**FINAL EXAM QUESTIONS**

**Extra Credit** **- 4 questions**

**Q1)** Suppose we run a factor regression for a stock fund to see which factors explain its return and get the following output:

A picture containing receipt, text

Description automatically generated

Where RM-RF is the excess market return, SMB is the Size factor, HML is the Value factor, RMW is the Robustness factor and CMA is the Conservative Factor.

Looking at the output of regression, which factor can you say is almost redundant and which factor explains the redundant factor most?

1. Robustness Factor (RMW) and Excess-Market return (RM-RF)
2. **Value Factor (HML) and Conservative Factor (CMA)**
3. Size Factor (SMB) and Value Factor (HML)
4. Conservative Factor (CMA) and Robustness Factor (RMW)
5. There is no redundant factor

**Solution:**

Value factor has the smallest magnitude of coefficient. Hence it has bare minimum effect on the regression model. The factor that explains it most is Conservative Factor (CMA) with a coefficient of 1.04 and t-stat of 23.

**Instructions for Q2 and Q3**

Please use the Facebook Ad dataset ***KAG\_conversion\_data\_wrangled.csv*** for the next set of questions. You should 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)

**Organic Impressions are ads that have generated impressions without any money spent on them. (i.e. Spent = 0).**

**Q2)**

**Which campaign (campaign\_id) has produced least number of ads with organic impressions?**

**Answer**: 1178

**Code**: data %>% filter(Spent == 0 & Impressions > 0) %>% group\_by(campaign\_id) %>% summarise(n\_ads = length(ad\_id)) %>% arrange(n\_ads)

**Q3)**

**Among ads with organic impressions, which ad (ad\_id) has the highest number of impressions?**

**Answer**: 1121094

**Code**: data %>% filter(Spent == 0 & Impressions > 0) %>% filter(Impressions == max(Impressions)) %>% select(ad\_id)

**Q4)** If the regression model log(Y) ~ + log () is implemented and the model summary is as presented below, which of the following statement is **always** true?

**Note:** These are few formulas which might come handy in case you are not aware of log operations.

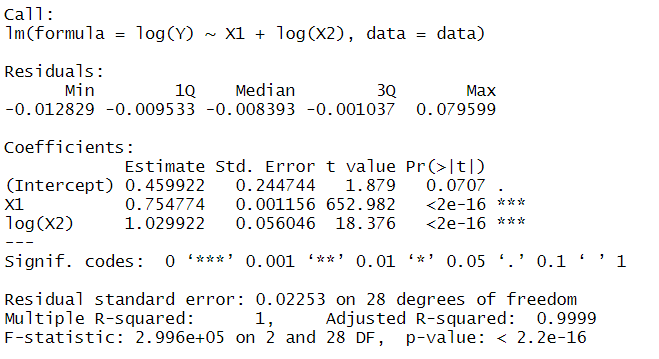
Log(A) – Log(B) = Log(A/B)

Log(A) + Log(B) = Log (AB)

Exp(A+B) = Exp(A) \* Exp(B)

Log(exp(A)) = A

Log(A) = B <==> A = Exp(B)



1. **1 unit increase in  and 1% increase in  leads to 114.9% increase in Y**
2. 1% increase in  and 1 unit increase in  leads to 106% increase in Y
3. 1 unit increase in and 1% increase in leads to 76.5% increase in Y
4. 1% increase in and 1% increase in leads to 87% increase in Y

**Solution:**

If we increase by 1 unit and by 1% the new Y will be

Subtracting equation 2 from equation 1 we have

Plugging b1 = 0.754774, b2 = 1.029922 in equation 3 we have

Therefore for 1 unit increase in X1 and 1% increase in X2 we observe a 114.9% increase in Y.