MATH 338

MIDTERM 1

WED/THURS, SEPTEMBER 27/28, 2017

Your name: ________________________

Your scores (to be filled in by Dr. Wynne):

Problem 1:	____/15
Problem 2:	____/11
Problem 3:	____/10
Problem 4:	____/19
Total:	____/55

You have 75 minutes to complete this exam.

You may refer to your (single-sided, prepared in advance) formula sheet. You may ask Dr. Wynne to clarify what a question is asking for. You may not ask other people for help or use any other resources.

For full credit, show all work except for final numerical calculations (which can be done using a scientific or graphing calculator).

1. The UPBEAT study investigated aerobic exercise as a treatment for depression in individuals with coronary heart disease. 37 participants were randomly assigned to a 16-week exercise regimen, 40 participants to take an antidepressant drug for 16 weeks, and 24 participants to take a placebo pill. The primary outcome measure was participants' score on the Hamilton Depression Rating Scale, which ranges from 0 to 52, with lower scores indicating less depression.

A) [1 pt] How many cases were there in this study?

37 + 40 + 24 = 101 cases

B) [2 pts] Identify the factor (experimental) variable in this study and its levels.

1 pt: the factor is the type of treatment for depression

1 pt: its levels are exercise, antidepressant, and placebo (or equivalents)

C) [1.5 pts] At the beginning of the study, investigators recorded the age, gender, and initial score on the Hamilton Depression Rating Scale for each person enrolled in the study. Classify each variable as categorical or numerical by <u>circling one answer for each variable</u> below.

Age:	categorical	numerical
Gender:	categorical	numerical
Score:	categorical	numerical

D) [2 pts] Why did the investigators caution that "patients who were not interested in exercise or taking an antidepressant were unlikely to have volunteered" to be in the study?

These 101 patients are not a representative sample of all patients with depression and coronary heart disease – only those that are interested in being treated for depression using exercise and/or drugs.

1. (Continued) Refer to the two-way table below for Parts E and F. The row variable is gender and the column variable is type of treatment.

	Exercise	Antidepressant	Placebo
Female	13	15	4
Male	24	25	20

E) [2 pt] Report the marginal distribution of gender.

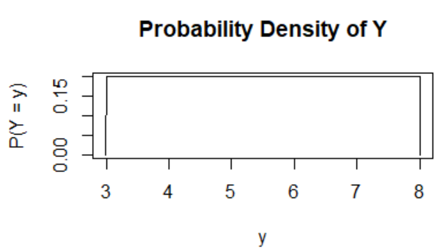
Female: 32/101; Male: 69/101

F) [2 pt] What is the probability that a randomly selected female subject received a placebo?

4/32 = 0.125

For Parts G through I, consider Y, a (continuous) uniform random variable with positive density between 3 and 8.

G) [1.5 pts] Sketch the density curve for Y.



1 pt for shape of curve, 0.5 pts for limits

H) [2 pts] Find P(6 ≤ Y < 9).

1 pt the height of the density curve is 0.2

1 pt P(6 ≤ Y < 9) is the area under the density curve between 6 and 9, which is (8-6)(0.2) = 0.4

I) [1 pt] The median of Y can be found, using the density curve, to be (circle <u>only one answer</u> below):

0.2 0.5 1 4 4.5 5 5.5 7.5 8

2. In a lab experiment, students estimate the concentration of an unknown salt in solution. The solution is known to have a true value of 5% weight by volume, but students' results are known to follow approximately a normal distribution with a standard deviation of 0.09 percentage points.

A) [4 pts] Assuming that the student measurements are unbiased and independent, what proportion of students will obtain a concentration estimate greater than 5.06%?

