GRADED HOMEWORK #3

Question 1. Beta Value of Company A = 0.75, Company B = 1.2 & Company C = 1.5. Which company's price changes the most as the market moves 1%?
    A. Company A
    B. Company B
    C. Company C
    D. Beta & Risk are independent

Answer: C
Explanation: Highest Beta implies most risky. 1% movement in market means 1.5% movement of company C (Lesson2)

Question 2. You want to evaluate three mutual funds using the Sharpe ratio for performance evaluation. The risk-free return during the sample period is 4%. The average returns and standard deviations for the three funds and the S&P 500 index are given below.

Fund	Average Return	Standard Deviation
Fund A	18%	38%
Fund B	15%	27%
Fund C	11%	24%
S&P500	10%	22%

The fund with the highest Sharpe ratio is:
    A. Fund A
    B. Fund B
    C. Fund C
    D. S&P 500

Answer: B
Explanation: Sharpe Ratio is the Ratio of Excess Return (that is, Average Return of the Fund – Risk Free Rate) divided by the Standard Deviation.
The highest value is clearly for Fund B which is equal to (15% - 4%) / 27% = 40.74%

Question 3. With respect to efficient markets, a company whose share price reacts gradually to the public release of its annual report most likely indicates that the market where the company trades is:
    A. Semi-Strong form of efficiency.
    B. Subject to behavioural biases.
    C. Receiving additional information about the company.

Answer: C

Explanation: If markets are efficient, the information from the annual reports is reflected in the stock prices; therefore, the gradual changes must be from the release of additional new information.

It is important to note that the question simply asked, "a company whose share price reacts *gradually* to the public release of its annual..." In most cases, the reason the market reacted as it did - including the word, gradually - is because it is reacting as it is receiving additional information about the company.

Question 4. How are factor returns for the Value (HML) and Size (SMB) factors in the Fama-French 3-factor model constructed?
- **A. By building factor-mimicking portfolios which try to emulate the factor**
- B. They are estimated using cross-sectional multivariate regression
- C. Using Principal Component Analysis

Answer: A **(We shall also provide grades for selecting B)**

Explanation: The Fama-French factors HML and SMB are built using factor mimicking portfolios and the returns of these factors equal the returns of such portfolios.

Let us take an example of a popular factor that we've recently built: QMJ (Quality Minus Junk).

This factor is about taking long positions on quality stocks and short positions on junk stocks. To think about it, there isn't any stock called QMJ in market, or we do not explicitly know what a quality or a junk stock is. To solve these issues and construct monthly QMJ factor returns, we follow a detailed process:

The method to find out this "quality" for each stock is quite cumbersome, wherein a lot of financials and price data for various individual stocks is used to construct this quality score for each stock on a monthly basis. Now, once we have these quality scores, we need a portfolio that mimics the essence of the factor: factor-mimicking portfolio to construct returns of QMJ factor. For this, we create a portfolio at every month with long 30% high quality stocks and short 30% worst quality stocks based on monthly scores, and we calculate returns using the changes in portfolio value over these months. This is how we end up with QMJ factor monthly returns. Same principle is followed while constructing factor returns like HML, UMD, etc.

Another important note, these factor portfolio returns are usually built in a cross-sectional fashion. In case of QMJ, we first divide all stocks in two parts based on firm size (using median market cap for US), then construct QMJ portfolios for small stock universe and large stock universe separately, find these portfolio returns and then combine them together into a single QMJ factor return (by giving 50% weight to small QMJ portfolio return and 50% to large QMJ portfolio return).

We understand that this process may have been confused with regression over cross sectional return data to compute data about the factors and hence option B will also be accepted.

One may usually get this information by reading the research paper such as Fama French 3 factor or 5 factor research papers. You can check all the research papers on Ken French Website: https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

Question 5. The "Value" factor belongs to which of the following category of factors?
    A. Macroeconomic Factors
    B. Statistical Factors
    C. Fundamental Factors
    D. None of the above – no need

Answer: C
Explanation: The Value factor belongs to the class of fundamental factors as it is based on the Book to Market ratio.

Instructions for Question 6-7

The following questions are based on the **Advertising** dataset (**Advertising.csv**). The sales are in thousands of units, while the advertising budgets (TV, Radio, Newspaper) are in thousands of dollars.

