Statistic Midterm

Spring 2018

This is a closed-book, closed-notes exam. You may use any calculator.

Please answer all problems in the space provided on the exam.

Read each question carefully and clearly present your answers.

Honor Code Pledge: "I pledge my honor that I have not violated the University Honor Code during this examination."

Signed: ____

Name: ____

Here are some useful formulas:

- $E(aX + bY) = aE(X) + bE(Y)$
- $Var(aX + bY) = a^2Var(X) + b^2Var(Y) + 2ab \times Cov(X, Y)$
- The standard error of $\bar{X}$ is defined as $s_{\bar{X}} = \sqrt{\frac{s_X^2}{n}}$ where s_X^2 denotes the sample variance of X.
- The standard error for the difference in the averages between groups a and b is defined as:

  $$s_{(\bar{X}_a - \bar{X}_b)} = \sqrt{\frac{s_a^2}{n_a} + \frac{s_b^2}{n_b}}$$

  where s_a^2 denotes the sample variance of group a and n_a the number of observations in group a.
- The standard error for a proportion is defined by: $s_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

Problem 1 (10 points each)

In recent years the NFL changed the extra point rule and moved the distance of the field goal attempt to 33 yards (as compared to the previous distance of 19 yards). A successful kick gives the team 1 extra point. Of course, teams have the option to try for 2 extra points by attempting a touchdown starting from the 2-yard line. Currently, the probability of successful kicks from 33 yards is 0.93 whereas the probability of getting a touchdown from the 2-yard line is 0.48.

1. As a general rule, over the course of a season, what should you do as a team? Go for 1 or 2 extra points? Justify your answer.

   Goal: maximizing expected points.

   $$X_1 = \begin{cases} 1 \text{ wp } 0.93 \\ 0 \text{ wp } 0.07 \end{cases}, \quad X_2 = \begin{cases} 2 \text{ wp } 0.48 \\ 0 \text{ wp } 0.52 \end{cases}.$$

   So

   $$E[X_1] = 0.93, \; E[X_2] = 0.96.$$

   Go for touchdown.

2. Now consider a situation where your team just scored a touchdown to tie the game. There is only 1 second left on the clock meaning that your extra point conversion is the last play of the game. What should you do now? Go for 1 or 2 extra points? Justify your answer and contrast to your answer to the previous question.

   Goal: maximizing winning probability.

   $$X_1 = \begin{cases} win \text{ wp } 0.93 \\ tie \text{ wp } 0.07 \end{cases}, \quad X_2 = \begin{cases} win \text{ wp } 0.48 \\ tie \text{ wp } 0.52 \end{cases}.$$

   So going for 1 extra point gives higher winning probability.

Problem 2 (10 points each)

The breathalyzers carried by cops have a false positive rate of 5%. However, they are perfect at detecting drunk drivers, i.e., the false negative probability is zero. Previous studies have concluded that on a Saturday night in Chicago, 1 in 500 drivers are drunk. Mayor Emmanuel decided to implement a policy of no-refusal, random traffic stops on Saturday's. That means that random cars will be pulled over and their drivers tested.

1. On your way home on a Saturday night, your Uber driver gets randomly stopped. She tests positive for being drunk. Given this information, what is the probability she is drunk?

Joint Probability	Drunk	Not Drunk
Test Positive	$1 \times 1/500$	$0.05 \times 499/500$
Test Negative	$0 \times 1/500$	$0.95 \times 499/500$

By Bayes' Rule,

$$\begin{aligned}
\Pr(drunk|positive) &= \frac{\Pr(drunk \,\&\, positive)}{\Pr(positive)} \\
&= \frac{\Pr(drunk)\Pr(positive|drunk)}{\Pr(positive|drunk)\Pr(drunk) + \Pr(positive|not\ drunk)\Pr(not\ drunk)} \\
&= \frac{1/500 \times 100\%}{100\% \times 1/500 + 5\% \times 499/500} \\
&= 0.038
\end{aligned}$$

2. Based on the numbers provided and your computation in question (1), do you agree or disagree with the Mayor's decision? Why or why not?

I disagree. Most of the positive results will be false positive.

Problem 3 (5 points each)

A construction company needs to complete a project within 11 weeks, or they will incur significant cost overruns, including penalties due to the client. The manager of the company has assessed that the project will take between 10 and 14 weeks to complete. The manager has also estimated the probability of each possible outcome:

Weeks to complete	Probability
10	0.075
11	0.65
12	0.2
13	0.05
14	0.025

1. What is the probability of completing the project on time?

   $\Pr(X \leq 11) = \Pr(X = 10) + \Pr(X = 11) = 0.725$

2. What is the expected value of the time to complete the project?

   $E[X] = \sum_{i=1}^{5} x_i \Pr(X = x_i) = 11.3$

3. The company must pay a penalty of \$5,000 for **every additional week (past 11 weeks) that they work**, plus a additional \$50,000 penalty if the work takes 14 weeks to complete. What is the expected value of the penalty incurred?

   $E[Y] = 5000 \times 0.2 + 10000 \times 0.05 + 65000 \times 0.025 = 3125$

4. Suppose you're the engineer in charge of bidding for the projects in this company (i.e. estimating the total cost of the job, plus overhead, potential overrun costs and profit). How would you use the information from the previous question to price this job?

   The price needs to be at least ($total\ cost + 3125$).

Problem 4 (10 points each)

I am trying to build a portfolio composed of SP500 and Bonds.
Assume $SP500 \sim N(11, 19^2)$ and $Bonds \sim N(4, 6^2)$. In addition, the covariance between SP500 and Bonds is -22.6.

1. Which portfolio is better: a 50-50 split between the two assets or a 60-40 split between SP500 and Bonds? Justify your criteria for comparison.

   $X = 0.5\ SP500 + 0.5\ Bonds$

   $$E[X] = 0.5 \times 11 + 0.5 \times 4 = 7.5$$
   $$Var[X] = 0.5^2 \times 19^2 + 0.5^2 \times 6^2 - 2 \times 0.5 \times 0.5 \times 22.6 = 87.95$$

   $Y = 0.6\ SP500 + 0.4\ Bonds$

   $$E[Y] = 0.6 \times 11 + 0.4 \times 4 = 8.2$$
   $$Var[Y] = 0.6^2 \times 19^2 + 0.4^2 \times 6^2 - 2 \times 0.6 \times 0.4 \times 22.6 = 124.87$$

   Then, Sharpe Ratio:

   $$SR_X = 0.7997,\ SR_Y = 0.7338.$$

   So 50-50 split is preferable. Although 50-50 split has lower expected value, it has much smaller risk than 60-40 split does.

2. Which of the two portfolios considered in item (1) has the largest probability of delivering a negative return?

   $$\Pr(X < 0) = \Pr\left(\frac{X - 7.5}{\sqrt{87.95}} < \frac{-7.5}{\sqrt{87.95}}\right) = \Pr(N(0,1) < -0.7997)$$

   $$\Pr(Y < 0) = \Pr\left(\frac{Y - 8.2}{\sqrt{124.87}} < \frac{-8.2}{\sqrt{124.87}}\right) = \Pr(N(0,1) < -0.7338)$$

   So 60-40 split has a large negative probability, simply looking at the standard normal density and area under curve.

Problem 5

I am trying to get into the insurance business. As a starter, I am going to sell the 2019 Booth MBA class an unemployment insurance policy that will pay a student \$100k if they don't get a job by the end of the program (by graduation time). From years and years of observing this program, I figured that the probability that a student does not get a job by the end of program is 0.5%. Based on this number I decided to sell each policy for \$850.

1. From the perspective of my insurance business, is that a good price? Justify your answer. (5 points)

   Let X be the profit.

   $$X = \begin{cases} 850 & \text{wp } 99.5\% \\ 850 - 100000 & \text{wp } 0.5\% \end{cases}$$

   $E[X] = 350$, so the expected profit is positive..

2. It turns out that I was able to sell this policy to 1,000 students in the 2019 class! What is my expected profit at graduation time? (5 points)

   Let X_i has the same distribution as X for $i = 1, \ldots, 1000$. Selling 1000 policies has expected profit

   $$E[Y] = E\left[\sum_{i=1}^{1000} X_i\right] = 350000.$$

3. The standard deviation for the payout of one policy is approximately \$7k. What is the probability (approximately) that I will lose money by graduation time? (hint: use the normal approximation...). (10 points)

$Y = \sum_{i=1}^{1000} X_i \sim N(350000, 7000^2 \times 1000)$, so

$$\begin{aligned} \Pr(Y < 0) &= \Pr\left(\frac{Y - 350000}{\sqrt{7000^2 \times 1000}} < \frac{-350000}{\sqrt{7000^2 \times 1000}}\right) \\ &= \Pr(N(0,1) < -1.5811) \\ &= 0.0569 \end{aligned}$$

