University of Chicago
Booth School of Business

41000: Business Statistics, Autumn 2021: Homework Assignment 1. Due in Week 3

Problem 1: Probability Answer the following statements TRUE or FALSE, providing a succinct explanation of your reasoning.

1. If the odds in favor of A are 3:5 then $\text{P}(A) = 0.4$.

   FALSE $P(A) = \frac{3}{3+5} = 0.375$

2. You roll two fair three-sided dice.

   The probability the two dice show the same number is 1/4.

   FALSE $P = \frac{1}{3} \times \frac{1}{3} \times 3 = \frac{1}{3}$.

3. If events A and B are independent and $\text{P}(A) > 0$ and $\text{P}(B) > 0$, then $\text{P}(A \text{ and } B) > 0$.

   TRUE $P(A \text{ and } B) = P(A) \times P(B) > 0$.

4. If $\text{P}(A \text{ and } B) \geq 0.5$ then $P(A) \leq 0.5$.

   FALSE Suppose A B are independent and $P(A) = 1$, $P(B) = 0.5$. $P(A \text{ and } B) \geq 0.5$

5. If two random variables have non-zero correlation, then they must be dependent.

   TRUE If two random variables are independent, they always have 0 correlation. Therefore if correlation is not 0, they cannot be independent.

6. If two random variables have zero correlation, then they must be independent.

FALSE Suppose X takes value in $\{-1, 0, 1\}$ with probability $\frac{1}{3}$ for each. $Y = X^2$, we have

$$X = \begin{cases} -1 & \textit{with prob } 1/3 \\ 0 & \textit{with prob } 1/3 \\ 1 & \textit{with prob } 1/3 \end{cases}, \quad Y = \begin{cases} 1 & \textit{with prob } 2/3 \\ 0 & \textit{with prob } 1/3 \end{cases} \tag{1}$$

It's easy to find that $Cov(X, Y) = 0$. But Y is a function of X, so they are not independent.

7. If two random variables are independent, then the correlation between them must be zero.

   TRUE Independence implies uncorrelation, thus correlation is 0.

Problem 2: Expectation and Strategy

An oil company wants to drill in a new location. A preliminary geological study suggests that there is a 20% chance of finding a small amount of oil, a 50% chance of a moderate amount and a 30% chance of a large amount of oil.

The company has a choice of either a standard drill that simply burrows deep into the earth or a more sophisticated drill that is capable of horizontal drilling and can therefore extract more but is far more expensive. The following table provides the payoff table in millions of dollars under different states of the world and drilling conditions

Oil	small	moderate	large
Standard Drilling	20	30	40
Horizontal Drilling	-20	40	80

Find the following

(a) The mean and variance of the payoffs for the two different strategies

(b) The strategy that maximizes their expected payoff

(c) Briefly discuss how the variance of the payoffs would affect your decision if you were risk averse

(d) How much are you willing to pay for a geological evaluation that would tell you with certainty the quantity of oil at the site prior to drilling?

Solution

(a) Using the plug-in rule, we get

$$\begin{aligned} E(Standard) &= 31 \\ E(Horizontal) &= 40 \\ Var(Standard) &= E(X^2) - [E(X)]^2 \\ &= 1010 - 961 \\ &= 49 \\ Var(Horizontal) &= E(X^2) - [E(X)]^2 \\ &= 2800 - 1600 \\ &= 1200 \end{aligned}$$

(b) Based on the above, the horizontal drilling maximizes the expected payoff.

(c) Note that the strategy with higher expected payoff has a substantially higher variance. Thus, a risk averse person may settle for a strategy with lower expected payoff and lower variance (standard drilling) while a risk seeking person will chose the horizontal drilling.

(d) If the geological evaluation tells us with certainty the type of oil, then we will be able to chose the strategy which maximizes our payoff under different types of oil. Thus, the amount one should be willing to pay is the expected payoff under this revised schedule of payoff minus the maximum payoff under the current schedule. Given this

$$WTP = 0.2 \times 20 + 0.5 \times 40 + 0.3 \times 80 - 40 = 48 - 40 = 8$$

Therefore, once should be willing to pay \$8 million for this information.

Problem 3: Normal Distribution

After Facebook's earnings announcement we have the following distribution of returns. First, the stock beats earnings expectations 75% of the time, and the other 25% of the time earnings are in line or disappoint. Second, when the stock beats earnings, the probability distribution of percent changes is normal with a mean of 10% with a standard deviation of 5% and, when the stock misses earnings, a normal with a mean of −5% and a standard deviation of 8%, respectively.

(a) Ahead of the earnings announcement, what is the probability that Facebook stock will have a return greater than 5%?

(b) Do you get the same answer for the probability that it drops at least 5%?

