Name:

University of Chicago

Business Statistics

Special Notes:

1. 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 1.5 hour exam.

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

1. Suppose X and Y are independent random variables and $Var(X) = 6$ and $Var(Y) = 6$. Then $Var(X + Y) = Var(2X)$.

   False. First, $Var(X + Y) = Var(X) + Var(Y)$ *(as* $Cov(X, Y) = 0$*) and so* $Var(X + Y) = 2Var(X)$. *Secondly* $Var(2X) = 4Var(X)$

2. If A and B are mutually exclusive events, then $P(A|B) = 0$.

   True. By definition, if A *and* B *are mutually exclusive events then* $P(A \cap B) = 0$ *and so* $P(A|B) = P(A \cap B)/P(B) = 0$.

3. The trimmed mean of a dataset is more sensitive to outliers than the mean.

   False. The trimmed mean is the average of the observations deleting the outer 5% in tha tails. Hence it is less sensitive to outliers than the sample mean

4. If X is normally distributed with mean 3 and variance 9 then the probability that X is greater than 1 is 0.254.

False. The stanardized Z-score is given by $Z = (1-3)/3 = -0.667$. *From tables* $P(X > 1) = P(Z > -0.667) = P(Z < 0.667) = 0.7486$

5. If $P(A \cap B) \geq 0.10$ then $P(A) \geq 0.10$.

True. $A \cap B$ *is a subset of* A *and so* $P(A) \geq P(A \cap B) = 0.10$.

6. Let investment X have mean return 5% and a standard deviation of 5% and investment Y have a mean return of 10% with a standard deviation of 6%. Suppose that the correlation between returns is zero. Then I can find a portfolio with higher mean and lower variance then X.

True. For example, consider the portfolio $P = 0.5X + 0.5Y$. *This has expected return* $\mu_P = 0.5 \times 5 + 0.5 \times 0.10 = 7.5\%$. *As the correlation is zero, its variance is given by* $\sigma_P^2 = 0.5^2 \times 0.0025 + 0.5^2 * 0.0036 = 0.001525 < Var(X)$.

7. Historically 15% of chips manufactured by a computer company are defective. The probability of a random sample of 10 chips containing exactly one defect is 0.15.

False. The probability of one defectuce is given by the binomial probability $10 \times 0.15 \times (0.85)^9 = 0.3474$.

8. A Normal distribution with mean 4 and standard deviation 3.6 will provide a good approximation to a Binomial random variable with parameters $n = 40$ and $p = 0.10$.

False. For $n \geq 40$, we can approximate a Binomial distribution $Bin(n, p)$ with a normal $N(np, np(1-p))$ In this case, we get $np = 4$ and variance $40 \times 0.1 \times 0.9 = 3.6$ *or a standard deviation of* 1.90.

9. The Central Limit Theorem allows us to approximate the distribution of a sample average by a Normal distribution.

True. The central limit theorem says that the distribution of the sample average is approximately normal, specifically $\bar{X} \approx N\left(\mu, \frac{\sigma^2}{n}\right)$.

10. If 27 out of 100 respondents to a survey state that they drink Pepsi then a 95% confidence interval for the proportion p of the population that drinks Pepsi is $(0.26, 0.28)$.

False. The 95% confidence interval for the proportion is given by $\hat{p} \pm 1.96\sqrt{\hat{p}(1-\hat{p})/n}$. *Which here gives* $0.27 \pm 1.96\sqrt{0.27 \times 0.73/100}$ *or the interval* $(0.18, 0.36)$.

Problem B. (20 points)
You are sponsoring a fund raising dinner for your favorite political candidate. There is uncertainty about the number of people who will attend (the random variable X), but based on past dinners, you think that the probability function looks like this:

x	100	200	300	400	500
$P_X(x)$	0.1	0.2	0.3	0.2	0.2

1. Calculate $E(X)$, the expected number of people who will attend.

2. The owner of the venue is going to charge you \$1500 for rental and other miscellaneous costs. You know that you will make a profit (after per person costs) of \$40 for each person attending. Calculate the expected profit after the rental cost.

3. The owner of the venue proposes an alternative pricing scheme. Instead of charging \$1500, she will charge you either \$5 per person or \$2100, whichever is smaller. So if 100 people come, you only pay \$500. If 500 come, you pay \$2100. Calculate the expected profit under this scheme (still assuming \$40 per plate profit before you pay the owner).

