SAMPLE MID
THEORY QNS

Q1) Consider a linear regression model with 2 independent variables (assume both are correlated with the response variable). If we add an interaction term between the independent variables to the model, how will the model be affected:

A) The R^2 will increase (or remain the same) with certainty while the adjusted R^2 can increase or decrease
B) Both the R^2 and adjusted R^2 will increase with certainty.
C) The R^2 will decrease (or remain the same) with certainty while the adjusted R^2 can increase or decrease
D) Both the R^2 and adjusted R^2 will decrease with certainty.

Solution A: The R^2 is bound to increase with the addition of new variables or stay the same if the interaction variable doesn't improve model. The adjusted R^2 adds a penalty term on the number of variables in the model, hence it may go down or up (if the new interaction variable offers significant predictive performance).

Q2) Which of the following is **NOT** a binary dependent variable?
(a). Whether a customer will default on his debt.
(b). Would a student pass a course.
(c). Change in value of an investment.
(d). If a firm would go bankrupt in the next year.
Answer – Option C (Week 4 Lesson 2)
Change in value of an investment can have more than two values. Rest all only take two values. Hence Option C is not a binary dependent variable

Q3) The odds for your team winning is 0.6 in the next game. What is the probability of your team losing in the next game?
(A) 0.4
(B) 0.375
(C) 0.6
(D) 0.625

Ans) (D) (Week 4 Lesson 1)
odds = 0.6 = p/(1-p)
This means that p = 0.6/1.6 = 0.375
Hence probability of team losing = 1-0.375 = 0.625

Q4) While calculating a difference in difference, we run a regression which is as follows: lm(y ~ d1 +d2 + d3)where d1 and d2 are dummy variables and d3 is their interaction term. We thus get its coefficients according to the below equation: Y = a + b*d1 + c*d2 + d*d3
What is the difference in difference estimator?

(A) a
(B) (d-c)-(b-a)

(C) a+b+c+d
(D) d

Answer: (D) d - Difference in difference estimator is given by coefficient of interaction term (Week 5 Lesson 5)

Q5) We want to observe a column “y” in dataset. We divide the observations into 2 parts, where y_0 is the set of observations of control group and y_1 is the set of observations of treatment group. (Let function mean(X) gives the mean value of X) What is the difference estimator given by?

(A) mean(y_1) – mean(y_0)
(B) Covariance(y_1, y_0)
(C)1 – mean(y_1)/mean(y_0)
(D) mean(y_0) - mean(y_1)

Answer: (A) (Week 5 lesson 3)
Difference estimator = Cov(y1,y0)/Cov(y0,y0) = mean(y_1) – mean(y_0)

Q6) Which of these asset classes has historically been the safest (least risky)?

a) Small Cap Stocks
b) Large Cap Stocks
c) Corporate Bonds
d) Treasury Bonds

Answer: D: Explanation: Treasury Bonds have historically been the safest asset (Week 6 Lesson 3)

Q7) 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?

a) $50
b) $100
c) $150
d) $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. (Week 6, Self Assessment & Lesson 1)

Q8) A speculative fund manager wants to take advantage of mis-pricing in the market, where he sees a stock trading at $13.00 after the closing bell at around 5:00 PM and decides to place a bet the next day once the market opens. Unfortunately, when he wakes up, he sees the the best bid and offer prices of the stock (in $) as follows:

```
Bid:  13.20 x200          Ask:  13.27 x1,000
```

What is the delay cost per share he will incur if he places a market order immediately?

a) $0.20

b) \$0.27

c) \$0.235

d) There is no delay cost and the order will get executed at \$13

Solution:

b) If the speculator wants to buy a stock on a market order, it'll be executed in the ask price, which is 13.27. Hence, delay cost of (13.27-13) = \$0.27.

Q9) The stock price of Tesla on 31st December 2019 was \$418.33 and on 25th February 2020 it is \$799.91. Its book value per share is \$41.25 which is same for both dates. Calculating the change in value factor (book value to price ratio) and momentum (percentage change in price). What does this imply about the value factor and momentum factor?

A. Value Factor and Momentum Factor both go up
B. Value Factor and momentum Factor both go down
C. Value Factor goes up and momentum Factor goes down
D. Value Factor goes down and momentum Factor goes up

Solution: D (Week 8 Lesson 1 slide 7,9)
A high increase in price results in high increase in momentum. Also, this increase in price results in a decrease in book to price ratio, thereby decreasing the value. There is strong negative relation between the two. Intuitively, increase in momentum => increase in price => decrease in book to price ratio => decrease in value.

Q10) 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*
Intercet	-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. Looking at the coefficients of the factors, this fund is most likely a:

A. Growth Fund
B. Large-Cap Fund
C. Small-Cap Fund
D. Value Fund

Answer: B (Week 8 Lesson 2 Slide 10,12)
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.

CODING QNS

Q1) 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?

a) 0.00549, 0.048
b) 0.0143, 0.022
c) 0.0398, 0.0385
d) 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?

a) Log-Linear
b) Log-Log
c) Linear-Linear
d) 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 Log-Log model.

a) 1% increase in Room.Board leads to (e^(0.0040568)-1)*100% decrease in Personal
b) 1 unit increase in Room.Board leads to 0.40568*100% decrease in Personal
c) 1 unit increase in Room.Board leads to (e^0.40568)*100% decrease in Personal
d) 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 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 Q4 – 5
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.

Q4) How to interpret the coefficient of gre?