1 pt sketch or otherwise indicate that students' results are N(5, 0.09)

1 pt z-score for 5.06% = (5.06 – 5)/0.09 = 2/3

1 pt cumulative proportion for z = 2/3 is about 0.7486

1 pt the proportion of students obtaining a concentration estimate <u>greater</u> than 5.06% is 1 - 0.7486 = 0.2514, or about 25% will obtain a concentration estimate greater than 5.06%

B) [6 pts] Assuming that the student measurements are unbiased and independent, what proportion of classes of 20 students will obtain a sample mean concentration estimate greater than 5.06%?

3 pts using CLT, we find that the sample mean $\bar{x} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right) = N\left(5, \frac{0.09}{\sqrt{20}}\right) = N(5, 0.02)$

1 pt the z-score for 5.06% is then (5.06 – 5)/0.02 = 3

2 pts the cumulative proportion for z = 3 = 0.9987, so the proportion of classes of 20 students that will obtain a sample mean concentration estimate greater than 5.06% is 0.0013, or about 0.13% of classes

OR

2 pts by the 68-95-99.7 rule, 99.7% of classes will have a z-score between -3 and 3, so the top 0.15% of classes will have a z-score above 3, so about 0.15% of classes will have a sample mean concentration estimate greater than 5.06%

I gave you 1 out of 6 points for finding the mean number of students, in a class of 20, who obtain a sample mean estimate greater than 5.06%

C) [1 pt] Circle the name of a plot that could be used to display the distribution of the concentrations obtained by the class of 20 students (circle <u>only one answer</u>, though more may be correct):

time plot bar plot pie chart stem plot histogram

1 pt for circling either stem plot or histogram, minus points for circling other stuff

3. According to the NHANES-III survey, the incidence of urinary tract infections in adult women is 13,320 per 100,000 women per year (13.32%). A dipstick nitride test for urinary tract infections has a sensitivity of only 26.7% (roughly 4/15), and a specificity of 93.3% (roughly 14/15) (Zaman *et al.*, 1998).

A) [6 pts] Report <u>and interpret</u> the estimated positive predictive value (PPV) and negative predictive value (NPV) for this dipstick nitride test.

1 pt for using a tree diagram, a two-way table, or Bayes's Rule

3 pts for filling in the probabilities/counts correctly. Letting A = have UTI and B = test positive for UTI,

$$P(A|B) = \frac{P(A)P(B|A)}{P(A)P(B|A) + P(A^C)P(B|A^C)} = \frac{(0.1332)\left(\frac{4}{15}\right)}{(0.1332)\left(\frac{4}{15}\right) + (0.8668)\left(\frac{1}{15}\right)}$$

$$P(A^C|B^C) = \frac{P(A^C)P(B^C|A^C)}{P(A^C)P(B^C|A^C) + P(A)P(B^C|A)} = \frac{(0.8668)\left(\frac{14}{15}\right)}{(0.8668)\left(\frac{14}{15}\right) + (0.1332)\left(\frac{11}{15}\right)}$$

OR

	Test +	Test -	Total
Have UTI	3552	9768	13,320
No UTI	5778.67	80901.33	86,680
Total	9330.67	90669.33	100,000

1 pt the ppv, or probability of having a UTI given the test comes back positive, is about 0.381 or 38.1%

1 pt the npv, or probability not having a UTI given the test comes back negative, is about 0.892 or 89.2%

B) [2 pts] If the incidence of urinary tract infections decreased, which of the following would change (circle <u>all values</u> that would change)?

sensitivity   specificity   positive predictive value   negative predictive value

1 pt per correct answer selected; -50% of earned points for each wrong answer selected

C) [2 pts] If the manufacturers of the test increase its sensitivity without changing its specificity, would the positive predictive value of the test increase or decrease? Explain your answer.

It would increase (1 pt) because the number/probability of true positives would increase, while the number/probability of false positives would stay the same (1 pt)

4. Melvin the Troll has just picked up the Bludgeon of Obfuscating Stupidity! The bludgeon has a chance to hit of 20%; that is, he only hits 20% of the time he swings. If he hits, he does 32 damage; otherwise, he misses and does no damage.

