options.**

**Q4)**

In the model log(Y) = b0 + b1\*log(X), the elasticity of Y is the percentage change in Y (the dependent variable), when X (the independent variable) increases by one unit.

1. False
2. True

**Answer: False (A)**

**Explanation: Elasticity is the percent change in Y when X increases by 1%**

**Q5)**

Odds for is the probability that the event will not happen divided by the probability that the event will happen.

1. True
2. False

**Answer: B (False)**

**Explanation: Odds for is the probability that the event will happen divided by probability that the event will not happen.**

**Q6)**

In the election of Whoville, a poll is conducted to see whether Alex or Bob will win. Accidentally, this poll is conducted in an area which has a vast majority of Bob's supporters. Based on the result, it looks like Bob is going to win the election. However, the poll's prediction was incorrect and Alex ends up winning the election. This is an instance of selection bias – true or false?

A. True

B. False

**Answer: True**

**Explanation: Since everyone answered the poll, and we made it clear that most of them were Bob’s supporters, it is a type of selection bias. It cannot be non response bias because Alex's supporters did participate in the poll but since there are more of Bob’s supporters, this bias does not exist.**

**Instructions for Q7 and Q8**

Load dataset “Wages” (not “Wages1”) from Package “Ecdat” in R. Run the following linear regression model:  
                     **lwage = b0 + b1\*ifsouth**  
where “ifsouth” is an indicator variable created from variable “south” and 1 indicates people who live in the south). More details on variable description can be found from CRAN <http://ftp.auckland.ac.nz/software/CRAN/doc/packages/Ecdat.pdf>

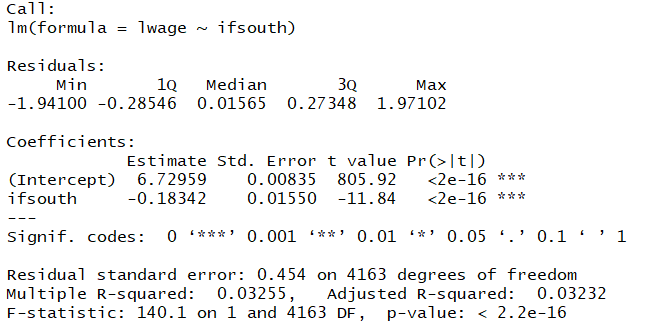
**Q7)**

What is the average of “lwage” for an individual who resides in the south?

* 1. 6.73
  2. 6.70
  3. 6.60
  4. 6.55

**Answer: D**

**Explanation: 6.73-0.18=6.55. With ifsouth=1, b0=6.73, b1=-0.18**



**Q8)**

What is the value of difference estimator, b1, in the model?

* 1. 0.01565
  2. 6.72959
  3. -0.18342
  4. 0.00835

**Answer: C**

**Explanation: From the regression output, b1=-0.18342**

**Q9)**

Beta value of Company A = 0.5, Company B = 1 and Company C = 1.5. Which company is expected to move **exactly along** with the market?

1. Company A
2. Company B
3. Company C
4. Company A & Company C

**Answer: B**

**Explanation: Beta Value of 1 means same variance as market (Lesson 2)**

**Q10)**

Which of these asset classes has historically been the safest (least risky)?

1. Small Cap Stocks
2. Large Cap Stocks
3. Corporate Bonds
4. Treasury Bonds

**Answer: D**

**Explanation: Treasury Bonds have historically been the safest asset (Lesson 3)**

**Q11)**

Fill in the blanks

On average, riskier investments have \_\_\_\_\_\_\_\_\_ average annual return & \_\_\_\_\_\_\_\_\_\_ standard deviation.

1. Low, low
2. Low, High
3. High, Low
4. High, High

**Answer: D (Lesson 3 statement), the question asks for on average or in general. Exceptions like BAB aren't considered. The question asks a direct concept from lesson 3 in week 6.   
The BAB strategy includes finding overvalued assets which have high beta and and shorting them and finding undervalued assets with low beta and taking a leveraged long position. This is a specific strategy used by some firms and the graph shown is only for a few assets. The question asks about the general market and any asset in general.**

**Q12)**

Bob deposited $100 in the bank. After 3 years, the savings compounded to $115.76. What was the annual rate of interest offered by the bank?