Solution *We have the following information:*

$$P(\textit{Beat Earnings}) = 0.75 \text{ and } P(\textit{Not Beat Earnings}) = 0.25$$

together with the following probabilities distributions

$$R_{Beat} \sim N(0.10, 0.05^2) \text{ and } R_{Not} \sim N(-0.05, 0.08^2)$$

We want to find the probability that Facebook stock will have a return greater than 5%:

$$\begin{aligned} &P(\textit{Beat}) \times P(R_{Beat} > 0.05) + P(\textit{Not}) \times P(R_{Not} > 0.05) \\ &= 0.75 \times (1 - F_{Beat}(0.05)) + 0.25 \times (1 - F_{Not}(0.05)) \end{aligned}$$

```
% > 0.75*(1-pnorm(0.05,0.1,0.05)) + 0.25*(1-pnorm(0.05,-0.05,0.08))
% [1] 0.657421
```

Therefore, the probability that Facebook stock beats a 5% return is 65.7%.

Similarly, for dropping at least 5%, we want:

$$\begin{aligned} &P(\textit{Beat}) \times P(R_{Beat} < -0.05) + P(\textit{Not}) \times P(R_{Not} < -0.05) \\ &= 0.75 \times (F_{Beat}(-0.05)) + 0.25 \times (F_{Not}(-0.05)) \end{aligned}$$

We can compute this is in R by the following commands:

```
> 0.75*pnorm(-0.05,0.1,0.05) + 0.25*pnorm(-0.05,-0.05,0.08)
[1] 0.1260124
```

Therefore, the probability that Facebook stock drops at least 5% is 12.6%.

Problem 4: Binomial Distribution

The Downhill Manufacturing company produces snowboards. The average life of their product is 10 years. A snowboard is considered defective if its life is less than 5 years. The distribution is approximately normal with a standard deviation for the life of a board of 3 years.

(a) What's the probability of a snowboard being defective?

(b) In a shipment of 120 snowboards, what is the probability that the number of defective boards is greater than 10?

[You can use `R` and simulation with `rbinom, rnorm` as an alternative]

Solution *Here we are given the following information:* $L \sim N(10, 3^2)$*. We want to find the probability of a snowboard being defective:*

$$P(L < 5) = F_L(5)$$

This is simply the cumulative distribution function. We can find the solution by using the following R commands:

```
> pnorm(5,mean=10,sd=3)
[1] 0.04779035
```

Thus, the probability a snowboard is considered defective is 4.78%.

Out of a shipment of 120 snowboards, the distribution of the number of defective boards is a Binomial Distribution parameterized as follows:

$$N_{def} \sim Bin(120, 0.04779035)$$

We want to find the ***exact*** *probability:*

$$P(N_{def} > 10)$$

```
>  1 - pbinom(10,120,0.0478)
[1] 0.02914742
```

For problem (b), you can also calculate probability by normal approximation. Note than when n *is large, binomial distribution* (n, p) *is close to normal distribution* $N(np, np(1-p))$*. In this case* $np = 120 \times 0.04779035 = 5.7348$, $\sqrt{np(1-p)} = 2.3368$*. So the* ***approximation*** *by normal distribution is*

```
>  1 - pnorm(10, mean = 5.7348, sd = 2.3368)
[1] 0.03398308
```

2. **Problem 5: Portfolio ETF**

You want to build a portfolio of exchange traded funds (ETFs) for your retirement strategy. You?re thinking of whether to invest in growth or value stocks, or maybe a combination of both. Vanguard has two ETFs, one for growth (VUG) and one for value (VTV). The R script hwk1.R script on the course web-page let's you download historical price data.

- Plot the historical price series for VUG vs VTV.
- Calculate the means and standard deviations of both ETFs.
- Calculate their covariance.
- Suppose you decide on a portfolio that is a 50 / 50 split. Calculate the new mean and variance of your portfolio.
- Which portfolio best suits you?
- What's the probability that growth (VUG) will beat value (VTV) in the future?

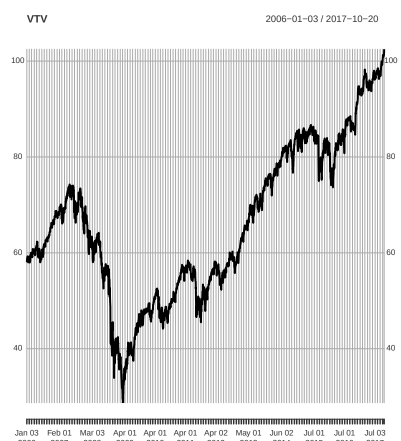
Hint: you will find the following formulas useful. Let P denote the return on your portfolio which is a weighted combination $P = aX + bY$. Then

$$E(P) = aE(X) + bE(Y) \tag{2}$$

$$Var(P) = a^2 Var(X) + b^2 Var(Y) + 2abCov(X, Y) \tag{3}$$

where $Cov(X, Y)$ is the covariance for$X$and Y .