A. If gre increases by 1 unit, the natural log of the odds of admission increases by 0.003.
B. If gre increases by 1 unit, the odds of admission increase by a factor of exp(0.003).
C. If gre increases by 1 unit, the odds of admission increase by roughly 100*0.003 percent.
D. 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)**

```
# Load dataset
mydata <- read.csv("binary.csv")
# Create Logistic Model
mylogit <- glm(admit ~ gre + gpa, data = mydata, family = "binomial")
# Model summary
summary(mylogit)
```

```
Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -4.949378   1.075093  -4.604 4.15e-06 ***
gre          0.002691   0.001057   2.544   0.0109 *
gpa          0.754687   0.319586   2.361   0.0182 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Q5)
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?

A. exp(-4.949 + 0.003*0.1 + 0.755*330)/[1 + exp(-4.949 + 0.003*0.1 + 0.755*330)]
B. exp(-4.949 + 0.003*0.1 + 0.755*330)
C. 0.0755
D. None of the above

Answer: C The question is about how the natural log changes, and so it is similar to Q27, and it simply changes by the value of the coefficient of GPA, and so the answer is **(C) (Week 4 Lesson 3)**

Instructions for Q6 to Q8

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

Q6) 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 (week 6 Lesson 3)

```
df <- read.csv("Berkshire.csv", header = TRUE)
df$ExcessPerformance<- df$BrkRet-df$MKT
mean(df$ExcessPerformance)
```

Q7)

$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 (Week 6 Lesson 1)

```
(prod(df$BrkRet+1)-1)*10000
```

Q8) What is Berkshire Hathaway's monthly Sharpe ratio?

a. 0.10
b. 0.55
c. 0.80
d. 0.23

Answer: D (Week 7 Lesson 1)

Instructions for Q9 to Q10

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.

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

Q9) The coefficient of HML for the three factor model for UPS suggests that:

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 1)

Q10) Based on their three factor model, which firm has a higher level of performance? What is this firm's return (performance level)?

A. UPS, 0.06% per month
B. UPS, 0.09% per month
C. KO, 0.2 % per month
D. KO, 0.2% per year

Answer: C (Week 8 Lesson 2)

Explanation for Q9 and Q10:

```
fulldata <- read.csv("UPS_KO.csv")
fulldata$KO_RF <- fulldata$KO - fulldata$RF
fulldata$UPS_RF <- fulldata$UPS - fulldata$RF
data <- fulldata[which(fulldata$Date <= 201908 & fulldata$Date >= 201409),]
factorUPS <- lm(UPS_RF~Mkt_RF+SMB+HML, data=data)
factorKO <- lm(KO_RF~Mkt_RF+SMB+HML, data=data)
stargazer(factorUPS, align = TRUE, type = 'text', out = "factormodelfull.html")
summary(factorUPS)
stargazer(factorKO, align = TRUE, type = 'text', out = "factormodelfull.html")
summary(factorKO)
```

KO:

```
> stargazer(factorKO, align = TRUE, type = 'text', out = "factormodelfull.html")

===============================================
                        Dependent variable:
                    ---------------------------
                              KO_RF
-----------------------------------------------
Mkt_RF                      0.540***
                             (0.104)

SMB                         -0.944***
                             (0.149)

HML                           0.008
                             (0.147)

Constant                      0.002
                             (0.004)

-----------------------------------------------
Observations                   60
R2                            0.489
Adjusted R2                   0.462
Residual Std. Error      0.027 (df = 56)
F Statistic          17.892*** (df = 3; 56)
===============================================
Note:               *p<0.1; **p<0.05; ***p<0.01
> summary(factorKO)

Call:
lm(formula = KO_RF ~ Mkt_RF + SMB + HML, data = data)

Residuals:
      Min        1Q    Median        3Q       Max
-0.072350 -0.023079  0.000607  0.020961  0.051588

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.001875   0.003697   0.507    0.614
Mkt_RF       0.540239   0.103711   5.209 2.82e-06 ***
SMB         -0.944037   0.148826  -6.343 4.18e-08 ***
HML          0.007514   0.146611   0.051    0.959
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.02735 on 56 degrees of freedom
Multiple R-squared:  0.4894,	Adjusted R-squared:  0.462
F-statistic: 17.89 on 3 and 56 DF,  p-value: 2.887e-08
```

UPS:

```
> stargazer(factorUPS, align = TRUE, type = 'text', out = "factormodelfull.html")

===============================================
                        Dependent variable:
                    ---------------------------
                              UPS_RF
-----------------------------------------------
Mkt_RF                       1.185***
                             (0.183)

SMB                           -0.050
                             (0.262)

HML                           0.386
                             (0.258)

Constant                     -0.0001
                             (0.007)

-----------------------------------------------
Observations                    60
R2                            0.457
Adjusted R2                   0.428
Residual Std. Error      0.048 (df = 56)
F Statistic          15.696*** (df = 3; 56)
===============================================
Note:               *p<0.1; **p<0.05; ***p<0.01
```

```
> summary(factorUPS)

Call:
lm(formula = UPS_RF ~ Mkt_RF + SMB + HML, data = data)

Residuals:
      Min        1Q    Median        3Q       Max
-0.132991 -0.028052 -0.001813  0.034269  0.139366

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -7.795e-05  6.508e-03  -0.012    0.990
Mkt_RF       1.185e+00  1.825e-01   6.489  2.4e-08 ***
SMB         -5.048e-02  2.620e-01  -0.193    0.848
HML          3.858e-01  2.581e-01   1.495    0.141
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.04814 on 56 degrees of freedom
Multiple R-squared:  0.4568,	Adjusted R-squared:  0.4277
F-statistic:  15.7 on 3 and 56 DF,  p-value: 1.584e-07
```