MATH 338

MIDTERM 1

WEDNESDAY, FEBRUARY 28, 2018

Your name: _____

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

Problem 1: ____/11

Problem 2: ____/8

Problem 3: ____/16

Problem 4: ____/10

Problem 5: ____/10

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 sex and body weight (in kg) of a sample of 144 cats was recorded. The table below shows some summary statistics. Use the table to answer the following questions:

	Female Cats	Male Cats
Count	47	97
Minimum Weight	2.00	2.00
Q1 (Weight)	2.15	2.50
Median Weight	2.30	2.90
Q3 (Weight)	2.50	3.20
Maximum Weight	3.00	3.90
Mean Weight	2.36	2.90
Std. Dev. of Weight	0.27	0.47

A) [1.5 pts] Name all variables in this study, and identify each variable as categorical or quantitative.

B) [1.5 pts] Can we find the overall median body weight for the 144 cats from this table? If so, find it. If not, explain why not.

C) [4 pts] In pounds (1 kg = 2.2 lbs), what is the overall mean body weight of the 144 cats?

D) [4 pts] If the distribution of male cat body weights were normal with mean and standard deviation given by the table, give three ranges of weights corresponding to, respectively, the lightest 95% of male cats, the heaviest 95%, and the middle 95%. Label which range corresponds to which group of cats.

2. A 2017 American Medical Association Report recorded the mortality rate (death rate) from various types of cancer in 3,047 counties and county-equivalents in the U.S.

A) [1 pt] What was a case in this study?

B) [2 pts] Should we treat the national mortality rate from liver cancer as a statistic or a parameter? Justify your answer.

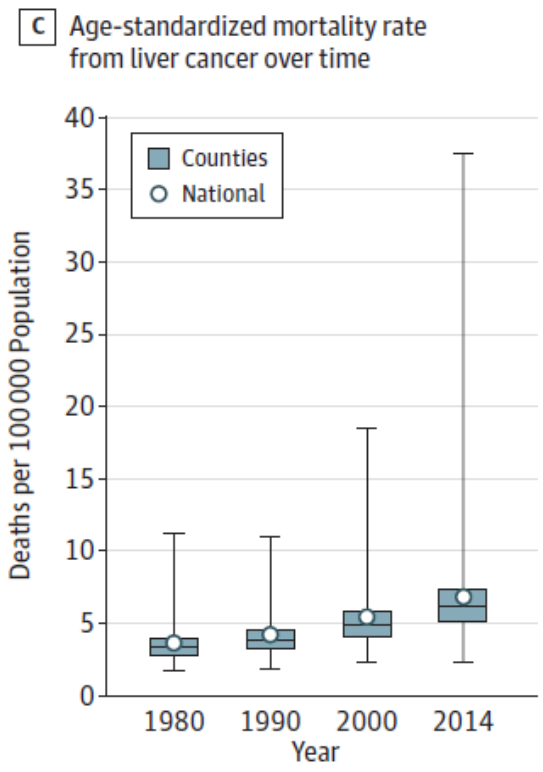
For parts C-E, refer to the figure to the right.

C) [1 pt] What plots are depicted in the figure (circle all correct answers)?

box plot histogram bar plot

D) [1 pt] In 1990, the distribution of liver cancer mortality rate among the counties was (circle the most likely correct answer):

skewed left symmetric skewed right



E) [3 pts] Which of these statements about the death rates from liver cancer are definitely true (circle the letter of each statement that can be proven from the figure)?

a. Over time, the death rate from liver cancer in the average American county has increased

b. Over time, the variability among counties in death rates from liver cancer has increased

c. In 2000, less than 75% of counties had death rates from liver cancer below 5%

d. In 2014, at least half of all counties had death rates between 5 and 10 deaths per 100,000 population

e. According to our rule for numerically finding outliers, some counties in 2014 have outlier death rates

f. The distribution of liver cancer death rates among counties is unimodal

3. In lecture we discussed how to lose money at American roulette. It turns out that in European roulette there is only one green 0, and so 18 out of 37 slots are red and 18 out of 37 slots are black. If you bet $1 on black, you win $1 if black shows up and lose $1 if it doesn't.

A) [3 pts] Suppose you bet $1 on black for each of 8 independent spins of a French roulette wheel. Explain why the number of bets you win can be modeled as a binomial random variable.

B) [3 pts] What is the probability that you win exactly 3 out of the 8 bets?

C) [5 pts] How much <u>total</u> money do you expect to win or lose over your 8 bets? Justify your answer.

D) [5 pts] We design an experiment to test whether a book's "system" for roulette is better than betting haphazardly on red or black. Briefly describe these terms <u>in the context of our experiment</u>.

Experimental (Factor) Variable:

Response Variable:

Control Group:

Random Assignment:

Repetition:

4. A 2010 study investigated cognitive-behavioral therapy in a sample of 109 children with obsessive-compulsive disorder (OCD). After treatment, they recorded the children's scores on the Yale-Brown Obsessive-Compulsive Scale (Y-BOCS) and whether the clinician believed the OCD to be in remission.

A) [8 pts] 73 of the 80 children in remission had a Y-BOCS score of 14 or below, while 26 of the 29 children judged not to be remission had higher scores. Suppose the researchers decide to use "Y-BOCS score of 14 or below" as a test to diagnose whether a child is in remission. From this sample, estimate the sensitivity, specificity, positive predictive value, and negative predictive value of this test. Round all estimates to the nearest percent.

Sensitivity: _____ Specificity: _____

PPV: _____ NPV: _____

Space to show work:

B) [2 pts] Suppose that the researchers instead decided to use "Y-BOCS score of 11 or below" to diagnose whether a child is in remission. How will this affect the sensitivity and specificity of the test? Circle the <u>most likely outcome</u> for both sensitivity and specificity.

Sensitivity will: increase decrease stay the same

Specificity will: increase decrease stay the same

5. Suppose that we plan to observe 25 independent realizations of the random variable $X \sim N(15, 5)$ and, independently, 100 independent realizations of the random variable $Y \sim N(12, 8)$.

A) [2 pts] Explain why we cannot, at this point, evaluate whether the statement $\bar{X} > \bar{Y}$ is true or false.

B) [8 pts] For each of the ten statements below, circle the equality or inequality sign (<, =, or >) that correctly describes the relationship between the quantity on the left and the quantity on the right. There is exactly one correct answer per row.

μ_X	<	=	>	μ_Y
σ_x	<	=	>	σ_Y
$E(X+Y)$	<	=	>	$E(X-Y)$
$Var(X+Y)$	<	=	>	$Var(X-Y)$
$E(\bar{X})$	<	=	>	$E(\bar{Y})$
$Var(\bar{X})$	<	=	>	$Var(\bar{Y})$
$P(X < 10)$	<	=	>	$P(X \leq 10)$
$P(X < 10)$	<	=	>	$P(\bar{X} < 10)$
$P(\bar{X} < 10)$	<	=	>	$P(\bar{Y} < 10)$
$P(\bar{Y} < 10)$	<	=	>	$P(\bar{Y} > 14)$

Space to show work:

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 ≤ 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: