Name: OUTLINE SOLUTIONS

University of Chicago
Graduate School of Business

Business 41000: Business Statistics
Solution Key

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 1 hr 45 min 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. The Binomial distribution can be approximated by a normal distribution when the number of trials is large.

   True, if n *is large enought, a reasonable approximation to* $B(n, p)$ *is* $N(np, np(1-p))$.

2. The Illinois state lottery introduces a new game. The numbers 1 to 20 are drawn at random without replacement (no two numbers can be the same). You win if you correctly identify all four numbers in exact order. The probability that you win is 1 in $124,750$.

   False, the probability that you win is 1 in 116,280 since $20 * 19 * 18 * 17 = 116,280$.

3. Suppose that a random variable X can take the values $\{0, 1, 2\}$ all with equal probability. Then the expected and variance of X are both 1.

False

$$E[X] = \frac{1}{3} * 0 + \frac{1}{3} * 1 + \frac{1}{3} * 2 = 1$$

$$Var[X] = \frac{1}{3}(0-1)^2 + \frac{1}{3}(1-1)^2 + \frac{1}{3}(2-1)^2$$
$$= \frac{1}{3} + 0 + \frac{1}{3} = \frac{2}{3}$$

4. Suppose that there's a 5% chance that it snows tomorrow and a 80% chance that the Chicago bears play their football game tomorrow given that it snows. The probability that they play tomorrow is then 80%.

False, 99%

$$P(Play) = P(Play|Snow)P(Snow) + P(Play|NoSnow)P(NoSnow)$$
$$= 0.8 * 0.05 + 1 * 0.95 = 0.99$$

5. Let $X \sim N(5, 10)$. Then $P(X > 5) = \frac{1}{2}$.

True, since the mean is 5, then half the density will be to the right of 5.

6. The maximum correlation is 1 and the minimum is 0.

False, the maximum is 1 and the minimum in -1.

7. If two events are independent, then $p(A|B) = p(B|A)$.

False, $p(A) = p(A|B) \neq p(B|A) = p(B)$

8. A kitchen has two cookie jars. The first jar contains 10 ginger snaps and 10 chocolate chip cookies. The second contains an unknown proportion of chocolate chip cookies. If you wish to eat a chocolate chip cookie you should be indifferent to selecting a cookie at random from either jar.

   True, we do not know the proportion of chocolate chip cookies in the second jar.

9. Suppose your website gets on average 2 hits per hour. Then the probability of at least one hit in the next hour is 0.135.

   False, $dpois(0, 2) = 0.135 = e^{-2}$ *so that* $P(hit > 1) = 1 - P(hit = 0) = 0.865$

10. A box has three drawers; one contains two gold coins, one contains two silver coins, and one contains one gold and one silver coin. Assume that one drawer is selected randomly and that a randomly selected coin from that drawer turns out to be gold. Then the probability that the chosen drawer contains two gold coins is 50%.

    False.

    Knowing that we have a gold coin, there is 2/3 chance of being in the 2 gold coin drawer and a 1/3 chance of being in the 1 gold coin drawer. Therefore, the probability that the chosen drawer contains two gold coins is 2/3.

Problem B. (20 points)

Netflix surveyed the general population as to the number of hours per week that you used their service. The following table provides the proportions of each category according to whether you are a teenager or adults.

Hours	Teenager	Adult
< 4	0.18	0.20
4 to 6	0.12	0.32
> 6	0.04	0.14

Calculate the following probabilities:

1. Given that you spend 4 to 6 hours a week watching movies, what's the probability that you are a teenager?
2. What is the marginal distribution of hours spent watching movies.
3. Are hours spent watching Netflix movies independent of age?

1. $\frac{0.12}{0.12+0.32} = 0.2727$

2. See Table:

Hours	*Probability*
< 4	*0.18+0.20=0.38*
4 *to* 6	*00.12+0.32=0.44*
> 6	*0.04+0.14=0.18*

3. No, the marginal distribution of hours given you are a teenager or adult are not the same:

Hours	*Teenager*	*Adult*
< 4	$\frac{0.18}{0.18+0.12+0.04} = 0.5294$	$\frac{0.20}{0.20+0.32+0.14} = 0.3030$
4 *to* 6	$\frac{0.12}{0.18+0.12+0.04} = 0.3529$	$\frac{0.32}{0.20+0.32+0.14} = 0.4848$
> 6	$\frac{0.04}{0.18+0.12+0.04} = 0.1176$	$\frac{0.14}{0.20+0.32+0.14} = 0.2121$

Problem C. (20 points)

The Chicago bearcats baseball team plays 60% of its games at night and 40% in the daytime. They win 55% of their night games and only 35% of their day games. You found out the next day that they won their last game

1. What is the probability that the game was played at night
2. What is the marginal probability that they will win their next game?

Explain clearly any rules of probability that you use.

We have the following info:

$$P(Night) = 0.6, \quad P(Day) = 0.4$$

$$P(Win|Night) = 0.55, \quad P(Lose|Night) = 0.45$$

$$P(Win|Day) = 0.35, \quad P(Lose|Day) = 0.65$$

1.

$$P(Night|Win) = \frac{P(Win|Night)P(Night)}{P(Win)}$$

$$= \frac{P(Win|Night)P(Night)}{P(Win|Day)P(Day) + P(Win|Night)P(Night)}$$

$$\frac{0.55 * 0.6}{0.35 * 0.4 + 0.55 * 0.6} = \frac{0.33}{0.47} = 0.7021$$

2.

$$P(Win) = P(Win|Day)P(Day) + P(Win|Night)P(Night)$$

$$= 0.35 * 0.4 + 0.55 * 0.6 = 0.47$$

Problem D. (20 points)

In a five year study of the effects of aspirin on Myocardial Infarction (MI), or heart attack, you have the following dataset on the reduction of the probability of getting MI from taking aspirin versus a Placebo, or control.

Treatment	with MI	without MI
Placebo	198	10845
Aspirin	104	10933

1. Find a 95% confidence interval for the difference in proportions $p_1 - p_2$.
2. Perform a hypothesis test of the null $H_0 : p_1 = p_2$ at a 1% significance level.

Converting to proportions:

Treatment	with MI	without MI	n
Placebo	0.01793	0.98207	11043
Aspirin	0.00942	0.99058	11037

1. *The CI for a difference in proportion is:*

$$(\hat{p}_1 - \hat{p}_2) \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$$

$$= (0.01793 - 0.00942) \pm 1.96\sqrt{\frac{0.01793(1-0.01793)}{11043} + \frac{0.00942(1-0.00942)}{11037}}$$

$$= 0.00851 \pm 1.96\sqrt{0.0000015945 + 0.0000008454} = 0.00851 \pm 1.96\sqrt{0.0000024399}$$

$$= 0.00851 \pm 1.96 * 0.00156205 = 0.00851 \pm 0.00306$$

Therefore, the CI for the difference in these proportions is:

$$(0.00545, 0.01157)$$

2. *At the 1% level, we would have a implied CI for the difference in proportions of:*

$$= 0.00851 \pm 2.58 * 0.00156205 = 0.00851 \pm 0.00403$$

Therefore, the CI for the difference in these proportions is:

$$(0.00448, 0.01254)$$

We can see that this CI does not contain 0, so therefore we can reject the null that $p_1 = p_2$ *at the 1% significance level.*