Note it is totally ok to leave the answer as $\Pr(N(0,1) < -1.5811)$ or $\Pr(Z < -1.5811)$

Problem 6 (10 points each)

Amy, a former Booth MBA student, now the manager of Windy City Trading Strategies, is trying to convince you to invest in the fund she runs. Based on the last two years of monthly returns, Amy provides you with the following summary of the performance of the fund:

Sample average	2% per month
Sample std. deviation	2% per month

1. Based on these results, if you decide to invest in her fund, what is the probability you will lose money next month?

   $\Pr(X < 0) = \Pr\left(\frac{X-2\%}{2\%} < -1\right) = \Pr(N(0,1) < -1) = 0.1587$

2. Before deciding to invest, you remember that Carlos told you that the monthly sharpe ratio (mean/std. dev) of the market is 0.4. Is the sharpe ratio of Amy's fund for sure better than that? (and "for sure" I mean, with 95% confidence). Hint: first build a 95% confidence interval for the mean of Amy's fund.

   $$s_{\bar{X}} = \frac{2\%}{\sqrt{24}} = 0.0041$$

   So 95% confidence interval of the mean is $[\bar{X} - 2s_{\bar{X}}, \bar{X} + 2s_{\bar{X}}] = [0.0118, 0.0282]$, and 95% confidence interval for Sharpe ratio is $[0.59, 1.41]$. So we should be fairly confident that Amy's fund performs better than the market, because 0.4 is smaller than $[0.59, 1.41]$

Problem 7 (20 points)

Your cognitive capacity is significantly reduced when your smartphone is within reach, even if it's off. That's the claim from a new study from the McCombs School of Business at The University of Texas at Austin.

Marketing Assistant Professor Adrian Ward conducted experiments with nearly 600 smartphone users in an attempt to measure, for the first time, how well people can complete tasks when they have their smartphones nearby even when they're not using them.

In one experiment, the researchers asked study participants to sit at a computer and perform a task that required full concentration in order to succeed. The tests were geared to measure participants' available cognitive capacity – that is, the brain's ability to hold and process data at any given time. Before beginning, participants were randomly instructed to place their smartphones either on the desk face down (group 1; 200 participants), in their pockets (group 2; 200 participants), or in another room (group 3; 200 participants). All participants were instructed to turn their phones to silent.

The researchers found that participants with their phones in another room significantly outperformed those with their phones on the desk, and they also slightly outperformed those participants who had kept their phones in the pocket.

Based on the results from the experiment (listed below), do you agree with the statement from the above (bold) paragraph? Justify your answer.

	group 1	group 2	group 3
Successful tasks	90	130	150

Sample mean: $\bar{X}_1 = .45$, $\bar{X}_2 = .65$, $\bar{X}_3 = .75$, so $\bar{X}_3 - \bar{X}_1 = 0.2$, and $\bar{X}_3 - \bar{X}_2 = 0.1$

Standard error: $s_{\bar{X}_3-\bar{X}_1} = \sqrt{\frac{\bar{X}_3(1-\bar{X}_3)}{200} + \frac{\bar{X}_1(1-\bar{X}_1)}{200}} = 0.0466$, $s_{\bar{X}_3-\bar{X}_2} = \sqrt{\frac{\bar{X}_3(1-\bar{X}_3)}{200} + \frac{\bar{X}_2(1-\bar{X}_2)}{200}} = 0.0456$.

Confidence interval: For $\bar{X}_3 - \bar{X}_1$, CI is $[\bar{X}_3 - \bar{X}_1 - 1.96 s_{\bar{X}_3-\bar{X}_1}, \bar{X}_3 - \bar{X}_1 + 1.96 s_{\bar{X}_3-\bar{X}_1}]$, i.e. $[0.1086, 0.2914]$. Similarly, for $\bar{X}_3 - \bar{X}_2$, CI is $[0.0107, 0.1873]$.

The implication is that, for both experiment designs, the effect of putting phones in another room is significant.

