Homework 0 sample student responses

Problem 1 – The Ludic Fallacy?

We throw two dice. Each one is a normal die with six equally-weighted sides. What is the probability that the sum of the two numbers is less than 7?

In this problem, I am going to denote the probability that the sum of the two numbers is less than 7 as "P(S < 7)". Since the dice are equally weighted, I can assume that all outcomes are equally likely and therefore use the counting principle. According to the counting principle:

$$P = \frac{\text{\# of Outcomes of Interest}}{\text{Total \# of Outcomes}}$$

Where:

1. The # of outcomes of interest is the number of sums that are less than 7
2. The total # of outcomes is the total number of sums possible through rolling the two dice.

After drawing out all of the possible sums, it was found that there are 36 total outcomes. This can also be found by multiplying the number of sides on each die by one another, i.e. 6 x 6 = 36. I then counted how many sums were less than 7 and found 15 different combinations of the two dice that added up to less than 7. Thus, using the counting principle, I divided 15/36 and found that the P(S < 7) = ~0.417.

After being able to visualize the sample space, I identified which of the sums would result in a value less than seven (Figure 2).

1 1	**2 1**	**3 1**	**4 1**	**5 1**	6 1
2 1	**2 2**	**2 3**	**2 4**	2 5	2 6
3 1	**3 2**	**3 3**	3 4	3 5	3 6
4 1	**4 2**	4 3	4 4	4 5	4 6
5 1	5 2	5 3	5 4	5 5	5 6
6 1	6 2	6 3	6 4	6 5	6 6

Figure 2: Sample space for rolling two die with events that result in a sum of less than seven bolded.

Here, we see that, of 36 possible outcomes, 15 of the 36 possible outcomes result in a sum less than seven. Thus, following the counting principle where the numerator is the number of outcomes of interest (a sum less than 7; a value of 15 in this problem) and the denominator is the total number of outcomes (36 in this sample space), we get:

$P(\text{sum} < 7) = 15/36 \approx .417$

Conclusion: The probability that the sum of the two numbers that are rolled with two die with six equally-weighted sides is less than seven is about 0.417.

Problem 2 – "The Iron Bank will have its due"

At the Iron Bank, 62% of customers have checking accounts, 24% have savings accounts, and 17% have both checking and savings accounts.

Part A Of the Iron Bank customers who hold checking accounts, what percentage also have a savings account?

Part B What is the probability that an Iron Bank customer has neither a checking account nor a savings account?

Part C What proportion of Iron Bank customers have a savings account but not a checking account?

Part A

$$P(\text{savings account}|\text{checking account}) = \frac{P(\text{savings account, checking account})}{P(\text{checking account})}.$$

Using the information from the question prompt:

1. P(checking account) = 0.62
2. P(savings account, checking account) = 0.17

So using the formula above:

$$P(\text{savings account}|\text{checking account}) = \frac{0.17}{0.62} = \sim 0.274 \text{ or } 27.4\%$$

Of the Iron Bank customers who hold checking accounts, 27.4% also have a savings account.

P (Savings | Checking)
= (0.17/0.62)
= 0.274

This answer tells us that 27.4% of those who have checking accounts also have saving accounts.

We are tasked with finding P(Saving| Checking)

Don't forget: P(A|B)= (P(A,B))/P(B)

We are given P(Saving,Checking)= .17 P(Checking)= .62

From here, we can simply plug in the formula for conditional probability.

(.17/.62)= P(Saving|Checking)= 0.2741935

Part B

P (neither checking nor savings account)
= 1- (0.62 + 0.24 - 0.17)
= 0.31

This answer to this tells us that 31% of customers have neither account.