Question 6. Now that we have our linear regression model, let's try to make a prediction for the sales given a new set of advertising budgets as followed:

```
new.dat <- data.frame(TV=200, Radio=10, Newspaper=20)
```

You are required to report the predicted distance as well as the lower and upper bound for the 95% *prediction* interval. What will you report?
    A. The predicted sales value is $13,543.06, with a 95% prediction interval of $10,210.25 and $16,875.87.
    B. The predicted sales value is $13,956.37, with a 95% prediction interval of $10,613.31 and $17,299.43.
    C. The predicted sales value is $15,852.04, with a 95% prediction interval of $12,508.44 and $19,195.64.
    D. The predicted sales value is $9,379.90 with a 95% prediction interval of $6,038.61 and $12,721.20.

Answer: B.
Explanation: Use the predict function in R and change the interval to "prediction" and level to "0.95"

```
new.dat <- data.frame(TV=200, Radio=10, Newspaper=20)
predict(lm, newdata=new.dat, interval='prediction', level=0.95)
       fit      lwr      upr
1 13.95637 10.61331 17.29943
```

Question 7. Use the confint() function in R and compute the confidence intervals for all three advertising budgets. Which of the following statement is correct?

    A. The confidence interval for TV is narrow and far from zero when compared to that of Newspaper.

    B. The confidence interval for Radio is narrow and far from zero when compared to that of Newspaper.

    C. The confidence interval for Newspaper includes zero, which indicates the variable is not statistically significant given the values of TV and Radio.

    D. All the above statements are correct.

Answer: D **(But we shall grade any choice for this question to be correct)**
Explanation: All the above statements are logical interpretation of the result of running the confidence intervals.

```
confint(lm)
                    2.5 %       97.5 %
(Intercept)    2.32376228  3.55401646
TV             0.04301371  0.04851558
Radio          0.17154745  0.20551259
Newspaper     -0.01261595  0.01054097
```

Question 8. Which of the following statements **is correct** with respect to Bounce Rate?

    A. Bounce Rate gives an indication of the proportion of visitors who did not interact with the website.

    B. Bounce Rate tells us how long, on average, visitors are staying on our website.

    C. Bounce Rate increases when someone loads a page and decreases after 30 minutes of inactivity.

    D. A high Bounce Rate generally indicates that the website entrance pages are very relevant to the website's visitors.

Answer: A
Explanation: Statement A is correct.
B) **Average Session Duration** defines how long, on average, visitors are staying on the website.
C) A **session** starts right away when someone loads a page and ends after 30 minutes of inactivity.
D) A high Bounce Rate generally indicates that the website entrance pages are not relevant to the website's visitors.

Question 9. A Vickrey auction is a type of sealed-bid auction. Bidders submit written bids without knowing the bid of other people in the auction. The highest bidder wins but the price paid is the (second-highest bid + $ 0.01). This type of auction gives bidders an incentive to bid their true value.
Following are the CPC (Cost per Click) bids of 4 digital advertisers for the keyword "candy":

1. Advertiser A: $0.92
2. Advertiser B: $0.87
3. Advertiser C: $1.08
4. Advertiser D: $0.93

Who wins the auction? What is the winner's CPC?
   A. Advertiser D, $0.92
   B. Advertiser D, $0.93
   C. **Advertiser C, $0.94**
   D. Advertiser C, $1.09

Answer: C
Explanation: Advertiser C is the highest bidder. The second highest bid is $0.93. Thus, winner is C and pays a CPC is ($0.93 + $0.01) = $0.94

Question 10. In which scenario will you not be able to run a successful A/B test on a landing page of a website?
   A. Changing website's background colour to attract more visitors, with all else unchanged
   B. Enlarging the website's sign-up button to increase the new leads, with all else unchanged
   C. **Redesigning the whole website at once, including logos, images, background colour, headings and button designs**
   D. In all cases above, the website can be improved through A/B testing

Answer: C
Explanation: In A/B testing, one needs a clear control(A) and test(B). In A, the difference between test and control is background color. B is sign-up button. C is testing too many elements at the same time, no clear control and test, and makes it hard to improve website from the statistical results. C can use multivariate testing instead.