1. 3%
2. 4%
3. 5%
4. 6%

**Answer: C**

**Explanation: 115.76=100(1+ x)^3. Solving for x, we get 5%**

**Q13)**

A company of market value $10 billion has a stock split of 2 for 1. Each share is valued at $100 before the stock split. What is the value of each share after the stock split?

1. $50
2. $100
3. $200

**Answer: A**

**Explanation: After 2:1 split. The prices are divided by 2 and number of shares is multiplied by 2. Therefore 100/2 = 50.**

**Q14)**

John notices DAB’s stock has beaten the market in September for the past five years consecutively, so he decides to purchase the stock at the beginning at September and sell it in the end of the month. However, DAB’s stock declined and had a lower return compared to the S&P 500. What is the minimum form of market efficiency that explains this situation?

* 1. Weak form efficiency
  2. Semi-strong form efficiency
  3. Strong form efficiency
  4. Inefficiency

**Solution: A**

**Explanation: The market is in weak form efficiency because John didn’t make an excess return by selecting the stock based on historical earnings data. Weak form efficiency indicates the current price is not influenced by the past information, which matches the situation mentioned in the question.**

**Q15)**

A technical trader placing a lot of weight on newly formed candles, making him lose track of the long-term trends. What behavioural bias(es) does this trader exhibit?

* 1. Overconfidence
  2. Loss aversion
  3. Recency effect
  4. Anchoring

**Answer: C - Definition of recency effect.**

**Q16)**

Let’s construct a CMV (Consonant Minus Vowel) factor by creating a factor-mimicking portfolio where we go long on stocks whose symbols begin with a consonant and short on stocks whose symbols begin with a vowel. Suppose it has provided an average annual premium of 3% from 1980 to 2018. Which of the following factor categories should this factor belong to?

A. Macroeconomic Factors

B. Statistical Factors

C. Fundamental Factors

D. None of the above

**Answer**: D

**Explanation: This factor does not belong to Macroeconomic, Statistical or Fundamental factors as it does not satisfy the criteria for any of these categories.**

**Instructions for Q17 and Q18**

Suppose we ran a factor regression for a stock fund to see which factors explains its returns and got the following output:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | *Coefficients* | *Standard Error* | *t Stat* | *P-value* |
| Intercept | -0.003 | 0.004 | -0.724 | 0.473 |
| Mkt-RF | 0.757 | 0.140 | 5.394 | 1.634E-06 |
| SMB | -0.721 | 0.159 | -4.543 | 3.238E-05 |
| HML | -0.056 | 0.165 | -0.338 | 0.736 |

Where Mkt-RF is the excess market return, SMB is the Size factor and HML is the Value factor

**Q17)**

This fund is most likely a:

1. Growth Fund
2. Large-Cap Fund
3. Small-Cap Fund
4. High Beta Fund

**Answer: B**

**Explanation: As the coefficient of the Size factor is negative and significant, this fund most likely is a Large-Cap Fund. Please note although the coefficient of Value factor is negative, but it is not significant. Hence, this fund in all probability will not be a Growth Fund.**

**Q18)**

What is the most likely interpretation of the intercept term from the regression output?

1. The fund manager generates a positive alpha which is statistically significant
2. The fund manager’s alpha is not statistically significant different from zero
3. The fund manager generates a negative alpha which is statistically significant
4. None of the above

**Answer: B**

**Explanation: As the coefficient for the intercept is not statistically significant, the answer is B.**

**Q19)**

Historically, the value premium has been positive but between 2009 and 2018, the average annual premium on value stocks was -2.33%. Which trait of the value factor is most likely represented by this underperformance?