Solution




```
> mean(VUG.return)
[1] 0.000391497
> mean(VTV.return)
[1] 0.0002707707
> sd(VUG.return)
[1] 0.01175392
> sd(VTV.return)
[1] 0.01248417
> cov(VUG.return, VTV.return)
              daily.returns
daily.returns    0.00013654
```

```
# portfolio mean and variance
> portfolio_mean_1 = 0.5 * mean(VUG) +
0.5 * mean(VTV)
> portfolio_var_1 = 0.5^2 * var(VUG) +
0.5^2 * var(VTV) + 2 * 0.5 * 0.5 * cov(VUG, VTV)
> portfolio_sd_1 = sqrt(portfolio_var_1)
```

```
> # Suppose you decide on a 50-50 split portfolio, calculate mean and variance of your portfolio.
> portfolio_return = 0.5 * VUG.return + 0.5 * VTV.return
> hist(portfolio_return, breaks = 50)
>
> # compute mean and variance of return of the portfolio
> portfolio_mean_1 = 0.5 * mean(VUG.return) + 0.5 * mean(VTV.return)
> portfolio_var_1 = 0.5^2 * var(VUG.return) + 0.5^2 * var(VTV.return) + 2 * 0.5 * 0.5 * cov(VUG.return,
> portfolio_sd_1 = sqrt(portfolio_var_1)

> print(portfolio_mean_1)
[1] 0.0003311338
> print(portfolio_var_1)
              daily.returns
daily.returns  0.0001417723
> print(portfolio_sd_1)
              daily.returns
daily.returns    0.01190682
```

You may say VUG or VTV best suits you, as long as you give an explanation such as you are risk aversion or not.

Now we consider mean and variance of VUG - VTV,

```
> portfolio_mean_2 = 1.0 * mean(VUG.return) - 1.0  * mean(VTV.return)
> portfolio_var_2 = 1.0^2 * var(VUG.return) + (-1.0)^2 * var(VTV.return) - 2 * cov(VUG.return, VTV.retu
> portfolio_sd_2 = sqrt(portfolio_var_2)
>
> portfolio_mean_2
[1] 0.0001207263
> portfolio_sd_2
              daily.returns
daily.returns   0.004574826
```

Therefore the probability that VUG - VTV < 0 *is*

```
> 1 - pnorm(0, portfolio_mean_2, portfolio_sd_2)
[1] 0.5105266
```

Problem 6: Google

Visitors to your website are asked to answer a single survey http://www.google.com question before they get access to the content on the page. Among all of the users, there are two categories

1. Random Clicker (RC)
2. Truthful Clicker (TC)

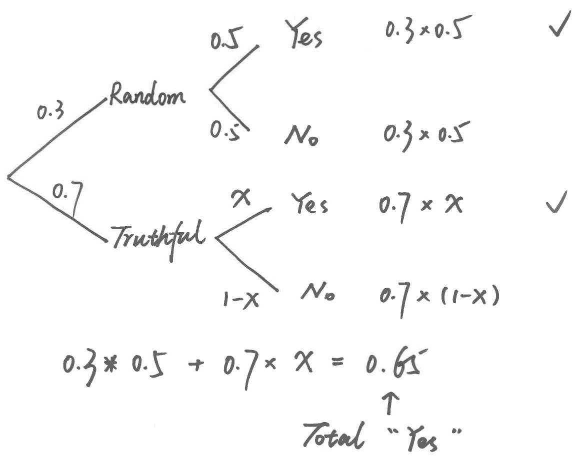
There are two possible answers to the survey: yes and no.

Random clickers would click either one with equal probability. You are also giving the information that the expected fraction of random clickers is 0.3.

After a trial period, you get the following survey results. 65% said Yes and 35% said No.

(a) How many people people who are truthful clickers answered yes?

Solution *Given the information that the expected fraction of random clickers is 0.3,*



$$P(RC) = 0.3 \text{ and } P(TC) = 0.7$$

Conditioning on a random clickers, he would click either one with equal probability.

$$P(Yes \mid RC) = P(No \mid RC) = 0.5$$

By the survey results, we know the proportion response of "Yes" and "No".

$$P(Yes) = 0.65 \text{ and } P(No) = 0.35$$

Therefore, the probability that a truthful clicker answer "Yes" is,

$$\begin{aligned} P(Yes \mid TC) &= \frac{P(Yes) - P(Yes \mid RC) * P(RC)}{P(TC)} \\ &= \frac{0.65 - 0.5 * 0.3}{0.7} \\ &= 0.71 \end{aligned}$$

Notice that, we use law of total probability in the first equation.