Problem 8 (3 points each)

Assume the model: $Y = 5 - 2X + \varepsilon, \quad \varepsilon \sim \ N(0, 9^2)$

1. What is expected value of Y if $X = 1$, i.e, $E[Y|X = 1]$?

   (a) 5

   (b) 3 - Correct

   (c) 4

   (d) 6

2. What is the $Var[Y|X = 0]$?

   (a) 9

   (b) 81 - Correct

   (c) 3

   (d) 6

3. What is the $Pr(Y > 10)$, given $X = 2$?

   (a) 13%

   (b) 68%

   (c) 16% - Correct

   (d) 2.5%

4. What is the $Pr(-6 < Y < 21)$, given $X = 1$?

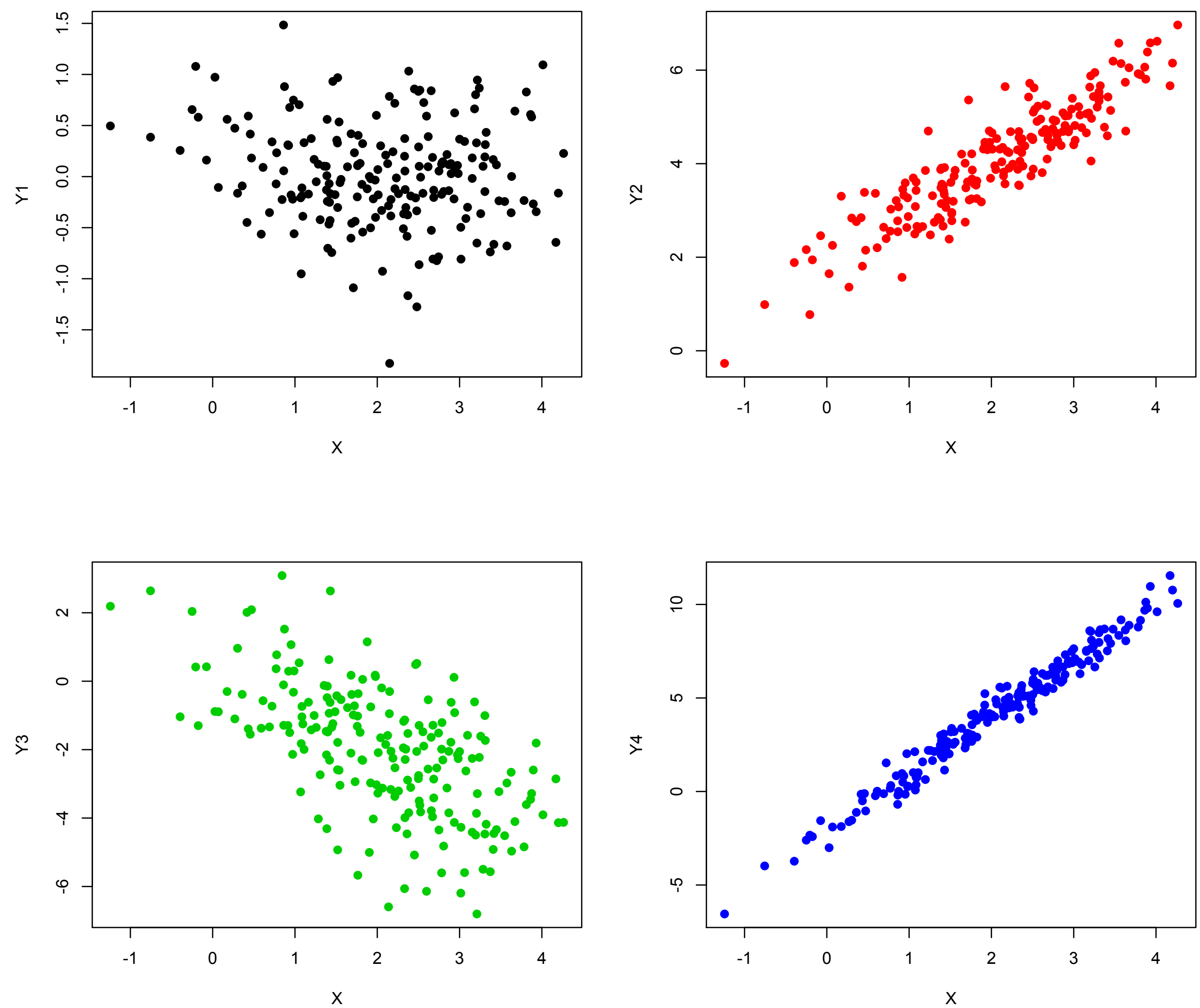
   (a) 5%

   (b) 16%

   (c) 81.5% - Correct

   (d) 34%

Problem 9 (5 points each)



In the above scatterplots, four different variables $Y1$, $Y2$, $Y3$ and $Y4$ were regressed onto the same X (in all four scatterplot we have the exact same $n = 200$ values for X). Carefully examine the plots and answer the questions below:

1. The estimates of the slope (b_1) for the four regressions are listed below. Which one belongs to the regression of $Y3$ onto X?