1. Cyclicality of the value factor over certain time horizons
2. Value factor is negatively correlated to the market factor
3. Value factor does not exist any more
4. Negative correlation of the value factor with size factor

**Answer: A**

**Explanation: The underperformance most likely represents the trait of cyclicality of systematic factors. The systematic factors have been sensitive to macro-economic and market forces and have underperformed the overall market for long periods of time.  
  
B and D are ruled out because correlations don’t explain under or over performance. C can be ruled out because a few years of underperformance don’t imply value factor doesn’t exist anymore. Cyclicality of factors is a quality shared by all factors and as they are cyclical, so some factors can have underperformance in the short term.**

**Q20)**

Which of the following is most likely to happen in the long-term if **excessive** investments are made into systematic factor funds?

1. The average annual return premiums of the factors will go down
2. The average annual return premium of the factors will go up
3. The average annual return premiums of the factors will not change
4. None of the above

**Answer: A**

**Explanation: As more and more money is poured into factor indexes, the factor premiums are most likely to diminish as the strategies will become overcrowded.**

**CODING QNS**

**Q21)**

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?

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

**Answer: C**

**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

**Q22)**

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
4. Linear-Log

**Answer: B**

**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

**Q23)**

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

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

**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*

**Q24)**

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

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^-07) 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.

**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*

**Q25)**

Interpret the coefficient of the independent variable for the Log-Log model.

1. 1% increase in Room.Board leads to 0.40568% decrease 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 in Personal

**Answer: A**

Explanation: 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 %

**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 Q26 – Q30**

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. 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.

**Q26)**

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**

**Q27)**

How to interpret the coefficient of gpa?

1. If gpa increases by 1 unit, the natural log of the odds of admission increases by 0.755.
2. If gpa increases by 1 unit, the odds of admission increase by 0.755.
3. If gpa increases by 1 unit, the odds of admission increase by exp(0.755).
4. All of the above.

**Answer: A**

**Q28)**

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**

**Q29)**

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\*0.1 + 0.755\*330)/[1 + exp(-4.949 + 0.003\*0.1 + 0.755\*330)]
2. exp(-4.949 + 0.003\*0.1 + 0.755\*330)
3. 0.0755
4. None of the above

**Answer: C**

**Q30)**

What is the value of area under the curve (AUC) for the model created? Please select the closest answer.

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

**Answer: C**

**Code:**

# Load dataset

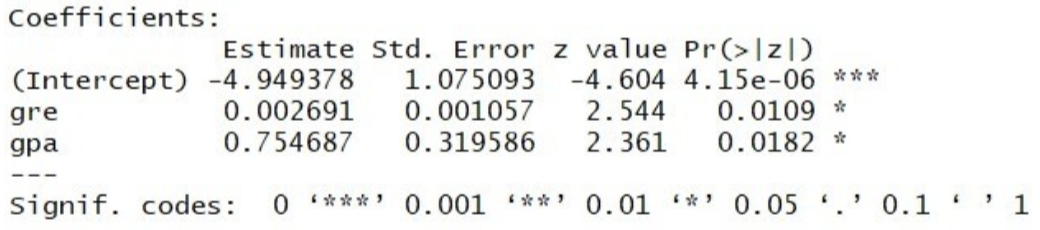
mydata <- read.csv("binary.csv")

# Create Logistic Model

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

# Model summary

summary(mylogit)

****

**Instructions for Q31 to Q38**

Use the dataset **Berkshire.csv** with the following variables.

* 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

**Q31)**

What is the standard deviation of Berkshire Hathaway over the sample period?

