Name: OUTLINE SOLUTION

<div align="center">

University of Chicago
Graduate School of Business

Business 41000: Business Statistics

</div>

Special Notes:

1. This is a closed-book exam. You may use an 8×11 piece of paper for the formulas.

2. Throughout this paper, $N(\mu, \sigma^2)$ will denote a normal distribution with mean μ and variance σ^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 zero, then the variables are independent.

   False. This is only a special case for normal distribution.

2. If $\mathbb{P}\left(A \text{ and } B\right) \leq 0.2$ then $\mathbb{P}(A) \leq 0.2$.

   False. We only know $\mathbb{P}\left(A \text{ and } B\right) \leq \mathbb{P}(A)$.

3. If $X \sim Poi(2)$ and $Y \sim Poi(3)$, then $X + Y \sim Poi(6)$.

False. By the properties of Poisson distribution,

$$E(X + Y) = E(X) + E(Y) = 2 + 3 = 5.$$

So, X+Y does not follow Poi(6), which has a mean 6.

4. At the Apple conference it was claimed that "97% of people love the iWatch".

From a market survey, you found empirically that 1175 out of a sample of 1400 people love the new iWatch. Statistically speaking, you can reject Apple's claim at the 1% significance level.

[Hint: You may use $pnorm(2.58) = 0.995$]

False. For $H_0 : p = p_0 = 0.97$,

$$\hat{p} = \frac{1175}{1400} = 0.839$$

$$s_{\hat{p}} = \sqrt{\frac{p_0(1 - p_0)}{1400}} = 0.00456$$

$$t - stat = \frac{\hat{p} - 0.97}{0.00456} = -28.7 < -2.58 = qnorm(0.005)$$

Yes, we can reject the null hypothesis at 1% significance level.

5. Assuming the Joe DiMaggio's batting average is 0.325 per at-bat and his hits are independent, then he has a probability of about 12% of getting more than 2 hits in 4 at-bats.

False. We need to calculate

$$P(hits > 2) \quad = \quad P(hits = 3) + P(hits = 4)$$

2

$$= \binom{4}{3}\hat{p}^3(1-\hat{p}) + \binom{4}{4}\hat{p}^4$$
$$= 14.85\%$$

6. Arsenal are playing Swansea tomorrow in an English Premier League (EPL). They are favourites to win. The number of goals they expect to score is Poisson with a mean rate of 2.2. Given this, the odds of Arsenal scoring at least one goal is greater than 60%.

   True.
   $$P(\#Goal > 0) = 1 - P(\#Goal = 0) = 1 - ppois(0, 2.2) = 0.89$$

7. The returns for Google stock on the day of earnings are normally distributed with a mean of 5% and a standard deviation of 5%. The probability that you will make money on the day of earnings is approximately 60%.

   False.
   $$P(ret > 0) = 1 - P(ret \leq 0) = 1 - pnorm(0, 0.05, 0.05) = 84.2\%$$

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.

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

This holds even if FG and FT are dependent.

9. Given a random sample of 1000 voters, 400 say they will vote for Donald Trump if the Republican nomination for the 2016 US Presidential Election. Given that he gets the nomination, a 95% confidence interval for the true proportion of voters that will vote for him includes 45%.

False. $\hat{p} = \frac{400}{1000} = 40\%$ *and*

$$s_{\hat{p}} = \sqrt{\frac{\hat{p}(1 - \hat{p})}{1000}} = 0.0072$$

The 95% confidence interval is $[0.4 \pm 0.0142]$, *which doesn't include 45%.*

10. In a sample of $100,000$ emails you found that 550 are spam. Your next email contains the word "bigger". From historical experience, you know that half of all spam email contains the word "bigger" and only 2% of non-spam emails contain it. The probability that this new email is spam is approximately 12%.

True.
$$P(bigger \mid spam) = 50\% \text{ and } P(bigger \mid non-spam) = 2\%$$
$$P(spam) = 0.55\% \text{ and } P(non-spam) = 99.45\%$$

$$
\begin{aligned}
P(spam \mid bigger) &= \frac{P(bigger \mid spam) * P(spam)}{P(bigger \mid spam) * P(spam) + P(bigger \mid non-spam) * P(non-spam)} \\
&= 12.1\%
\end{aligned}
$$

Problem B. (20 points)

Bayes Rule: Hit and Run Taxi.