   (a) -0.054

   (b) 3.020

   (c) 1.031

   (d) -1.094 - Correct

2. The estimates of the slope (b_1) for the four regressions are listed below. Which one belongs to the regression of $Y1$ into X?

   (a) -0.054 - Correct

   (b) 3.020

   (c) 1.031

   (d) -1.094

3. The estimates of the intercept (b_0) for the four regressions are listed below. Which one belongs to the regression of $Y4$ into X?

   (a) -2.044 - Correct

   (b) 0.212

   (c) 0.135

   (d) 1.951

4. The standard errors of the regressions (s) are listed below. Which one belongs to $Y3$?

   (a) 0.52

   (b) 0.49

   (c) 1.51 - Correct

   (d) 0.51

5. The R^2 of the regressions are listed below. Which one belongs to $Y2$?

   (a) 37%

   (b) 81% - Correct

   (c) 97 %

   (d) 1 %

6. What is the correlation between Y_1 and X?

   (a) -0.12 - Correct

   (b) 0.91

   (c) 0.97

   (d) -0.60

7. Using all the information provided so far, give a rough approximation of the 95% prediction interval for $Y4$ when $X = 3$.

   Approximately, $E[Y_4|X = 3] = 7$ and $\sigma = 1$, so $CI = [5, 9]$.

8. Give an approximation for $Pr(Y_3 > 0|X = 2)$

   Approximately, $E[Y_3|X = 2] = -2$ and $\sigma = 1.5$, so $Y_3|X = 2 \sim N(-2, 1.5)$.

$$\begin{aligned}\Pr(Y_3 > 0|X = 2) &= \Pr\left(\frac{Y_3 + 2}{1.5} > \frac{2}{1.5}\middle| X = 2\right) \\ &= \Pr(N(0,1) > 2/1.5) \\ &= 0.0912\end{aligned}$$

Problem 10 (5 points each)

In trying to understand how good of an investor Warren Buffet is, I collected data on the annual returns (in percentage terms) of his entire portfolio and decided to compare it to the market (again, in percentage terms). In doing so I ran the following regression:

$$WB = \alpha + \beta Market + \epsilon \qquad \epsilon \sim N(0, \sigma^2)$$

where WB refers to the annual returns on Buffet's portfolio and $Market$ refers to the annual returns on the U.S. stock market. The result of this regression is in the table below:

Regression Statistics	
Multiple R	0.628697
R Square	0.39526
Adjusted R	0.382113
Standard E	11.4114
Observatio	48

ANOVA

	df	*SS*	*MS*	*F*	*ignificance F*
Regression	1	3915.157401	3915.157	30.06571	1.71E-06
Residual	46	5990.121766	130.22		
Total	47	9905.279167			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	14.94673					
Market	0.524621					

1. What is the expected value for the returns on Buffet's portfolio when the market is up 10%?

   $E[WB|Market = 10\%] = 14.95 + 0.525 \times 10 = 20.2$

2. What is the standard deviation for the returns on Buffet's portfolio when the market is up 10%?

   $\hat{\sigma} = \sqrt{\frac{1}{46}\sum_i \hat{e}_i^2} = \sqrt{\frac{1}{46}SSR} = \sqrt{130.22} = 11.41$. This is also the Standard E in the table.

3. Give a 99% prediction range for the returns on Buffet's portfolio when the market is up 10%?

   Approximately, $X \sim N(20, 11.4^2)$. Suppose the 99% confidence interval is $[20 - 3 \times 11.4, 20 + 3 \times 11.4]$.

4. Provide an interpretation for both the intercept and the slope of this regression. What do these quantities tell us about Warren Buffet's performance, relative to the U.S. stock market?

   Intercept: When the market return is zero, Warren Buffet has 14.95% return on average.

   Slope: When the market return is 1% higher, Warren Buffet has 0.5% higher return on average.

   The coefficient tells us on average Warren Buffet beats the market, because $14.96 + 0.52Market > Market$ (unless $Market > 31.17$, which is rare).