1. 6.75%
2. 6.81%
3. 6.86%
4. 6.90%

**Ans: A (code given below)**

**Q32)**

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

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

**Ans: B (code given below)**

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

sd1 <-sd(data$BRKRet, na.rm = TRUE) #0.0675

mean<- mean(data$BRKRet, na.rm = TRUE) #0.0189

**Q33)**

Relative to the aggregate market, Berkshire Hathaway has:

a. Underperformed the market

b. Outperformed the market by 0.25% to 0.50% per month on average

c. Outperformed the market by greater than 0.75% per month on average

**Answer: C**

**Q34)**

$10,000 invested in Berkshire Hathaway at the start of the sample period would have grown to \_\_\_\_\_\_ by the end of the sample period

a. $900,000

b. $10,000,000

c. $25,000,000

d. Over $30,000,000

**Answer: D**

**Q35)**

What is Berkshire Hathaway’s monthly Sharpe ratio?

a. 0.10

b. 0.55

c. 0.80

d. 0.23

**Answer: D**

**Q36)**

Berkshire Hathaway’s Sharpe Ratio is \_\_\_\_\_\_\_\_ than the aggregate stock market?

a. Higher

b. Lower

**Answer: A**

**Q37)**

What is Berkshire Hathaway’s estimated beta?

a. 0.50

b. 1.25

c. 0.70

d. Greater than 1.25

**Answer: C**

**Q38)**

On a monthly basis, what is Jensen’s alpha for Berkshire Hathaway?

a. 0.52%

b. 1.08%

c. 0.25%

d. -0.50%

**Answer: B**

**Instructions for Q39 to Q40**

In this question, we will determine the factors explaining the returns for the HiTec industry portfolio (as defined by Fama and French [here](https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/det_10_ind_port.html))

We will build a factor regression model using the data in the **Factor\_HiTech.csv** file to answer the questions below.

In the file,

* Mkt\_rf: Monthly excess return on the aggregate stock market
* RF: Risk Free rate
* SMB: Size Factor
* HML: Value Factor
* QMJ: Quality Factor
* BAB: Betting against beta factor
* Mom: Momentum factor
* HiTec: Monthly return on the HiTec industry portfolio

**Q39)**

What is the coefficient of the Value factor in the factor regression (choose the closest answer)? What does the result for the value factor say of the investment style of the portfolio (choose the most likely answer)?

1. 1.11, the portfolio invests in growth stocks
2. -0.09, the portfolio invests in value stocks
3. -0.54, the portfolio invests in growth stocks
4. -0.04, we cannot say as coefficient is not statistically significant

**Answer: C**

**Q40)**

What is the correlation between the Value factor and the Momentum factor over the entire period of the data set? What does the value of correlation suggest about the diversification benefits to investors in this portfolio?

1. -0.211, negative correlation does not suggest a diversification benefit
2. -0.211, negative correlation suggests a diversification benefit
3. 0.211, positive correlation does not suggest a diversification benefit
4. -0.411, negative correlation does not suggest a diversification benefit

**Answer: B**

**Explanation for Q39 and Q40:**

**Code**

factor\_hitec <- read.csv("Factor\_HiTec.csv")

factor\_hitec$HiTec\_rf <- factor\_hitec$HiTec - factor\_hitec$RF

factor\_model\_hitec <- lm(HiTec\_rf ~ Mkt\_rf + SMB + HML + Mom + BAB + QMJ, factor\_hitec)

summary(factor\_model\_hitec)

cor(factor\_hitec$Mom, factor\_hitec$HML)

**Output**

Call:

lm(formula = HiTec\_rf ~ Mkt\_rf + SMB + HML + Mom + BAB + QMJ,

data = factor\_hitec)

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 \*\*\*

Mom -0.093566 0.030698 -3.048 0.002438 \*\*

BAB -0.286339 0.040218 -7.120 4.25e-12 \*\*\*

QMJ -0.038604 0.073603 -0.524 0.600193

---

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

cor(factor\_hitec$Mom, factor\_hitec$HML)

[1] -0.2114656

**Answer to Q39** – From the above output, we can see that the coefficient of HML is -0.544. As the coefficient is negative and statistically significant, thus the portfolio has a growth tilt.

**Answer to Q40 –** From the above code, we see that the correlation between HML and Mom is -0.211, which implies a diversification benefit as negatively correlated factors provide a great diversification benefit.

High correlation (i.e. high positive number) are less desirable for portfolio construction as the diversification benefits are lower than with low (i.e. low positive number ) or negative correlations (i.e. low or high negative number) so the small the low magnitude of correlation is not an issue.