It is important to consider that option C is the only one where a clear vision has not been defined, apart from the fact that multiple elements are being tested at the same time, which is a pre-requisite of A/B testing. While redesigning the whole website could be considered as the control, we have not been given why exactly is it being done.

The following explains why testing too many elements may cause discrepancies:

Testing too many elements together:

Industry experts caution against running too many tests at the same time. Testing too many elements of a website together makes it difficult to pinpoint which element influenced the success or failure of the test most. Apart from this, more the elements tested, more needs to be the traffic on that page to justify statistically significant testing. Thus, prioritization of tests is indispensable for successful A/B testing.

(Source: https://vwo.com/ab-testing/)

Question 11. The objective of Conversion Rate Optimization (CRO) is to:
   A. Increase number of website visitors
   B. Increase website sales
   C. Enhance engagement
   D. **All the above**

Answer: D
Explanation: From the slides, we learnt that CRO is performed for all the following –

Conversion rate optimization (CRO) is the systematic process of increasing the percentage of website visitors who take a desired action. Desired actions are different based on website goals. So above choices are general goals of all websites, and they are CRO goals.

Question 12. A website that uses Google Analytics wants to know the percentage of visitors that do not interact with the website. Which metric should be used?
   A. Page per sessions
   B. Pageviews
   C. Users
   D. **Bounce Rate**

Answer: D
Explanation: The following is how every choice may be defined –
A: Page Per Session - dividing the total number of pageviews by the total number of sessions. It is good indicator of overall user engagement.
B: Pageviews - any view of a page that is being tracked by Google Analytics.
C: Users - Total number of unique visitors to the website
D: Bounce Rate - It is the number of single-page sessions (bounces) divided by the total number of sessions. It shows the proportion of visitors who did not interact with the website.

Questions from Case Study: Chase (Questions 13 and 14)

Question 13. In the Chase case, Chase segmented customers based on the types of rewards they preferred. Which segmentation strategy does Chase use?
   A. **Behavioural method**

B. Demographic method
C. Psychographic method
D. All of the above

Answer: A
Explanation: The following line from the case study may be used to answer the above question
- *"Behavioral/attitudinal segmentation provided insight into how consumers used their cards and how much they valued rewards and/or what types of rewards they preferred (cashback, miles, points) as well as their channel preferences."*

Question 14. A complete economics of credit card transaction includes:
A. Card Issuer; Merchant Acquirer; Merchant
B. Card Issuer; Cardholder; Merchant Issuer; Merchant
C. Card Issuer; Cardholder; Merchant; Merchant Acquirer; Credit Card Network
D. Card Issuer; Cardholder; Merchant; Merchant Issuer; Credit Card Network

Answer: C
Explanation:
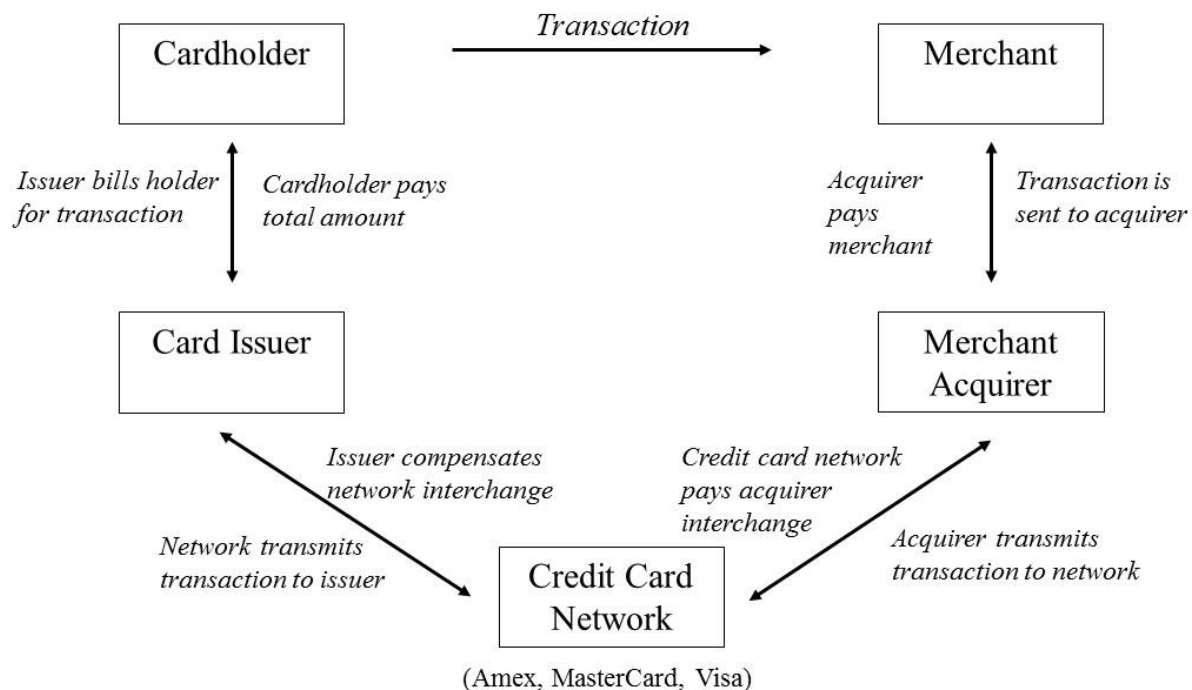
Instructions for Questions 15 to 20

