STAT 371 S22 Assignment #1

(Submission deadline: 11:59 pm May. 24)

1) The Capital Asset Pricing model (CAPM) is commonly used to describe the relationship between the expected return and risk of a security (e.g. stock) in a given market. All investors assume a certain *systematic* or market risk associated with the market return, r_m, for which they expect to be compensated above the *risk free* rate of return, r_f (e.g. the return on a 10 year government bond). The difference between the expected market return and the risk free return, $E(r_m)-r_f$, is known as the market risk premium. By investing in a security, investors assume an additional risk and thus expect a higher return on their asset.

   In the standard CAPM model, this risk-reward relationship for any given stock is described by

$$E(r) = r_f + \beta(E(r_m) - r_f)$$

   Or equivalently,

$$E(r) - r_f = \beta(E(r_m) - r_f)$$

   where $E(r)$ is the expected rate of return of the stock.

   This model states that the risk premium of the stock, given by $E(r)-r_f$, is linearly related to the market risk premium $E(r_m)-r_f$. The slope parameter, β, is known as the stock's *beta*, and is a measure of the stock's risk or volatility relative to that of the market as a whole. For example, a beta value of 1.1 indicates that the stock's volatility is 110% that of the market, whereas a stock with a beta of 0.5 suggests the stock is less risky than the market as a whole, possessing only 50% of the market's volatility. Investors routinely use beta values in their investment decisions, and will select stocks based in part on the consistency between a stock's beta and the investor's level of risk-aversion.

   The beta value for a given stock can be estimated by fitting the simple linear regression model

$$r_t - r_{ft} = \alpha + \beta(r_{mt} - r_{ft}) + \varepsilon_t$$

   which can be expressed in the form

$$y_t = \alpha + \beta x_t + \varepsilon_t \quad t = 1, 2, \ldots, T$$

   where $y_t = r_t - r_{ft}$ is the rate of return of the stock in excess of the risk-free rate at time *t*, and $x_t = r_{mt} - r_{ft}$ is the market rate of return in excess of the risk-free rate at time *t*. Note that α can be loosely interpreted as the performance of a stock relative to that of the benchmark index.

   For this assignment we will fit a SLR model to monthly return rates over the past five years ($t = 1, 2, \ldots, 60$) to estimate the beta of Bank of Montreal (BMO), a stock in the financial services sector on the S&P TSX Composite index. We will use the S&P TSX composite index (TSXC) as a proxy for the market as a whole, and the yield on a 10-year (Canadian) government bond rate (GVB10y) as a proxy for the risk-free rate of return.

   To begin, download the *capmA1* csv file that contains monthly closing prices for BMO and TSXC, as well the annualized monthly rate of return for GVB10y, from Sept. 2015 to Aug. 2020.

   Before (or after) importing the dataset into R, you will need to transform the stock and index prices to monthly rates of return, where, for example, the rate of return of BMO for month *t* can be expressed as

$$r_t = \frac{P_t - P_{t-1}}{P_{t-1}}$$

   As well, the government bond annualized return rate at time *t*,, r_{at}, will need to be converted to the monthly risk free rates:

$$r_{ft} = \left(\frac{r_{at}}{100} + 1\right)^{1/12} - 1$$

Note that by converting the prices to rates, the dataset will be reduced to 59 observations. You will also need to remove the first observation (Sept. 2015) from the government bond rates before you calculate x_t and y_t.

Once the dataset has been saved as a csv (or text) file, it can be imported into R using the *read.csv* (or *read*.txt) function. For example,

```
>capmA1=read.csv("capmA1.csv",header = TRUE)
```

(For R to find the *capmA1*file, you must make sure that the file is stored in your current working directory in R. To find the current working directory on your computer, use the command `>getwd()`)

a) Before we fit the SLR model $y_t = \alpha + \beta x_t + \varepsilon_t$, create a scatterplot of $\{x_t, y_t\}$. Does a linear model seem appropriate? What other characteristics of the plot do you see?

b) Calculate the correlation coefficient, r, between x_t and y_t using R (a quick search will provide you with the appropriate function, if necessary). Use R to verify this value by calculating r directly from the definition provided in the lessons.