P(checking)= .62 P(saving)= .24 P(both)= .17

To account for overlap, 1 - (P(neither)) = 1-(P(saving)+P(checking)-P(both))

1-((.62)+(.24)-(.17)) = P(neither checking or savings)= .31

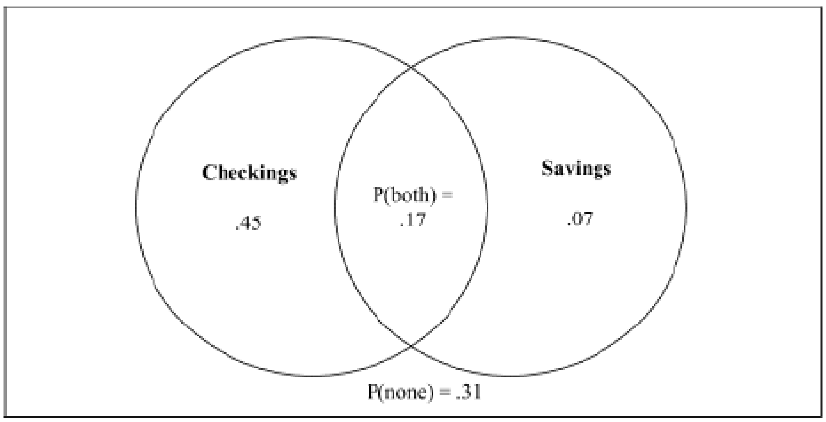


Figure 3: Venn diagram showing probabilities of different combinations regarding holding a checking or savings account at Iron Bank.

Thus, to find the probability that someone has neither, we need to first calculate the probability that someone has either one; we will calculate P(checking or savings account) using the addition rule:

P(checking or savings account) = P(checking account) + P(savings account) - P(checking and savings account) = .62 + .24 - .17 = .69

Now, we can utilize the negation rule to find the probability of someone having neither account type as one minus the probability of them having either account type.

P(neither checking nor savings account) = 1 - P(checking or savings account) = 1 - 0.69 = 0.31

Conclusion: The probability that an Iron Bank customer has neither a checking nor a savings account is 0.31.

To start off this problem, we have to find the probability of an Iron Bank customer having either a checking account or a savings account. Since these events are not mutually exclusive, we can find this through the use of the addition rule. Via the addition rule:

P (savings account or checking account) = P(savings account) + P(checking account) - P(both)

Since we would be including the probability of someone having a checking account and a savings account in each of the two individual probabilities, we subtract P(both) to ensure that this population isn't counted twice. Then, using the numbers provided in part C, we can plug in the respective values and solve:

P(savings account or checking account) = 0.24 + 0.62 - 0.17 = 0.69

Now, using the negation rule, we can find the probability of a customer having neither a checking nor a savings account:

P(neither a checking account nor a savings account) = 1 - P(savings account or checking account) = 1 - 0.69

P(neither a checking account nor a savings account) = 0.31

Part C

This problem is best understood through the use of a venn diagram. If you look at the circle indicated 'checking' in Figure 1, you can see that two numbers makeup the 62% of customers that have a checking account. As stated in the prompt, if 17% of customers have both a checking and a savings account, then 45% of customers have just a checking account. Similarly, if the percentage of customers that have both a checking and a savings account is 17%, and 24% of customers have a savings account in general, then: 0.24 - 0.17 = 0.07, indicating that 7% of Iron Bank customers have just a savings account and not a checking account. Thus, P(saving, no checking) = 0.07.

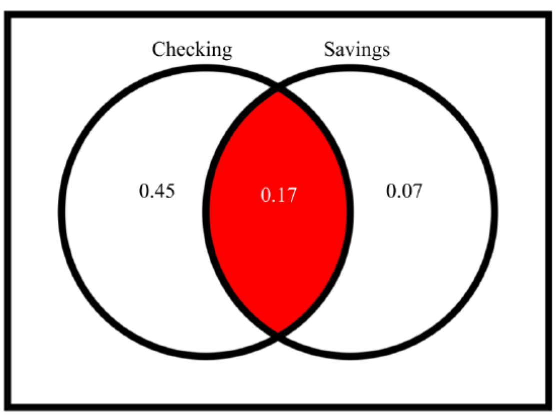


Figure 1: Venn Diagram showing how many Iron Bank Customers have a checking account, a savings account or both.

P (savings account | no checking account)
= 0.24- 0.17
=0.07

This answer tells us that 7% of those who don't have a checking account have savings accounts.

Same numbers as C. Find P(Savings but no checking)

We know 24% of all people use savings accounts, but 17% of all people use both saving and checking accounts. Thus,we must subtract the 17% from the 24% to note the amount that use savings without also using checking

(.24)-(.17) = P(savings but no checking)= **.07**

Problem 3 – Top Gun

If an aircraft is present in a certain area, a radar correctly registers its presence with probability 0.99. If it is not present, the radar falsely registers an aircraft presence with probability 0.10. Suppose that on average across all days, an aircraft is present with probability 0.05.

