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.

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. If we drop the largest observation from a sample, then the sample mean and variance will both be reduced.

   False. For example, $100, 101, 102$ *has mean* 101 *and variance* 0.333*, but* $100, 101$ *has mean* 100.5 *and variance* 0.5*.*

2. If $P(A \cap B) = 0.5$ and $P(A) = 0.1$, then $P(B|A) = 0.1$.

   False. $P(B|A) = P(A \cap B)/P(A)$ *would be greater than one.*

3. The probability of at least one head when tossing a fair coin 4 times is 0.9375.

True. $P(\text{at least one head}) = 1 - (0.5)^4 = 0.9375.$

4. In a group of students, 45% play golf, 55% play tennis and 70% play at least one of these sports. Then the probability that a student plays golf but not tennis is 15%.

True. $P(G) = 0.45, P(T) = 0.55, P(G \cup T) = 0.70$ *implies* $P(G \cap \bar{T}) = 0.15$

5. Advertising costs for a 30-second commercial are assumed to be normally distributed with a mean of $10,000$ and a standard deviation of 1000. Then the probability that a given commercial costs between 9000 and $10,000$ is 50%.

False. Required probability is $0.5 - P(Z < -1) = 0.3413$

6. In a sample of 120 Zagat's ratings of Chicago restaurants, the average restaurant had a rating of 19.6 with a standard deviation of 2.5. If you randomly pick a restaurant, the chance that you pick one with with a rating over 25 is less than 1%.

False. $P(X > 25) = 1 - P(Z < 2.16) = 0.0154$

7. A hospital finds that 20% of its bills are at least one month in arrears. A random sample of 50 bills were taken. Then the probability that less than 10 bills in the sample were at least one month in arrears is 50%

True. Using normal approximation, $n = 50, p = 0.2$*, have* $X \sim N(10, 8)$.

8. The following probability table related age with martial status

Age	Single	Not Single
Under 30	0.55	0.10
Over 30	0.20	0.15

Given these probabilities, age and martial status are independent.

False. $P(X > 30 \cap Y = \text{married}) = 0.15 \neq -.35 \times 0.25 = P(X > 30)P(Y = \text{married})$

9. A Chicago radio station believes 30% of its listeners are younger than 30. Out of a sample of 500 they find that 250 are younger than 30. This data supports their claim at the 1% level.

False. $n = 500$*,* $\hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right)$ *and* $H_0 : p = 0.3$ *versus* $H_1 : p \neq 0.3$*. Yields a* 99% *confidence interval with* $qnorm(0.995) = 0.2575$ *of* $(0.184, 0.417)$*. Sample proportion doesn't lie in confidence interval so the data does not support their claim. Or could use a* Z*-test.*

10. As the t-ratio increases the p-value of a hypothesis test decreases.

True. If $t > 0$ *moving out into the tails decreases the p-value. If* $t < 0$ *it will increase the p-value as the t is getting less negative.*

Problem B. (20 points)
A screening test for high blood pressure, corresponding to a diastolic blood pressure of 90mm Hg or higher, produced the following probability table

	Hypertension	
Test	Present	Absent
+ve	0.09	0.03
-ve	0.03	0.85

1. What's the probability that a random person has hypertension?
2. What's the probability that someone tests positive on the test?
3. Given a person who tests positive, what is the probability that they have hypertension?
4. What would happen to your probability of having hypertension given you tested positive if you initially thought you had a 50% chance of having hypertension.

1. $P(H) = 0.09 + 0.03 = 0.12$
2. $P(+) = 0.09 + 0.03 = 0.12$
3. $P(H|+) = P(+ \text{ and } H)/P(+) = 0.09/0.12 = 0.75$
4. From the joint probability table, $P(+|H) = 0.09/(0.03 + 0.09) = 0.75$ and $P(+|\bar{H}) = 0.03/0.88 = 0.034$.

   Now re-calculate $P(+) = P(+|H)P(H) + P(+|\bar{H})P(\bar{H}) = 0.784$. Then, by Bayes rule

$$P(H|+) = \frac{P(+|H)P(H)}{P(+)} = \frac{0.75 \times 0.5}{0.784} = 0.48 \;.$$

Problem C. (20 points)
Pfizer introduced Viagra in early 1998 and during 1998 of the 6 million Viagra users 77 died from coronary problems such as heart attacks. Pfizer claimed that this rate is no more than that in the general population.
You find from a clinical study of $1,500,000$ men who were not on Viagra that 11 of then died of coronary problems in the same length of time during the 77 Viagra users who dies in 1998.

1. Do you agree with Pfizer's claim that the proportion of Viagra users dying from coronary problems is no more than that of other comparable men?

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

   Can do a confidence interval or a Z-score test.

   With Viagra, $\hat{p}_1 = 77/6000000 = 0.00001283$ and without Viagra $\hat{p}_2 = 11/1500000 = 0.00000733$. Need to test whether these are equal.

   With a 95% confidence interval for $(p_1 - p_2)$ using the hint, you get an interval $(0.00000549, 0.0000055)$ which doesn't contain zero. Hence evidence that the proportion is higher.

   Measured very accurately due to the large sample sizes.

   With testing might use a one-sided test and an α of 0.01.

Problem D. (20 points)
The following probability table relates Y the number of TV shows watched by the typical student in an evening to the number of drinks X consumed.

	Y			
X	0	1	2	3
0	0.07	0.09	0.06	0.01
1	0.07	0.06	0.07	0.01
2	0.06	0.07	0.14	0.03
3	0.02	0.04	0.16	0.04

1. What is the probability that a student has more than two drinks in an evening?
2. What is the probability that a student drink more than the number of TV shows they watch?
3. What's the conditional distribution of the number of TV shows watched given they consume 3 drinks?
4. What's the expected number of drinks given they do not watch TV
5. Are drinking and watching TV independent?

1. $P(X > 2) = 0.26$.
2. $P(X > Y) = 0.42$
3. $P(Y|X = 4)$ are given by $(1/13, 2/13, 8/13, 2/13)$.
4. $P(X|Y = 0)$ are given by $(7/22, 7/22, 1/11, 3/11)$ and so $E(X|Y = 0) = 25/22 = 1.13$
5. No. $P(X = 2 \cap Y = 2) = 0.14 \neq 0.43 \times 0.30 = P(X = 2)P(Y = 2)$