Please use the Facebook Ad dataset ***KAG_conversion_data_wrangled.csv*** for the next set of questions. We advise to solve these questions using R (preferably using *dplyr* library wherever applicable) after reviewing the code provided for Week 11 and other resources provided for learning *dplyr* in R Learning Guide.

Load the dataset as:

```
data <- read.csv("KAG_conversion_data_wrangled.csv",stringsAsFactors =
FALSE)
```

Question 15. What are the number of unique interests that the dataset cover?

Answer: **40**

Code: `length(unique(data$interest))`

Question 16. Which gender was targeted by the ad which has the most Impressions among all the ads?

    A. **Male**
    B. Female

Answer: A
Code: `data %>% filter(Impressions == max(Impressions)) %>%`
`select(gender)`

Question 17. How many Impressions does the ad which has the most Approved Conversion have?

Answer: **2080666**

Code: `data %>% filter(Approved_Conversion == max(Approved_Conversion))`
`%>% select(Impressions)`

Question 18. Which ad (provide `ad_id` as the answer) among the ads that have the least CPC led to the most Impressions?

 Answer: **1121094**

Code: `data %>% filter(CPC == min(CPC)) %>% filter(Impressions ==`
`max(Impressions)) %>% select(ad_id)`

Question 19. Which campaign (provide `campaign_id` as the answer) had spent least efficiently on brand awareness on an average (i.e. most Cost per mille or `CPM`)?

    A. 916
    B. **936**
    C. **1178**

Answer: B and C
Code:

Approach 1:
```
Code: data %>% group_by(campaign_id) %>% summarise(n_ads =
length(ad_id),campaign_CPM = mean(CPM)) %>%
arrange(desc(campaign_CPM))

>1178
```

Approach 2 (More Accurate):
```
Code: data %>% group_by(campaign_id) %>%
summarise(Total_Impressions = sum(Impressions), Total_Spent =
sum(Spent)) %>% mutate(CPM = Total_Spent/Total_Impressions*1000) %>%
 arrange(desc(CPM))

>936
```

Approach 1 takes the average of CPM at a campaign level, across all the ads belonging to a particular CPM. However, an alternative way to do this is to re-calculate CPM at the campaign level (total cost for the campaign/total impressions in thousands) for comparison. First approach CPM acts as the measure acts as a non-weighted average (we are averaging CPM of both big and small ads with same priority/weight) while the second approach doesn't have this ambiguity. Another way of doing it is to take weighted average of CPMs (weights could be decided using impressions) which would give the same answer as second approach.

Which approach to use? This is dependent on the nature of the measure. If giving equal weights to each subcomponent (ad) while averaging is not ambiguous for the measure, approach 1 is more suitable. Approach 2 is more useful when there is a need to assign importance to subcomponent (ad) based on its scale (Impressions).

Question 20. How many unique ads by campaign 1178 which have never caused any approved conversion at all were targeted towards the age of 32?

 Answer: **65**

Code: `data %>% filter(campaign_id == 1178 & Approved_Conversion ==0 ) %>% group_by(age) %>% summarise(count=length(ad_id))`