

Name: OUTLINE SOLUTIONS

University of Chicago  
Graduate School of Business  
Business 41000: Business Statistics

**Special Notes:**

1. This is a closed-book exam. You may use an  $8 \times 11$  piece of paper for the formulas.
2. Throughout this paper,  $N(\mu, \sigma^2)$  will denote a normal distribution with mean  $\mu$  and variance  $\sigma^2$ .
3. This is a 2 hr exam.

**Honor Code:** By signing my name below, I pledge my honor that I have not violated the Booth Honor Code during this examination.

**Signature:**

**Problem A. True or False:** Please Explain your answers in detail. Partial credit will be given (50 points)

1. If the sample covariance between two variables is one, then there must be a strong linear relationship between the variables

*False.*  $Cov(X, Y) = Corr(X, Y)\sqrt{Var(X) * Var(Y)}$ , so knowing  $Cov(X, Y) = 1$  doesn't inform you about the correlation

2. If  $X$  and  $Y$  are independent random variables, then  $Var(2X - Y) = 2Var(X) - Var(Y)$ .

*False.*  $Var(2X - Y) = 4Var(X) + Var(Y)$

3. For any normal random variable,  $X$ , we have  $\mathbb{P}(\mu - \sigma < X < \mu + \sigma) = 0.64$ .  
[Hint: You may use  $\text{pnorm}(1) = 0.841$ ]

*False.*  $\text{pnorm}(1) = 0.841$  and  $2 \times \text{pnorm}(1) - 1 = 0.6828 > 0.64$ .

4. The  $p$ -value is the probability that the Null hypothesis is true.

*False.* The  $p$ -value is the probability of the observing something more extreme than the observed sample assuming the null hypothesis is true.

5. Suppose that you toss a fair coin with probability 0.5 a head. The probability of getting five heads in a row is less than three percent.

*False.* Given a fair coin with 50% probability for a head,  $0.5^5 = 0.03125 > 0.03$

6. Arsenal are playing Burnley at home in an English Premier League (EPL) game this weekend. They are favourites to win. They have a Poisson distribution for the number of goals they will score with a mean rate of 2.5 per game. Given this, the odds of Arsenal scoring at least two goals is greater than 50%.

*True.* Here  $S \sim \text{Pois}(2.5)$ . We need  $P(S \geq 2) = 1 - P(S \leq 1) = 1 - 0.2873 = 0.7127$ . In R we have  $\text{ppois}(1, 2.5) = 0.28729$ .

7. Suppose that the annual returns for Facebook stock are normally distributed with a mean of 15% and a standard deviation of 20%. The probability that Facebook has returns greater than 10% for next year is 60%

*True.* Let  $R$  denote returns. Convert to standard normal and compute

$$\mathbb{P}\left(\frac{R - \mu}{\sigma} > \frac{0.1 - \mu}{\sigma}\right) = \mathbb{P}(Z > -0.25) = 0.5987$$

where  $\mu = 0.15$  and  $\sigma = 0.2$ .

8. LeBron James makes 85% of his free throw attempts and 50% of his regular shots from the field (field goals). Suppose that each shot is independent of the others. He takes 20 field goals and 10 free throws in a typical game. He gets one point for each free throw and two points for each field goal assuming no 3-point shots. The number of points he expects to score in a game is 28.5.

*True.*

$$E(\text{points}) = 2 * E(\#FGmade) + 1 * E(\#FTmade) = 2 \times 20 \times 0.5 + 1 \times 10 \times 0.85 = 28.5$$

*This holds even if FG and FT are dependent.*

9. Given a random sample of 2000 voters, 800 say they will vote for Hillary Clinton in the 2016 US Presidential Election. At the 95% level, I can reject the null hypothesis that Hillary has an evens chance of winning the election.

*True.* The  $t$ -statistic is  $T = (\hat{p} - 0.5) / \sqrt{\frac{\hat{p} \times (1 - \hat{p})}{n}} = -9.13 < \text{where } qnorm(0.025) = -1.96$ .

10. The Central Limit Theorem states that the distribution of a sample mean is approximately Normal.

*True.* The standard deviation of the sample mean is  $\frac{\sigma}{\sqrt{n}}$ . The CLT has (approximately)  $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$ , where  $\mu$  and  $\sigma^2$  are the population true value.

**Problem B.** (20 points)

The Chicago Cubs are having a great season. So far they've won 72 out of the 100 games played so far. You also have the expert opinion of Bob the sports analysis. He tells you that he thinks the Cubs will win. Historically his predictions have a 60% chance of coming true.

- Calculate the probability that the Cubs will win given Bob's prediction
- Suppose you now learn that it's a home game and that the Cubs win 60% of their games at Wrigley field.

What's your updated probability that the Cubs will win their game?

1. *First we have*

$$P(\text{Win}|\text{Bob}) = \frac{0.72 \times 0.6}{0.72 \times 0.6 + 0.28 \times 0.4} = 0.79$$

2. *Learning the new information, we use the previous posterior as a prior for the next Bayes update*

$$P(\text{Win}|\text{home}, \text{Bob}) = \frac{0.79 \times 0.60}{0.79 \times 0.60 + 0.21 \times 0.40} = 0.85$$

*it's highly likely that the Cubs will win!*

**Problem C.** (20 points)

You are given the following R output for daily returns data on Apple (AAPL) and Google's (GOOG) stock from January 2005 to the end of December 2013.

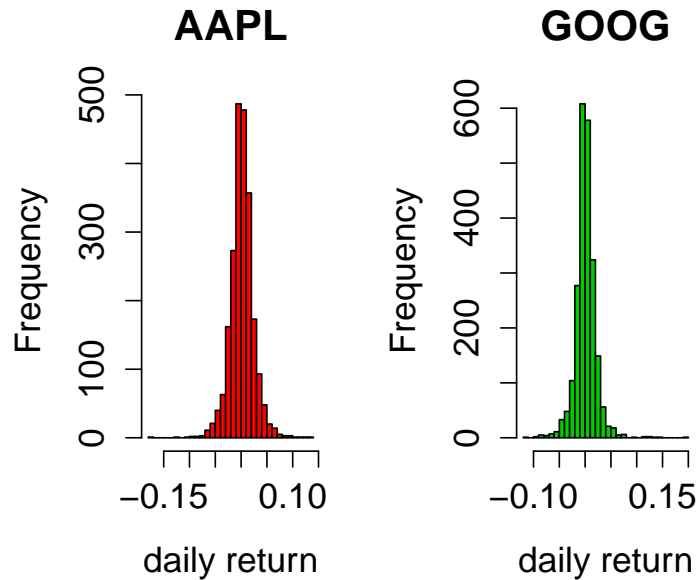


Figure 1: Apple vs Google

```
> summary(aapl)
      Min.   1st Qu.   Median     Mean   3rd Qu.   Max.
-0.179200 -0.010400  0.001304  0.001544  0.013820  0.139000
> summary(goog)
      Min.   1st Qu.   Median     Mean   3rd Qu.   Max
-0.1161000 -0.0082710  0.0004700  0.0009843  0.0104900  0.1999000

> skewness(aapl)
[1] -0.04116922
> skewness(goog)
[1] 0.7518

> kurtosis(aapl)
[1] 13.10311
> kurtosis(goog)
[1] 7.454387
```

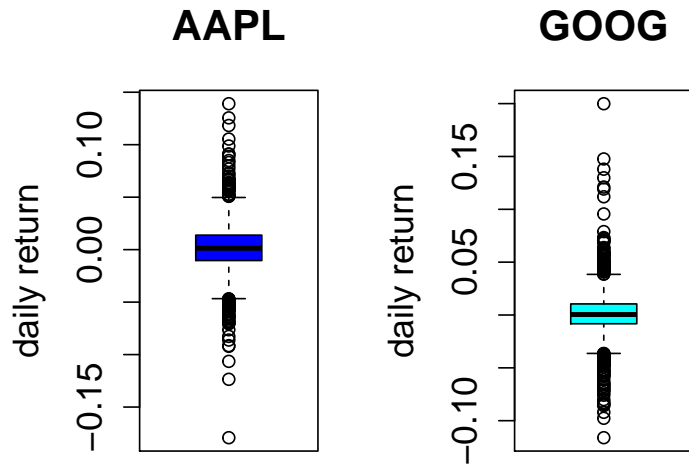


Figure 2: Boxplots

- Describe clearly what you learn from the **summary** and **skewness** and **kurtosis** empirical data.

*From the box plots, we can see many extreme observations out of 1.5 IQR (inter-quarter-range) of the lower and upper quartiles, which are confirmed by the high kurtosis values.*

*Both kurtosis  $> 3$  and that's why we can also see their heavy tails in the histograms.*

*By the skewness values, GOOG is positive skew (mean much higher than median) and AAPL is not "significantly" skew (mean close to median).*

- Over this period use the return distribution and box plots to describe the statistical properties of the investment returns from Apple and Google.

How would you make a relative comparison?

*The average daily return of AAPL is 0.13% that is almost triple of GOOG 0.047%. AAPL almost doubles its kurtosis value of GOOG (13 v.s. 7.5) and there are more extreme values of AAPL than GOOG.*

**Problem D.** (20 points)

Amazon is test marketing a new package delivery system. It wants to see if same-day service is feasible for packages bought with Amazon prime. In a small study of a hundred thousand delivers they find the following times for delivery

	Deliveries	Mean-Time (Hours)	Standard Deviation-Time
new system	80,000	4.5	2.1
old system	20,000	5.6	2.5

1. Find a 95% confidence interval for the decrease in delivery time.

Follow the hint, the answer is (-1.14, -1.06).

```
> (4.5-5.6)-1.96*sqrt(2.1^2/80000+2.5^2/20000)
```

```
[1] -1.14
```

```
> (4.5-5.6)+1.96*sqrt(2.1^2/80000+2.5^2/20000)
```

```
[1] -1.06
```

Hence, the 95% confidence interval is (-1.14, -1.06).

2. If they switch to the new system, what proportion of deliveries will be under 5 hours which is required to guarantee same day service.

As we have a large sample size, we can assume a normal distribution with given sample mean and standard deviation,

$$P(T \leq 5) = 0.594$$

```
> pnorm(5,mean=4.5,sd=2.1)
```

```
[1] 0.594
```

Hence, 59.4% of deliveries will be under 5 hours.

[Hint: a 95% confidence interval for a difference in means  $\mu_1 - \mu_2$  is given by  $(\bar{x}_1 - \bar{x}_2) \pm 1.96\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$  ]