A certain town has two taxi companies: Blue Birds, whose cabs are blue, and Uber, whose cabs are black. Blue Birds has 15 taxis in its fleet, and Uber has 75. Late one night, there is a hit-and-run accident involving a taxi.

The town's taxis were all on the streets at the time of the accident. A witness saw the accident and claims that a blue taxi was involved. The witness undergoes a vision test under conditions similar to those on the night in question. Presented repeatedly with a blue taxi and a black taxi, in random order, they successfully identify the colour of the taxi 4 times out of 5.

- Which company is more likely to have been involved in the accident?

We need to know $P(Blue \mid identified\ Blue)$ and $P(Black \mid identified\ Blue)$.
First of all, write down some probability statements given in the problem.

$$P(Blue) = 16.7\% \ and \ P(Black) = 83.3\%$$

$$P(identified\ Blue \mid Blue) = 80\% \ and \ P(identified\ Black \mid Blue) = 20\%$$

$$P(identified\ Black \mid Black) = 80\% \ and \ P(identified\ Blue \mid Black) = 20\%$$

Therefore, by Bayes Rule,

$$
\begin{aligned}
P(Blue \mid identified\ Blue) &= \frac{P(identified\ Blue \mid Blue) * P(Blue)}{P(identified\ Blue)} \\
&= \frac{P(identified\ Blue \mid Blue) * P(Blue)}{P(identified\ Blue \mid Blue) * P(Blue) + P(identified\ Blue \mid Black) * P(Black)} \\
&= 44.5\%
\end{aligned}
$$

$$P(Black \mid identified\ Blue) = 1 - P(Blue \mid identified\ Blue) = 55.5\%$$

Therefore, even though the witness said it was a Blue car, the probability that it was a Black car is higher.

5

Problem C. (20 points)

Berkshire Realty is interested in determining how long a property stays on the housing market. For a sample of 800 homes they find the following probability table for length of stay on the market before being sold as a function of the asking price

Days until Sold	Under 20	20-40	over 40
Under $250K	50	40	10
$250-500K	20	150	80
$500-1M	20	280	100
Over $1 M	10	30	10

1. *What is the probability of a randomly selected house that is listed over 40 days before being sold?*

$$P(over40) = \frac{10 + 80 + 100 + 10}{800} = 25\%$$

2. *What is the probability that a randomly selected initial asking price is under 250K?*

$$P(under250K) = \frac{50 + 40 + 10}{800} = 12.5\%$$

3. *What is the joint probability of both of the above event happening?*

$$P(over\ 40 \cap under\ 250K) = \frac{10}{800} = 1.25\%$$

4. *Assuming that a contract has just been signed to list a home for under $500K, what is the probability that Berkshire realty will sell the home in under 40 days?*

$$P(under\ 40|under\ 500K) = \frac{50 + 40 + 20 + 150}{450} = 57.8\%$$

Problem D. (20 points)

In the very famous study of the benefits of Vitamin C, 279 people were randomly assigned to a dose of vitamin C or a placebo (control of nothing). The objective was to study where vitamin C reduces the incidence of a common cold. The following table provides the responses from the experiment

Group	Colds	Total
Vitamin C	17	139
Placebo	31	140

1. Is there a significant difference in the proportion of colds between the vitamin C and placebo groups?

2. Find a 99% confidence interval for the difference. Would you recommend the use of vitamin C to prevent a cold?

[Hint: a 95% confidence interval for a difference in proportions $p_1 - p_2$ is given by $(\hat{p}_1 - \hat{p}_2) \pm 1.96\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$]

To test

$$H_0 : p_1 - p_2 \neq 0,$$

we could perform a two-sample t-test.

$$\hat{p}_1 = \frac{17}{139} = 12.23\% \ and \ \hat{p}_2 = \frac{31}{140} = 22.14\%$$

$$\hat{\sigma} = \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}} = 4.48\%$$

$$t - stat = \frac{\hat{p}_1 - \hat{p}_2}{\hat{\sigma}} = -2.21 < -1.96 = qnorm(0.025)$$

Yes, we can reject the null hypothesis and there is a significant difference at 5% level.

To form the 99% confidence interval of the difference, we use qnorm(0.995) = 2.58.

$$0 \in [-9.91\% - 2.58 * 4.48\%, -9.91\% + 2.58 * 4.48\%]$$

We cannot recommend the use of vitamin to prevent a cold at the 1% significance level.