The radar today just registered the presence of an aircraft. What is the probability that an aircraft is actually present? Make sure to show your work.

P(R | A) = 0.99
P(R | not A) = 0.10
P(A) = .05

Thus, by the negation rule, we know that:
P(not A) = 0.95

Now, we must consider what we are trying to find: P(A | R), or the probability of an aircraft being present, given that the radar registered its presence. If we consider our Bayes' Rule equation for this, we see that:
P(A | R) = [P(A) * P(R | A)] / P(R)

The only value that we don't know from this equation is P(R). Using the total rule of probability, we can solve for this value:
P(R) = P(R, A) + P(R, not A)
P(R) = P(A) * P(R | A) + P(not A) * P(R | not A)
P(R) = .05 * .99 + .95 * .1 = .0495 + .095 = .1445

Thus, we can now plug this into our Bayes' Rule equation:
P(A | R) = [.05 * .99] / .1445 = .343

<u>Conclusion:</u> Given that the radar registered the presence of an aircraft today, the probability that an aircraft is actually present is about 0.343.

For this problem, we would use Bayes' Rule to find the posterior probability that an aircraft is actually present, given that the radar registered the presence of one. According the Bayes' Rule:

$$P(\text{Aircraft Present}|\text{Radar Registered}) = \frac{P(\text{Aircraft Present}) * P(\text{Radar Registered}|\text{Aircraft Present})}{P(\text{Radar Registered})}$$

1. Given that a radar correctly registers an aircraft's presence if one is present in a certain area with a probability of 0.99, we know that P(Radar Registered|Aircraft Present) = 0.99
2. Given that a radar falsely registers an aircraft's presence if one is not present with a probability of 0.10, we know that P(Radar Registered| Aircraft Not Present) = 0.10
3. Through the use of the negation rule, we further know that P(Radar Didn't Register| Aircraft Not Present) = 1 - P(Radar Registered| Aircraft Not Present) = 1 - 0.10 = 0.90
4. Given that on average across all days, an aircraft is present with probability 0.05, we know that P(Aircraft Present) = 0.05

5. Via the negation rule, we know that P(Aircraft Not Present) = 1 - P(Aircraft Present) = 1 - 0.05 = 0.95

With these probabilities in hand, now we can start to solve. Starting with P(Radar Registered), we would use the Rule of Total Probability to break the probability that a radar registers the presence of an aircraft into the sum of the probabilities for all the different ways in which this event can happen:

P(Radar Registered) = P(Radar Registered,Aircraft Present) + P(Radar Registered, Aircraft Not Present)

Via the multiplication rule:
P(Radar Registered) = P(Aircraft Present) * P(Radar Registered|Aircraft Present) +
P(Aircraft Not Present) * P(Radar Registered|Aircraft Not Present)

Now if we plug in the numbers from above:
P(Radar Registered) = (0.05 * 0.99) + (0.95 * 0.10) = 0.1445

Going back to our equation from Bayes' Rule, we now have all of the information we need to solve for the posterior probability: P(Aircraft Present|Radar Registered)

$$P(\text{Aircraft Present}|\text{Radar Registered}) = \frac{(0.05 * 0.99)}{0.1445}$$

Once we plug all of that into a calculator, we get that the posterior probability that an aircraft is actually present given that the radar registered the presence of one as P(Aircraft Present|Radar Registered) = ~ 0.343.

Problem 4 – "Picture me rollin"'

Download the `s550.csv` data set from the course Canvas site. This file contains a sample of 1501 Mercedes S-Class S550 cars offered for sale in the U.S. via cars.com. There are several variables in the data set, but for this problem the three relevant ones are `mileage`, `price`, and `year`.

Using the `ggplot` function within the `ggplot2` R library, make a scatterplot of price (y) versus mileage (x), faceted by year. (That is, there should be a separate panel for each year.) Format your figure so that it has 2 rows of 4 panels each, so that it looks wider than it is tall.

All you need to include in your write-up for this problem is the figure itself. To copy and paste from RStudio directly into a Word file or Google doc, click the `Zoom` button in RStudio under the plots tab. This should give you a resizeable window that you can copy/paste directly from. Pick an aspect ratio for the figure that looks good in your write-up.

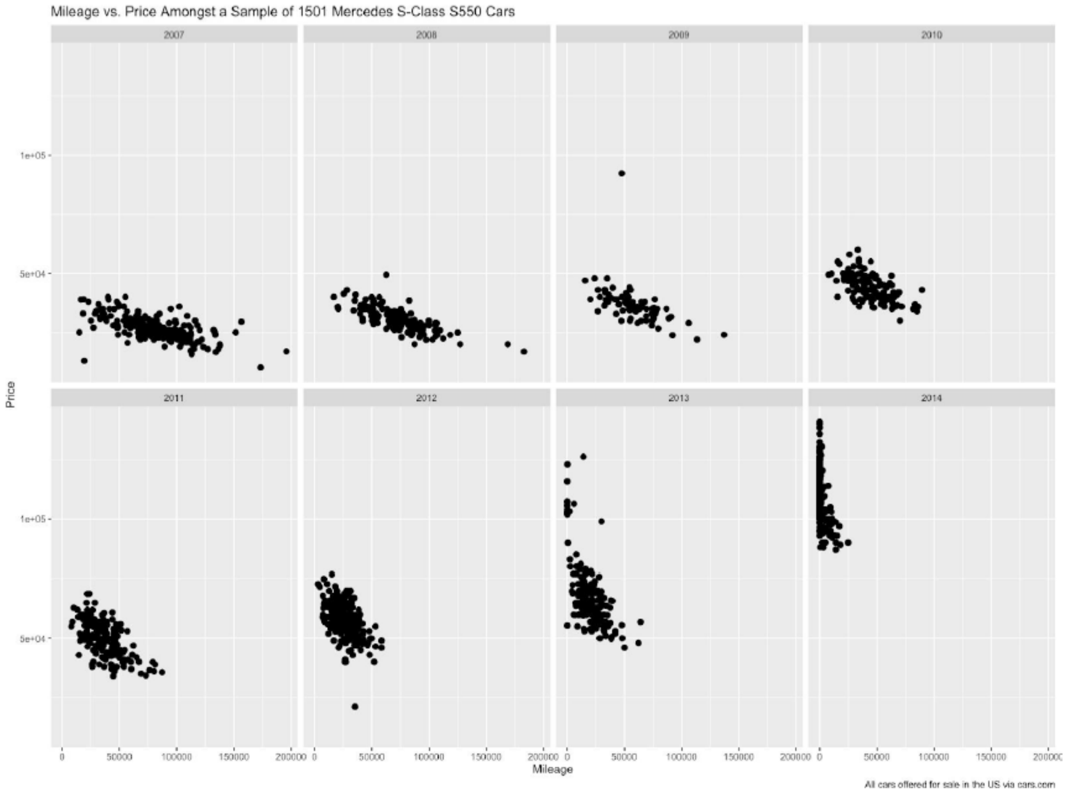
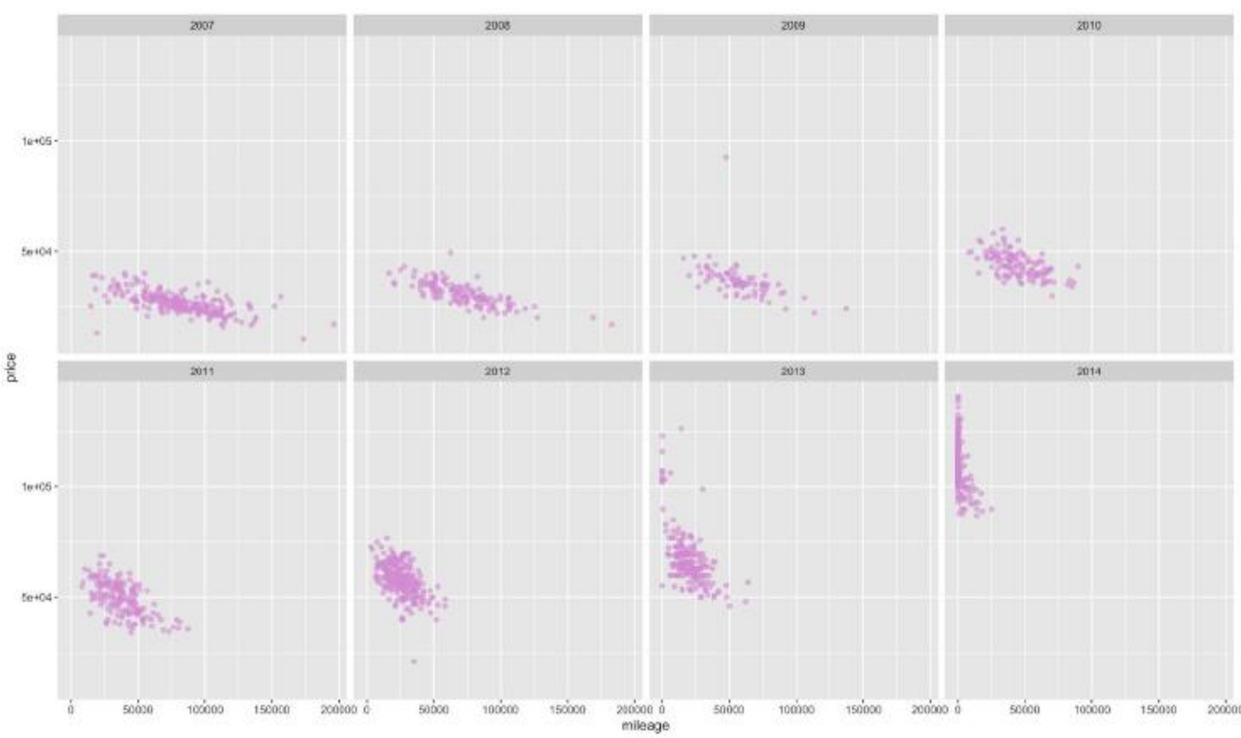