4. Let Y_1 be your profit under the first scheme and Y_2 be your profit under the second. If you do the calculations, it turns out that the standard deviations of these profits are:

$$\sigma_{Y_1} = 4996 \qquad \sigma_{Y_2} = 4488$$

Using the expected values calculated above explain which of the two scenarios you prefer.

1. $E(X) = 100(.1) + 200(.2) + 300(.3) + 400(.2) + 500(.2) = 320$

2. $Y = 40X - 1500$ $E(Y) = E(40X - 1500) = 40E(X) - 1500 = 40(320) - 1500 = 11300$

3. *Profit* $= 40X - \min(2100, 5X)$. *Note that the profit formula is not linear (just like in the newspaper example). Consequently, you need to re-calculate profit for each of the five cases to get the expected value.*

x	*100*	*200*	*300*	*400*	*500*
$p_X(x)$	*0.1*	*0.2*	*0.3*	*0.2*	*0.2*
costs	*500*	*1000*	*1500*	*2000*	*2100*
gross	*4000*	*8000*	*12000*	*16000*	*20000*
profit = gross – costs	*3500*	*7000*	*10500*	*14000*	*17900*

$E(\text{profit}) = 3500(.1) + 7000(.2) + 10500(.3) + 14000(.2) + 17900(.2) = 11280$

4. The profit and standard deviation are larger in (b) than in (c). If you are risk averse, go for (c) (less uncertainty). If you are a gambler, go for (b) (higher average return). Since the difference in σ is larger than in the mean.

Problem C. (20 points)
An oil company has purchased an option on land in Alaska. Preliminary geologic studies have assigned the following probabilities of finding oil

$$P(\text{high quality oil}) = 0.50 \quad P(\text{medium quality oil}) = 0.20 \quad P(\text{no oil}) = 0.30$$

After 200 feet of drilling on the first well, a soil test is taken. The probabilities of finding the particular type of soil identified by the test are as follows:

$$P(\text{soil} \mid \text{high quality oil}) = 0.20 \quad P(\text{soil} \mid \text{medium quality oil}) = 0.80 \quad P(\text{soil} \mid \text{no oil}) = 0.20$$

- What are the revised probabilities of finding the three different types of oil?
- How should the firm interpret the soil test?

1. *Using Bayes' Theorem, we get*

$$P(\text{high quality oil} \mid \text{soil}) = \frac{P(\text{soil} \mid \text{high quality oil})P(\text{high quality oil})}{P(\text{soil})}$$

Using the law of total probability

$$P(\text{soil}) = 0.5 \times 0.2 + 0.2 \times 0.8 + 0.3 \times 0.2 = 0.32$$

Hence

$$P(\text{high quality oil} \mid \text{soil}) = \frac{0.5 \times 0.2}{0.32} = 0.3125$$

Similarly

$$P(\text{medium quality oil} \mid \text{soil}) = \frac{0.8 \times 0.2}{0.32} = 0.50$$

$$P(\text{no oil} \mid \text{soil}) = \frac{0.3 \times 0.2}{0.32} = 0.1875$$

2. *The firm should interpret the soil test as increasing the probability of medium quality oil, but making both no oil and high quality oil less likely than their initial probabilities.*

Problem D. (20 points)
Cooper Realty is a small real estate company located in Albany, New York, specializing primarily in residential listings. They have recently become interested in determining the likelihood of one of their listings being sold within a certain number of days. An analysis of recent company sales of 800 homes in produced the following table:

Days Listed until Sold	Under 20	31-90	Over 90	Total
Under \$50K	50	40	10	100
\$50-\$100K	20	150	80	250
\$100-\$150K	20	280	100	400
Over \$ 150K	10	30	10	50

1. Estimate the probability that a home listed for over 90 days before being sold
2. Estimate the probability that the initial asking price is under \$50K.
3. What the the probability of both of the above happening? Are these two events independent?
4. Assuming that a contract has just been signed to list a home that has an initial asking price of less than \$100K, what is the probability that the home will take Cooper Realty more than 90 days to sell?

Let A be the event that it takes more than 90 days to sell.
Let B denote the event that the initial asking price is under \$50K.

1. $P(A) = (10 + 80 + 100 + 10)/800 = 200/800 = 0.25$
2. $P(B) = 100/800 = 0.125$
3. $P(A \cap B) = 10/800 = 0.0125$
4. *First,* $P(< \$100K) = 350/800 = 0.4375$. *Secondly,*

   $$P(A| < \$100K) = P(< \$100K|A)P(A)/P(< \$100K) = (90/200)\times(200/800)/0.4375 = 0.2571$$