

Name:

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 1 hr 30 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 kurtosis of a distribution is not sensitive to outliers.

*False.* Kurtosis depends on each observation in the distribution and is extremely sensitive to the outliers.

2. Suppose that  $P(A) = 0.4$ ,  $P(B) = 0.5$  and  $P(A \cup B) = 0.7$  then  $P(A \cap B) = 0.3$ .

*False.*  $P(A \cap B) = P(A) + P(B) - P(A \cup B) = 0.2$

3. The central limit theorem says that the distribution of the sample mean is approximately normal.

**True.** *CLT says that for large sample sizes, the mean of the sample is distributed normally and approaches the true population mean*

4. The expected value of the sample mean is the population mean, that is  $E(\bar{X}) = \mu$ .

**True.** *The expected value of the sample does equal the true value.*

5. If the correlation between  $X$  and  $Y$  is zero then the standard deviation of  $X + Y$  is the square root of the sum of the standard deviations of  $X$  and  $Y$ .

**False.** *Using the plug in rule, the standard deviation of  $X + Y$  is the square root of the sum of the variances of  $X$  and  $Y$ .*

6. The probability of observing three heads out of five tosses of a fair coin is 0.6.

**False.**  $P(3 \text{ heads}) = 0.5^5 = 0.03125$ . *10 such combinations are possible so the probability is 0.3125.*

7. Zagats rates restaurants on food quality. In a random sample of 100 restaurants you observe a mean of 20 with a standard deviation of 2.5. Your favorite restaurant has a score of 25. This is statistically different from the population mean at the 5% level.

**True.** We can calculate the  $t$  statistic as  $(25 - 20)/(2.5/10) > 1.96$ . Hence we reject the null hypothesis.

8. Shaquille O'Neal has a 55% chance of making a free throw in Basketball. Suppose he has 900 free throws this year. Then the chance he makes more than 500 free throws is 45%

**False.**  $\mu = np = 495$ ,  $\sigma = \sqrt{np(1-p)} = 14.92$ . Then  $P(X > 500) = P(Z > 0.335) = 1 - 0.631 = 36.9\%$

9. The  $t$ -score is used to test whether a null hypothesis can be rejected.

**True.** The  $t$  score can be used to test whether a null hypothesis can be rejected or not. Alternately, a  $z$  score can be used.

10. A mileage test for a new electric car model called the "Pizzazz" is conducted. With a sample size of  $n = 30$  the mean mileage for the sample is 36.8 miles with a sample standard deviation of 4.5. A 95% confidence interval for the population mean is (32.3, 41.3) miles.

**False.**  $CI = 36.8 \pm 1.96 \frac{4.5}{\sqrt{30}} = (35.19, 38.41)$

**Problem B.** (20 points)

An oil company wants to drill in a new location. A preliminary geological study suggests that there is a 20% chance of finding a small amount of oil, a 50% chance of a moderate amount and a 30% chance of a large amount of oil.

The company has a choice of either a standard drill that simply burrows deep into the earth or a more sophisticated drill that is capable of horizontal drilling and can therefore extract more but is far more expensive. The following table provides the payoff table in millions of dollars under different states of the world and drilling conditions

Oil	small	moderate	large
Standard Drilling	20	30	40
Horizontal Drilling	-20	40	80

Find the following

1. The mean and variance of the payoffs for the two different strategies
2. The strategy that maximizes their expected payoff
3. Briefly discuss how the variance of the payoffs would affect your decision if you were risk averse
4. How much are you willing to pay for a geological evaluation that would tell you with certainty the quantity of oil at the site prior to drilling?

**Part 1**

Using the plug-in rule, we get

$$\begin{aligned}
 E(\textit{Standard}) &= 31 \\
 E(\textit{Horizontal}) &= 40 \\
 \textit{Var}(\textit{Standard}) &= E(X^2) - [E(X)]^2 \\
 &= 1010 - 961 \\
 &= 49 \\
 \textit{Var}(\textit{Horizontal}) &= E(X^2) - [E(X)]^2 \\
 &= 2800 - 1600 \\
 &= 1200
 \end{aligned}$$

**Part 2**

Based on the above, the horizontal drilling maximizes the expected payoff.

### Part 3

Note that the strategy with higher expected payoff has a substantially higher variance. Thus, a risk averse person may settle for a strategy with lower expected payoff and lower variance (standard drilling) while a risk seeking person will chose the horizontal drilling.