c) Fit the SLR model

$$y_t = \alpha + \beta x_t + \varepsilon_t \quad \varepsilon_t \sim N(0,\sigma^2) \ \ ind. \quad t = 1,2,...,59$$

to the data. Be sure to include the *summary* output in your assignment. See R output used in the lessons for additional assistance with relevant R code.

d) Compare your estimate of beta for BMO with that stated in the BMO profile at https://www.theglobeandmail.com/investing/markets/stocks/BMO-T/. Note that there will be slight discrepancies between betas on different sites, depending on the proxies used and other details.

e) Can we conclude that the BMO stock is more (or less) volatile than the market?
   i) Answer this question by calculating a 95% confidence interval for the relevant parameter using the *confint* function in R
   ii) Verify this interval by calculating, using R, the confidence interval directly from the expression provided in lessons.

f) According to the standard CAPM, alpha should be zero. Is the result of your regression analysis consistent with an alpha = 0? Answer this question by carrying out an appropriate hypothesis test (no calculations are required – just refer to the appropriate values in the output).

g) Confirm the following output values by calculating these values in R directly from the definitions provided in lessons (be sure to show your R output for these calculations):

   Relevant R functions:

   `>residuals(name.lm)` #gives you the fitted residuals of the fitted model, name.lm.

   `>fitted(name.lm)` #gives you the fitted values, $\hat{\mu}_i$, of the fitted model, name.lm.

   `>sum(x^2)` #gives you the sum of the squares of the data vector, *x*.

   `>pt(t, df)` #gives you $P(t_{df} < t)$, where t_{df} is a *t* random variable with degrees of freedom, *df*.

   i) $\hat{\sigma}$ ii) $\hat{\alpha}$ iii) $\hat{\beta}$ iv) $se(\hat{\beta})$ v) the p-value associated with $H_0 : \beta = 0$

2) Simulation is a useful method for understanding regression concepts and investigating properties of parameter estimators. We will use R to generate random values of the response variable, *Y*, for the model

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \varepsilon_i \quad \varepsilon_i \sim N(0, \sigma^2), ind.$$

based on a subset of the explanatory variates associated with the *audit* dataset. These explanatory variates can be found in the *A1_variates* dataset.

. *A1_variates* can be imported into R using the *read.table* function:

```
.> name=read.table("C:/Desktop/.../A1_variates.txt",header=T)

#specify the location on your computer of your text file that you wish to
bring into R
```

We will use parameter values -250000 (intercept), 50 (size), -200 (age), 12000 (employees), 200000 (col), and 15000 for $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4$ and σ, respectively, to generate random values of *Y* associated with the given set of values of your explanatory variates for this model. Include these values in your output. (Note that these somewhat arbitrary values for the (unknown) parameters are consistent with the fitted model of overhead vs. these variates) Some R functions you may find useful:

- `rnorm(n,mu,sigma)    #generates n random values from N(mu,sigma)`
- `c(-250000, 50, -200, 12000, 200000) #creates the Beta vector`
- `as.matrix(cbind(rep(1,n),A1_variates)) #creates X matrix (i.e binds a column of ones to the matrix of explanatory variates)`
- `A%*%B #yields the matrix product (i.e., the inner product) of matrices A, B.`

a) **Using your student ID# as the seed***, generate (pseudo) random values of the response, overhead, for the given parameter values and set of explanatory variates.

   (**This can be accomplished with the command* `>set.seed(your ID#)`*. This will not only ensure that each of you will generate a different set of values, but will also allow you to reproduce the results of your simulation if necessary*)

b) Fit the linear model described above to the dataset containing the response variate and explanatory variates you generated.

   - `A1sim.lm = lm(y~x1+x2+x3+x4) #fits a linear model to the data, where x1,...,x4 correspond to the explanatory variates.`
   - `summary(A1sim.lm) #displays a summary of the fit`

c) Confirm the parameter estimators given in the output by using R to calculate $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$. As always, be sure to include your R code.

d) Based on your fitted model, provide a 95% confidence interval for β_1. Is this interval consistent with what you know about the 'true' model?