Figure 2: Scatterplot of price vs. mileage for a sample of 1501 Mercedes S-Class S550 Cars

---



Problem 5 – CAPM

An important model that finance professionals use to understand asset prices is called the Capital Asset Pricing Model (CAPM). You will learn more about the CAPM in a future finance course. But the basic assumption of the model is that the rate of return on an individual stock is linearly related to the rate of

return on the overall stock market. That is, each stock's rate of return is assumed to follow a linear regression model:

$$Y_t^{(k)} = \beta_0^{(k)} + \beta_1^{(k)} X_t + e_t^{(k)} ,$$

where $Y_t^{(k)}$ is the rate of return of an individual stock (k) in some given time period t; X_t is the rate of return of the entire stock market in that same time period; and $e_t^{(k)}$ is the residual for stock k in that time period. The superscript (k)'s here are simply denoting the different stocks (Apple, Target, etc), while the subscript t's are denoting the different time periods. Note that the market rate of return (X_t) is a predictor common to all stocks. (The rate of return can be interpreted similarly to an interest rate. For example, if a stock was worth \$100 yesterday and \$102 today, then it gained 2%, for an implied daily rate of return of 0.02.)

The β_1 (slope) term in this regression model is super important to finance professionals; they just call it "beta", and they refer to the β_0 (intercept) term as "alpha." Please watch this short YouTube video to understand how beta is used to think about different stocks.

Once you've watched the video, please turn to the data in `marketmodel.csv`, which contains information on the daily returns for the S&P 500 stock index, denoted SPY, along with the returns for 6 individual stocks: Apple (AAPL), Google (GOOG), Merck (MRK), Johnson and Johnson (JNJ), Wal-Mart (WMT), and Target (TGT). (We can think of the return of the S&P 500 as a proxy for the whole market.) The data start from the beginning of 2019. The entries are interpretable as percentage returns, expressed on a 0-to-1 decimal scale—for example, if the S&P 500 gained 1.5% in value on a given day, the corresponding entry in the data frame would be 0.015.

Regress the returns for each of the 6 stocks individually on the return of S&P 500 (which is like X_t, the market return, in the equation above). Make a clean, professional looking table (e.g. in Excel) that shows the ticker symbol, intercept, slope, and R^2 for each of the 6 regressions.

In your write-up, you should include:

- a two-to-three paragraph introduction, in your own words, on what the "beta" of a stock is measuring and how it is calculated. (Watch the video and summarize it in your own words, making sure to connect it to the regression model we've written down above—this is a bridge you will have to make yourself, using what you know about regression models.) A reasonable aim for your summary is about 250 words here, but this is approximate; nobody on our end is breaking out the word counter.

- the table itself, along with an informative caption below the table, no more than 2-3 sentences in length, to give readers the information necessary to interpret the table.

- a conclusion that answers two questions: in light of your analysis, which of these six stocks has the *lowest* systematic risk? And which has the ***highest*** systematic risk? (Again, watch the video to understand how this is measured using the regression model.)

The "beta" of a stock measures how much the rate of return of an individual stock responds to a one percent change in the rate of return of the overall stock market. For example, if the "beta" value for a specific stock is 0.33, then when the overall market increases by one percentage point, the rate of return for that specific stock would increase by 0.33 of a percentage point. When the “beta” of an individual stock is greater than one, the individual stock goes up or goes down more than the market. Thus, “beta” is a measure of systematic risk, with a “beta” of zero equaling no systematic risk.

In order to calculate the “beta” of a stock, you would find a regression model for the individual stock through fitting an equation in RStudio, where the x-axis was the market’s rates of return and the y-axis was that specific stocks' rates of return. Then looking at the regression model, the slope, or B_1, would represent the “beta” of that individual stock. In the regression model stated in the prompt, we can see why the slope represents the “beta” value. Given that the predictor variable represents the rate of return of the entire stock market at some given time, we can see that for each value of X, the response variable, or the individual stock’s rate of return, will increase by the value of the “beta” times however many percentage points the stock market increases by. But of course no regression model is perfect, so we also include a residual at the end of the equation to account for any unpredictable variation that the “beta” cannot explain.

	Ticker Symbol	**Intercept: “alpha”**	**Slope: “beta”**	**R^2**
Apple	AAPL	0.009	1.066	0.013
Google	GOOG	0.000	0.997	0.648
Merck	MRK	0.000	0.714	0.484
Johnson and Johnson	JNJ	0.000	0.677	0.502
Wal-Mart	WMT	0.001	0.519	0.285
Target	TGT	0.002	0.708	0.248

The ticker symbol is used to identify shares of a specific stock in a specific market while the intercept tells us the rate of return of an individual stock at some given time when the rate of return of the entire stock market (S&P 500) at that same time is equal to 0. The slope, or “beta”, tells us the percentage change in an individual stock given a 1% change in the rate of return of the entire stock market and, R^2 tells us what fraction of variability in the rate of return of an individual stock is predictable in terms of the rate of return of the entire stock market. The rates of return can be interpreted similarly to interest rates.

In light of my analysis, Wal-Mart has the lowest systematic risk because its beta value is the closest to 0, and Apple has the highest systematic risk because its beta value is above 1 and the greatest in magnitude. When the market goes up, the rate of return of Apple goes up the most, and when the market goes down, the rate of return of Apple falls the farthest. On the other end, the rate of return of Wal-Mart goes up the lowest and falls the lowest when there’s a change in the market.

When we reference the "beta" of a stock, this value refers to how volatile a stock is in comparison to the rest of the market. The specific value of the beta is the percentage change of a stock's return given that there was a 1% change in the market portfolio (typically, the S&P 500 index). In this sense, beta is an indicative measure of the systematic risk of a firm's stock. A stock typically has systematic and unsystematic risk; the systematic risk is risk that is related to the overall market whereas unsystematic risk is risk that is related to the individual firm. The reason beta indicates the systematic risk for a stock is because it indicates how the stock changes when the market changes. If a beta value is higher, this demonstrates a higher systematic risk because that stock reacts quite a bit based on the market; if a beta value is lower, this demonstrates a lower systematic risk because that stock doesn't react quite as much to the market.

To calculate the "beta" of a stock, we run a linear regression on the returns of the firm's stock (y-variable) versus the returns of the overall stock market (S&P 500; x-variable). In this regression model that represents each stock's rate of return, the beta is the slope; the slope represents the change in y for each 1 unit change in x; thus, for this model, the slope represents the change in the return of the firm's stock (y) for each 1% change in the overall stock market (S&P 500 returns; x). This is exactly what the slope represents and the exact definition of what the beta value is; thus, the beta is equivalent to the slope.

Ticker Symbol	**Intercept**	**Slope**	**R^2**
AAPL	0.009	1.066	0.013
GOOG	0.000	0.997	0.648
MRK	0.000	0.714	0.484
JNJ	0.000	0.677	0.502
WMT	0.001	0.519	0.285
TGT	0.002	0.708	0.248

The first column is the y-intercept of the linear regression model for each stock's rate of return (also known as the alpha), which indicates what the stock's returns would be if the returns of the overall stock market (S&P 500) were 0. The second column represents the slope of the linear regression model, which indicates the beta value of the stock, or the percentage change of a stock's return given that there was a 1% change in the market portfolio (S&P 500). The last column represents the r-squared value for each *stock, which indicates what fraction of the variation in the firm's stock is predictable in terms of the returns of the S&P 500, with values closer to 1 representing a larger proportion of systematic variation).*

Conclusion: In light of my analysis, the stock of the six listed with the lowest systematic risk is Walmart (WMT) because it has the smallest slope (which represents the lowest beta value) and the stock of the six listed with the highest systematic risk is Apple (AAPL) because it has the largest slope (which represents the highest beta value).

The "*beta*" (***β_1***) of a stock is known as the slope of the linear regression model, and it measures how much an individual stock's rate of return is influenced by the overall stock market's rate of return. The higher the value of beta is, the more susceptible the stock is to change when the overall stock market experiences changes, meaning that beta is an indicator of systematic risk. When the market portfolio experiences a change of 1%, beta indicates the percentage change of the individual stock's return. In order to calculate beta, we would need to fit a linear regression model for each individual stock against the S&P 500 Stock Index and identify the slope, or in other words, the change in the individual stock's return (*Y, the response variable*) for every 1% change in the overall stock market's return (*X, the predictor variable*).

Beta allows us insight into the volatility of the individual stocks, rendering a better estimate of which stocks tend to be riskier investments than others. From there, people can choose to invest in certain individual stocks presumably on how they foresee the performance of the market to be, and predict the magnitude of the changes in the upward and downward trends based on the magnitude of the specific beta value. When the overall stock market experiences changes, an individual stock with a beta closer to 0 suggests less drastic changes, and an individual stock with a beta closer to 1 suggests more drastic changes — the latter is usually much riskier to invest in, but can be rewarding if the investor is wanting a quicker return on their investment while the market is experiencing upward trends.

Company Name	Ticker Symbol	Intercept (β_0)	Slope (β_1)	R^2
Apple	**AAPL**	0.0092	1.066	0.013
Google	**GOOG**	0.0002	0.997	0.648
Merck	**MRK**	-0.00015	0.714	0.484
Johnson & Johnson	**JNJ**	-0.00002	0.677	0.502
Wal-Mart	**WMT**	0.0007	0.519	0.285
Target	**TGT**	0.0016	0.708	0.248

The table shown above provides the company name with its denoted stock ticker symbol, along with a y-intercept, slope, and r-squared value. The r-squared value answers the question of what proportion of the rate of return of an individual stock is predictable in terms of the rate of return of the entire stock market in that same time period, and the higher that number is, the more systematic it is. The slope, known as beta, and the intercept, known as alpha, can be set up as $\beta_1(X_t) + \beta_0 + e_t = Yt$ *, which means that in order to find the* Y *(the rate of return of an individual stock), we must multiply the beta by the market rate of return, and add the intercept (alpha) and the residual (difference between observed and expected value) for the individual stock in that time period.*

In light of our analysis, I believe Wal-Mart stock has the lowest systematic risk due to its small beta value of .519, out of the six stocks in this dataset. When interpreted, we can assume that Wal-Mart stock has low systematic risk because the company provides goods that are necessities and is less affected by the trends of the overall stock market, versus other consumer goods (that are deemed normal goods, as opposed to inferior goods). This would mean that the stock with the highest beta value to be the stock with the greatest systematic risk, which we observe in this dataset to be Apple. Apple appears to be highly influenced by the market with a beta of 1.066, as it has consumer goods that are considered more luxurious wants as opposed to dire needs, so its stock will vary greatly depending on how the overall market is performing.