A) [1 pt] What is the probability that Melvin <u>misses</u> (does not hit) on his next swing?

1 – 0.2 = 0.8 or 80%

B) [3 pts] Suppose Melvin swings 10 times, and every one of Melvin's swings is independent of the others. Explain why we can model the number of Melvin's hits as a binomial random variable.

1 pt per assumption:

B: he either hits or he misses. Hits = success, misses = failure

N: The number of swings is fixed at n = 10

S: The probability of success is fixed at 0.2 for each swing

C) [3 pts] What is the probability that Melvin hits exactly 2 times in 10 swings?

2 pts Let X = number of hits, then X ~ B(10, 0.2) and we want to find P(X = 2)

1 pts $P(X = 2) = \frac{10!}{2!8!}(0.2)^2(0.8)^8 = 0.302$

D) [3 pts] Which of the following statements are correct about the mean amount of damage Melvin does in 1000 swings? Circle the letter of <u>each correct statement</u>.

a. This mean is constant, and we can find its value even before Melvin swings 1000 times.

b. This mean is constant, and we don't know its value until we observe Melvin swing 1000 times.

c. This mean is random, and we can find its value even before Melvin swings 1000 times.

d. This mean is random, and we don't know its value until we observe Melvin swing 1000 times.

e. If its value is 6, we can express that fact using the equation $\mu = 6$.

f. If its value is 6, we can express that fact using the equation $\bar{x} = 6$.

g. The shape of its sampling distribution is approximately normal.

h. The mean of its sampling distribution is the same as if he took 10 swings.

i. The standard deviation of its sampling distribution is the same as if he took 10 swings.

j. It does not have a sampling distribution.

1 pt for circling d (but not a, b, or c), 0.5 pts for f (but not e), 0.5 pts for g, 1 pt for h (but not i or j)

4 (continued). Currently, Melvin has equipped the Club of Ad Hominem. 20% of the time he swings the club, he misses and does no damage; 70% of the time, he hits and does 8 damage; and the remaining 10% of the time, he hits and does 12 damage.

E) [2 pts] Let X be the amount of damage Melvin does on a single swing of the Club of Ad Hominem. Write the probability distribution of X.

x	P(X=x)	x*p	$(x-\mu)^2$	$(x-\mu)^2p$
0	0.2	0	(-6.8)^2 = 46.24	9.248
8	0.7	5.6	(1.2)^2 = 1.44	1.008
12	0.1	1.2	(5.2)^2 = 27.04	2.704

The probability distribution of X is given in the left two columns of the table. If you give me the entire table without any indication of which part is the probability distribution, that's okay.

F) [5 pts] Using your distribution from Part (E), find the mean, variance, and standard deviation of X.

2 pts fill in the third column above, and sum down the third column to get E(X) = 6.8

2 pts fill in the fourth and fifth columns, and sum down the fifth column to get Var(X) = 12.96

1 pt sd(X) = sqrt(12.96) = 3.6

G) [2 pts] With which weapon will Melvin expect to do more damage per swing, the Bludgeon of Obfuscating Stupidity or the Club of Ad Hominem? Justify your answer.

1 pt calculate mean damage of Bludgeon to be 0(0.8) + 32(0.2) = 6.4

1 pt he expects to do more damage with the club because 6.8 > 6.4

Extra Space. The tables below show a number of critical values *z* for the standard normal variable $Z \sim N(0, 1)$ and the corresponding cumulative proportions, corresponding to $P(Z \leq z)$.

z-score	Cumulative Proportion
-3.00	0.0013
-2.50	0.0062
-2.00	0.0228
-1.65	0.0495
-1.28	0.1003
-1.00	0.1587
-0.67	0.2514

z-score	Cumulative Proportion
0.67	0.7486
1.00	0.8413
1.28	0.8997
1.65	0.9505
2.00	0.9772
2.50	0.9938
3.00	0.9987

The rest of this space to be used for extra work: