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 8×11 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 30 min exam.

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

1. The sample mean of a dataset must be larger than its standard deviation

   False. Many datasets have a negative mean and the standard deviation must be positive.

2. If $P(A \cup B) = 0.5$ and $P(A \cap B) = 0.5$, then $P(A) = P(B)$.

   True. We know that $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ *and so* $P(A) + P(B) = 1$. *Also* $P(A) \geq P(A \cap B) = 0.5$ *and so* $P(A) = P(B) = 0.5$

3. Suppose that the random variable $X \sim N(-2, 4)$ then $-2X \sim N(4, 16)$.

   True. We have $-2X \sim N(4, 16)$.

4. It is always true that the standard deviation is less than the variance

   False. $s_X \geq Var(X)$ *if the* $Var(X) \leq 1$.

5. If the correlation between X and Y is $r = -0.81$ and if the standard deviations are $s_X = 20$ and $s_Y = 25$, respectively, then the covariance is $Cov(X, Y) = -401$.

   False. Using the formula relating covariance and correlation, namely $Cov(X, Y) = rs_X s_Y$ *gives* $Cov(X, Y) = -0.81 \times 20 \times 25 = -405$.

6. Mortimer's steak house advertises that it is the home of the 16 ounce steak. They claim that the weight of their steaks is normally distributed with mean 16 and standard deviation 2. If this is so, then the probability that a steak weights less that 14 ounces is 16%.

   True. We have $P(X < 14) = P\left(\frac{X-16}{2} < \frac{14-16}{2}\right) = P(Z < -1) = 0.1587$

7. A mortgage bank knows from experience that 2% of residential loans will go into default. Suppose it makes 10 such loans, then the probability that at least one goes into default is 95%.

False. Using $P(\text{at least one}) = 1 - P(\text{none})$ *gives* $1 - (0.98)^{10} = 0.1829$.

8. Jessica Simpson is not a professional bowler and 40% of her bowling swings are gutter balls. She is planning to take 90 blowing swings. The mean and standard deviation of the number of gutter balls is $\mu = 36$ and $\sigma = 3.65$.

False. From a Binomial distribution $E(X) = np = 90 \times 0.4 = 36$ *and the standrad deviation* $s_X = \sqrt{np(1-p)} = \sqrt{21.6} = 4.65$.

9. The following data on age and martial status of 140 customers of a Bondi beach night club were taken

Age	Single	Married
Under 30	77	14
Over 30	28	21

Given this data, age and martial status are independent.

False. We can compute conditional probabilities as $P(S|U_{30}) = 0.8462, P(M|U_{30}) = 0.1638$ *and* $P(S|O_{30}) = 0.5714, P(M|O_{30}) = 0.4286$. *The events are not independent as the conditional probabilities are not the same.*

10. An oil company introduces a new fuel that they claim has on average no more than 100 milligrams of toxic matter per liter. They take a sample of 100 liters and find that $\bar{X} = 105$ with a given $\sigma_X = 20$. Then there is evidence at the 5% level that their claim is wrong.

True. This is a one-side test $H_0 : \mu_x \leq 100$ *versus* $H_1 : \mu_X > 100$. *The Z-score is* $Z = \frac{105-100}{20/\sqrt{100}} = 2.5$. *The critical value for a 1-sided test at the* 5% *level is* 1.64. *Therefore you can reject* H_0 *are the* 5% *level.*

Problem B. (20 points)
Suppose that a hypothetical baseball player (call him "Rafael") tests positive for steroids. The test has the following sensitivity and specificity

1. If a player is on Steroids, there's a 95% chance of a positive result.
2. If a player is clean, there's a 10% chance of a positive result.

A respected baseball authority (call him "Bud") claims that 1% of all baseball players use Steriods. Another player (call him "Jose") thinks that there's a 30% chance of all baseball players using Steriods.

- What's Bud's probability that Rafael uses Steroids?
- What's Jose's probability that Rafael uses Steroids?

Explain any probability rules that you use.