### Part 4

If the geological evaluation tells us with certainty the type of oil, then we will be able to chose the strategy which maximizes our payoff under different types of oil. Thus, the amount one should be willing to pay is the expected payoff under this revised schedule of payoff minus the maximum payoff under the current schedule. Given this

$$WTP = 0.2 \times 20 + 0.5 \times 40 + 0.3 \times 80 - 40 = 48 - 40 = 8$$

Therefore, once should be willing to pay \$8 million for this information.

**Problem C.** (20 points)

A marketing firm is studying the effects of background music on people's buying behavior. A random sample of 150 people had classical music playing while shopping and 200 had pop music playing. The group that listened to classical music spent on average \$ 74 with a standard deviation of \$ 18 while the pop music group spent \$ 78.4 with a standard deviation of \$ 12.

1. Test whether there is any significant difference between the difference in purchasing habits. Describe clearly your null and alternative hypotheses and any test statistics that you use.
2. Is there a difference between using a 5% and 1% significance level.

**Part 1**

We are given the following:  $\bar{x} = 74, s_x = 18, n_x = 150, \bar{y} = 78.4, s_y = 12, n_y = 200$

$$H_0 : \mu_x - \mu_y = 0$$

$$H_1 : \mu_x - \mu_y \neq 0$$

In order to test for this hypothesis, we calculate the z score and compare it with the z score at 5% and 1% confidence level.

$$z = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{s_x^2}{n_x} + \frac{s_y^2}{n_y}}} = \frac{74 - 78.4}{\sqrt{\frac{324}{150} + \frac{144}{200}}} = -2.592$$

Since  $z > z_{0.025}$ , we can reject the null hypothesis at 95% confidence level and say that there exist significant differences between purchasing habits.

**Part 2**

For  $\alpha = 0.05, z_{\alpha/2} = 1.96$  and for  $\alpha = 0.01, z_{\alpha/2} = 2.575$ . Thus, the answer does not change when we move from 95% to 99% confidence level.

Alternately, the above question can be solved by creating a confidence interval based on the differences. Again, we will reach the same conclusion.

**Problem D.** (20 points)

A Breathalyzer test is calibrated so that if it is used on a driver whose blood alcohol concentration exceeds the legal limit, it will read positive 99% of the time, while if the driver is below the limit it will read negative 90% of the time. Suppose that based on prior experience, you have a prior probability that the driver is above the legal limit of 10%.

1. If a driver tests positive, what is the posterior probability that they are above the legal limit?
2. At Christmas 20% of the drivers on the road are above the legal limit. If all drivers were tested, what proportion of those testing positive would actually be above the limit?
3. How does your answer to part 1 change. Explain

Let the events be defined as follows:

E - Alcohol concentration exceeds legal limit

NE - Alcohol concentration does not exceed legal limit

P - Breathalyser reads positive

N - Breathalyser reads negative

$$\begin{aligned}P(P|E) &= 0.99 \\P(N|NE) &= 0.90 \\P(E) &= 0.10\end{aligned}$$

**Part 1**

Based on above, we have  $P(P|NE) = 1 - P(N|NE) = 0.10$  and  $P(NE) = 1 - P(E) = 0.90$ . Then,

$$\begin{aligned}P(E|P) &= \frac{P(P|E)P(E)}{P(P)} \\&= \frac{P(P|E)P(E)}{P(P|E)P(E) + P(P|NE)P(NE)} \\&= \frac{0.99 \times 0.10}{0.99 \times 0.10 + 0.10 \times 0.90} \\&= 52.38\%\end{aligned}$$

## Part 2

Now, we have  $P(E) = 0.20$ . Thus,  $P(NE) = 1 - P(E) = 0.80$ . We now calculate

$$\begin{aligned} P(E|P) &= \frac{P(P|E)P(E)}{P(P)} \\ &= \frac{P(P|E)P(E)}{P(P|E)P(E) + P(P|NE)P(NE)} \\ &= \frac{0.99 \times 0.20}{0.99 \times 0.20 + 0.10 \times 0.80} \\ &= 71.22\% \end{aligned}$$

Thus, 71.22% of all the drivers who test positive would be above the legal limit.

## Part 3

Compared to Part 1, the posterior probability increases due to the fact that the probability of a driver testing positive and exceeding the legal limit increases as does the probability of testing positive. However, the increase in the probability of testing positive and exceeding the legal limit is greater than the increase in the probability of testing positive which results in an increase in the posterior probability.