Name: OUTLINE SOLUTIONS

University of Chicago
Graduate School of Business

Business 410: 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.

Honor Code: By signing my name below, I pledge my honor that I have not violated the GSB 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. Selection bias is not a problem when you are estimating a population mean.

   False. When sampling to estimate a population mean, you need to be certain to select a random sample, or selection bias may effect your results. The Dewey/Truman election example is the classic example of selection bias

2. The sample mean, $\bar{x}$, approximates the population mean for large random samples.

   True. The CLT states that the $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$, *or the best guess for the population mean is* $\bar{x}$.

3. The expectation of X minus $2Y$ is just the expectation of X minus twice the expectation of Y, that is $E(X - 2Y) = E(X) - 2E(Y)$.

   True. The plug-in rule states that $E(aX + bY) = aE(X) + bE(Y)$. *Plug in* $A = 1$ *and* $B = -2$. *Hence* $E(X - 2Y) = E(X) + E(-2Y) = E(X) - 2E(Y)$

4. In a random sample of 100 NCAA basketball games, the team leading after one quarter won the game 72 times.

   Then a 95% confidence interval for the proportion of teams leading after the first quarter that go on to win is approximately $(0.6, 0.84)$.

   False. The 95% CI is approximately equal to

$$.72 \pm 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.72 \pm 1.96 \times 0.045 = (.63, .81)$$

5. For the same sample, a 95% prediction interval for a particular team winning is also $(0.6, 0.84)$.

False. The prediction interval is equal to

$$0.72 \pm 2\sqrt{\hat{p}(1-\hat{p})}\sqrt{1+1/n}$$

In this case, that contains the entire interval (0,1).

6. The probability that a standard normal distribution is more than 1.96 standard deviations from the mean is 0.05.

True. In a normal distribution, 2.5% of the population is more than 1.96 standard deviations below the mean, and 2.5% is more than 1.96 standard deviations above the mean. Hence 5% of the distribution is more than 1.96 standard deviations from the mean.

7. In playing poker in Vegas, from 100 hours of play, you make an average of $50 per hour with a standard deviation of $10. A 95% confidence interval for your mean gain per hour is approximately $ $(48, 52)$.

True. The CI is given by $50 \pm 2 \times \frac{10}{\sqrt{100}} = (48, 52)$

8. Thirty six different kinds of ice cream can be found at Ben and Jerry's. There are $58,905$ different combinations of four choices of ice cream.

True. Combinations are given by $36!/(36-6)!6! = 58,905$.

9. If $P(A \cap B) = 0.4$ and $P(B) = 0.8$, then $P(A|B) = 0.5$.

True. $P(A|B) = P(A \cap B)/P(B)$ *by Bayes rule.*

10. If two events A and B are independent then both $P(A|B) = P(A)$ and $P(B|A) = P(A)$.

False. $P(B|A) = P(A)$. *Note that the statement is true if (and only if)* $P(A) = P(B)$.

Problem B. (20 points)

1. A real estate firm in Florida offers a free trip to Florida for potential customers. Experience has shown that of the people who accept the free trip, 5% decide to buy a property.

   - If the firm brings 1000 people, what is the probability that at least 125 will decide to buy a property?

   In order to find the probability that *at least* 125 decide to buy, the binomial distribution would require calculating the probabilities for 125-1000. Instead, we use the normal approximation for the binomial.

   $$\mu = np = 50.$$

   $$\sigma = \sqrt{np(1-p)} = \sqrt{1000 \times .05 \times .95} = \sqrt{47.5} = 6.89$$

   Calculating the Z-score for 125: $Z = \frac{125-50}{6.89} = 10.9$. and $P(Z \geq 10.9) = 0$.

2. In October 1992, the ownership of the San Francisco Giants considered a sale of the franchise that would have resulted in a move to Florida. A survey from the *San Francisco Chronicle* found that in a random sample of 625 people, 50.7% would be disappointed by the move.

   - Find a 95% confidence interval of the population proportion
     The 95% CI is found by

   $$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = .507 \pm 1.96\sqrt{\frac{.507 \times .493}{625}} = .507 \pm 2 \times .020 = (.467, .547)$$

Problem C. (20 points)
In the game Chuck-a-Luck you pick a number from 1 to 6. You roll three dice. If your number doesn't appear on any dice, you lose \$1. If your number appears exactly once, you win \$1. If your number appears on exactly two dice, you win \$2. If your number appears on all three dice, you win \$3.
Hence every outcome has how much you win or lose on the game, namely $-1, 1, 2$ or 3.

- Fill in the blanks in the pdf and cdf values

X	P(X)
-1	?
1	?
2	?
3	?

X	F(X)
-1	?
1	?
2	?
3	?

Explain your reasoning carefully.

1. PDF: This is a binomial experiment with n=3 and p=1/6. Plugging in for $x = 0$:

$$P(x = 0) = (5/6)^3 = .5787$$
$$P(x = 1) = 3(1/6)^1 \times (5/6)^2 = .3472$$
$$P(x = 2) = 3(1/6)^2 \times (5/6)^1 = .0694$$
$$P(x = 3) = (1/6)^3 = .0046$$

2. CDF: The values for this are the probabilities that X is less than or equal to a given value:

$$F(x = 0) = P(x \leq 0) = .5787$$
$$F(x = 1) = P(x \leq 1) = P(x = 0) + P(x = 1) = .9259$$
$$F(x = 2) = .9954$$
$$F(x = 3) = 1$$

- Compute the expected value of the game, $E(X)$.

The expected value is

$$\sum_x xP_X(x) = -1 \times 0.5787 + 1 \times 0.3472 + 2 \times 0.0694 + 3 \times 0.0046 = -\$.08.$$

You expect to lose 8 cents per game.

Problem D. (20 points)
A market research survey finds that in a particular week 28% of all adults watch a financial news television program; 17% read a financial publication and 13% do both.

- Fill in the blanks in the following joint probability table

	Watches TV	Doesn't Watch	
Reads	.13	.04	.17
Doesn't Read	.15	.68	.83
	.28	.72	1.00

  All probabilities come from the given probabilities and the fact that the sum of the column/row probabilities must add to 1.

- What is the probability that someone who watches a financial TV program read a publication oriented towards finance?

$$P(reads|TV) = P(reads \cap TV)/P(TV) = .13/.28 = .4643$$

- What is the probability that someone who reads a finance publication watches a financial TV program.

$$P(TV|reads) = P(reads \cap TV)/P(reads) = .13/.17 = .7647$$

- Why aren't the answers to the above questions equal?

The reason for the difference is the denominators of the equations are different. This is an example of a base rate issue, it is more likely that someone who reads watches TV because fewer people read. This is not a condition of independence.