Let T and $\bar{T}$ be positive and negative test results. Let S and $\bar{S}$ be using and not using Steriods, respectively. We have the following conditional probabilities

$$P(T|S) = 0.95 \text{ and } P(T|\bar{S}) = 0.10$$

For our prior distributions we have $P_{Bud}(S) = 0.01$ and $P_{Jose}(S) = 0.30$.
From Bayes rule and we have

$$P(S|T) = \frac{P(T|S)P(S)}{P(T)} \tag{1}$$

and by the law of total probability

$$P(T) = P(T|S)P(S) + P(T|\bar{S})P(\bar{S}) \tag{2}$$

Applying these two probability rules gives

$$P_{Bud}(S|T) = \frac{0.95 \times 0.01}{0.95 \times 0.01 + 0.10 \times 0.99} = 0.0876$$

and

$$P_{Jose}(S|T) = \frac{0.95 \times 0.3}{0.95 \times 0.3 + 0.10 \times 0.7} = 0.8028$$

Problem C. (20 points)
A super market carried out a survey and found the following probabilities for people who buy generic products depending on whether they visit the store frequently or not

	Purchase Generic		
Visit	Often	Sometime	Never
Frequent	0.10	0.50	0.17
Infrequent	0.03	0.05	0.15

1. What is the probability that a customer who never buys generics visits the store?
2. What is the probability that a customer often purchases generic?
3. Are buying generics and visiting the store independent decisions?
4. What is the conditional distribution of purchasing generics given that you frequently visit the store?

1. $P(F \cap N) + P(\bar{F} \cap N) = 0.17 + 0.15 = 0.32$
2. $P(F \cap O) + P(\bar{F} \cap O) = 0.10 + 0.03 = 0.13$
3. $P(F|S) = 0.9091, P(\bar{F}|S) = 0.0909$ and $P(F|O) = 0.7692, P(\bar{F}|O) = 0.2308$. The probabilties are not the same so the events are not independent.
4. $P(O|F) = 0.1299, P(S|F) = 0.6494, P(N|F) = 0.2208$.

Problem D. (20 points)
On September 24, 2003, Pete Thamel in the New York Times reported that the Boston Red Sox had been accused of cheating by another American League Team. The claim was that the Red Sox had a much better winning record at home games than at games played in other cities.
The following table provides the wins and losses for home and away games for the Red Sox in the 2003 season

	Record			
Team	Home Wins	Home Losses	Away Wins	Away Losses
Boston Red Sox	53	28	42	39

1. Is there any evidence that the proportion of Home wins is significantly different from home and away games?

Discuss any other issues that are relevant.

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

The home and away winning proportions are given by

$$p_1 = \frac{53}{81} = 0.654 \text{ and } \text{Var} = \frac{p_1(1-p_1)}{n_1} = \frac{53 \times 28}{81^3} = 0.0028$$

and

$$p_2 = \frac{42}{81} = 0.518 \text{ and } \text{Var} = \frac{p_2(1-p_2)}{n_2} = \frac{42 \times 39}{81^3} = 0.0031$$

Can either do the problem as a confidence interval or a hypothesis test.
For the CI, we get

$$\left(\frac{53}{81} - \frac{42}{81}\right) \pm 1.96\sqrt{\frac{53 \times 28}{81^3} + \frac{42 \times 39}{81^3}} = (0.286, -0.014)$$

As the confidence interval contains zero there is no significant difference at the 5% level.
For the hypothesis test we have the null hypothesis that $H_0 : p_1 = p_2$ *versus* $H_1 : p_1 \neq p_2$. *Then the test statistic is approximately normally distributed (large* $n = n_1 + n_2$*) is and given by*

$$Z = \frac{0.654 - 0.518}{\sqrt{0.0059}} = 1.875$$

At the 5% level the critical value is 1.96 and so there's not statistical significance.
The major issue here is whether you should test whether the proportions from home and away are different. There might be a significant home team bias in general and we should test to se if the Red Sox advanatage is significantly different from that. All things else equal this will reduce the significance of the Red Sox home advantage bias.