Name:

University of Chicago
Booth School of Business
41000: Business Statistics, Winter 2020: Midterm. February, 12.

Special Notes:

1. You may use an 8×11 piece of paper for the formulas.
2. You may use a simple calculator. No internet.
3. This is a 3 hr 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:

GOOD LUCK!!

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

1. Let X and Y be independent random variables. Then the variance of the sum is given by $Var(X+Y) = Var(X) + Var(Y)$.

2. If $P(X|Y) = 0.5$ and $P(Y) = 0.5$, then we must have $P(X) = 0.5$

3. If a researcher has set the significance level at 5% and the test statistic yields a p-value of 0.06, the researcher should reject the null hypothesis.

4. If X has a normal distribution with mean 3 and standard deviation 5, then $Z = \frac{X-3}{5}$ has a standard normal distribution.

5. As the sample size (nunber of data observations) increases, the confidence interval of sample mean narrows, holding all else the same.

6. Manchester City (the EPL champions in 2017) are playing Arsenal. Suppose number of goals of two teams are independent. Both teams are expected on average to score two goals. Then the probability of a $1-1$ draw is approximately 30%. (Hint: Poisson model, density of Poisson(λ) distribution is $P(X = k) = e^{-k}k^2/k!$ where $k! = 1 \times 2 \times .. \times k$.)

7. A semi-conductor company knows from experience that 0.2% of chips will have imperfections. Suppose it makes 1000 such chips, then the probability that **at least one** is imperfect is over 95%.

8. Suppose that the annual returns for Tesla stock are normally distributed with a mean of 20% and a standard deviation of 10%. The probability that Tesla has returns greater than 20% for next year is approximately 30%

9. A friend claims she can tell the difference between Evian and Dasani bottled water. Suppose p is the probability she can identify Evian correctly. In a random experiment with 100 repeated tests, the proportion that she can correctly identified the Evian water is $p = 0.6$. Then you can reject the null hypothesis that $p = \frac{1}{2}$ at the 95% level.

Problem B: Bayes (20 points)

Shipments from an online retailer take between 1 and 7 days to arrive, depending on where they ship from, when they were ordered, the size of the item, etc.

Suppose the distribution of delivery times has the following distribution function:

x	1	2	3	4	5	6	7
P$(X = x)$							
P$(X \leq x)$	0.10	0.20	0.70	0.75	0.80	0.90	1

1. Fill in the above probability table.
2. What is the conditional probability of a delivery arriving on day four given that it did not arrive in the first three days? (Hint: find $P(X = 4 \mid X >= 4)$)

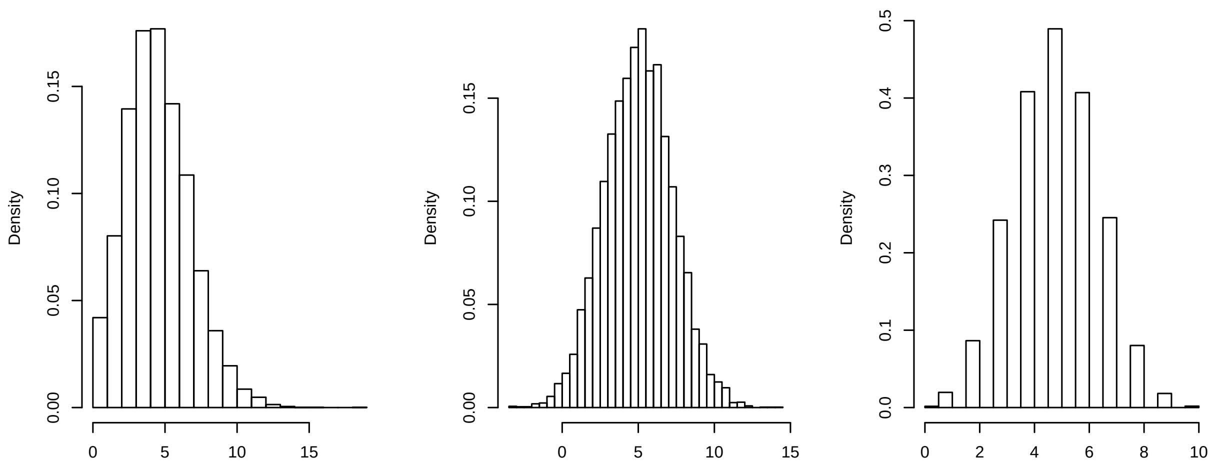
Problem C: A/B Testing. (20 points)

During a recent breakout of the flu, 850 out 6,224 people diagnosed with the virus presented severe symptoms.

During the same flu season, a experimental anti-virus drug was being tested. The drug was given to 238 people with the flu and only 6 of them developed severe symptoms.

Based on this information, can you conclude, for sure, that the drug is a success?

Problem D: Match the Distribution (20 points)



- Left histogram shows the data collected by the new coronavirus hospital in Wuhan. It shows the number of patients admitted every hour.
  a) What distribution would you use to describe this data?
  b) What is your best guess (based on the histogram) about parameter(s) of this distribution?
- Central histogram plots observed differences between the body temperature of the patients infected with coronavirus from the body temperature and a healthy person (98.6 F).
  c) What distribution would you use to describe this data?
  d) What is your best guess (based on the histogram) about parameter(s) of this distribution?
- You split checked-in patients into groups of 10. The right histogram shows distribution of the patients admitted to the hospital form each group of 10. The rest are not actually sick and get sent home.
  e) What distribution would you use to describe this data?
  f) What fraction of the patients get admitted?
  g) What is the probability that two patients out of 10 get admitted?