



University of Colorado at Boulder

Department of Mathematics

395 UCB

Boulder, Colorado 80309-0395

(303) 492-7664

Fax: (303) 492-7707

Math 2510-001

MIDTERM 1

PRACTICE PROBLEMS

Q1-Q16

normal
distribution

Q17-Q31

example
midterm
CU Boulder

Q32-Q57

example
midterms
(extra)

Q58-Q65

challenge
problems

ASK COLTON FOR
SOLUTIONS IN CLASS
OR VIA EMAIL.

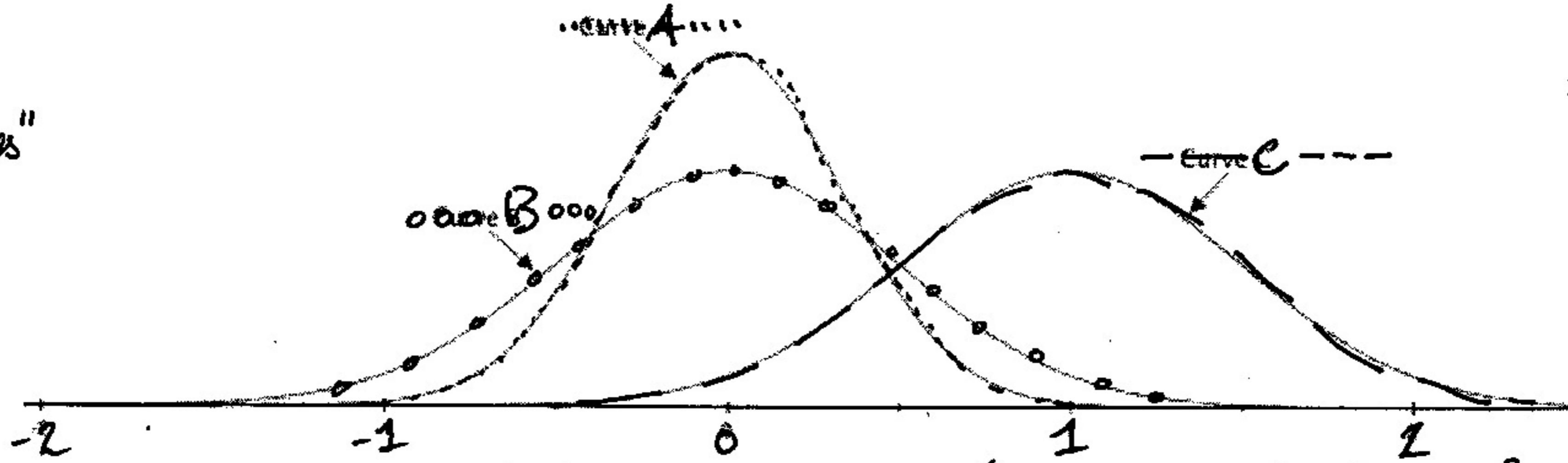
Please make sure to show your work. If you use a function in your calculator to derive the answer, please write what was entered into your calculator and the output of the function that you used.

On my honor, as a University of Colorado at Boulder student, I have neither given nor received unauthorized assistance on this work.

SIGNATURE:

A Q1

- 1) Consider the graph of three normal distributions, given below, to answer the following questions:



which distribution A, B, or C has mean not at zero?

- (a.) (1 point) Write the letter of the distribution which has a non-zero mean.

A B C

which distribution A, B, or C is the "tightest"?

- (b.) (1 point) Write the letter of the distribution which has the smallest standard deviation.

A B C

"normal
percentiles"
Q2

- (3 points) The average SAT score for a high school senior in Colorado is 1065 with a standard deviation of 169. Assume that SAT scores are approximately normally distributed. If a student scores in the 76th percentile, what was their SAT score? Round your answer to two decimal places.

Q3

The body weight of a fully grown female golden retriever is normally distributed with mean 27.5 kg and standard deviation 2.2 kg.

"quickly"
★

a.) (1 point) What is the z-score corresponding to a 24 kg fully grown female golden retriever? Round your answer to two decimal places.



b.) (2 points) What is the probability that a fully grown female golden retriever selected at random weighs more than 28 kg? Round your answer to four decimal places.

c.) (2 points) Two fully grown female golden retrievers are selected. Bella has a body weight corresponding to the z-score 1.2 and Daisy has a body weight corresponding to the z-score -0.4. Which dog has a body weight which is considered more unusual with respect to the rest of the population? Explain your reasoning.

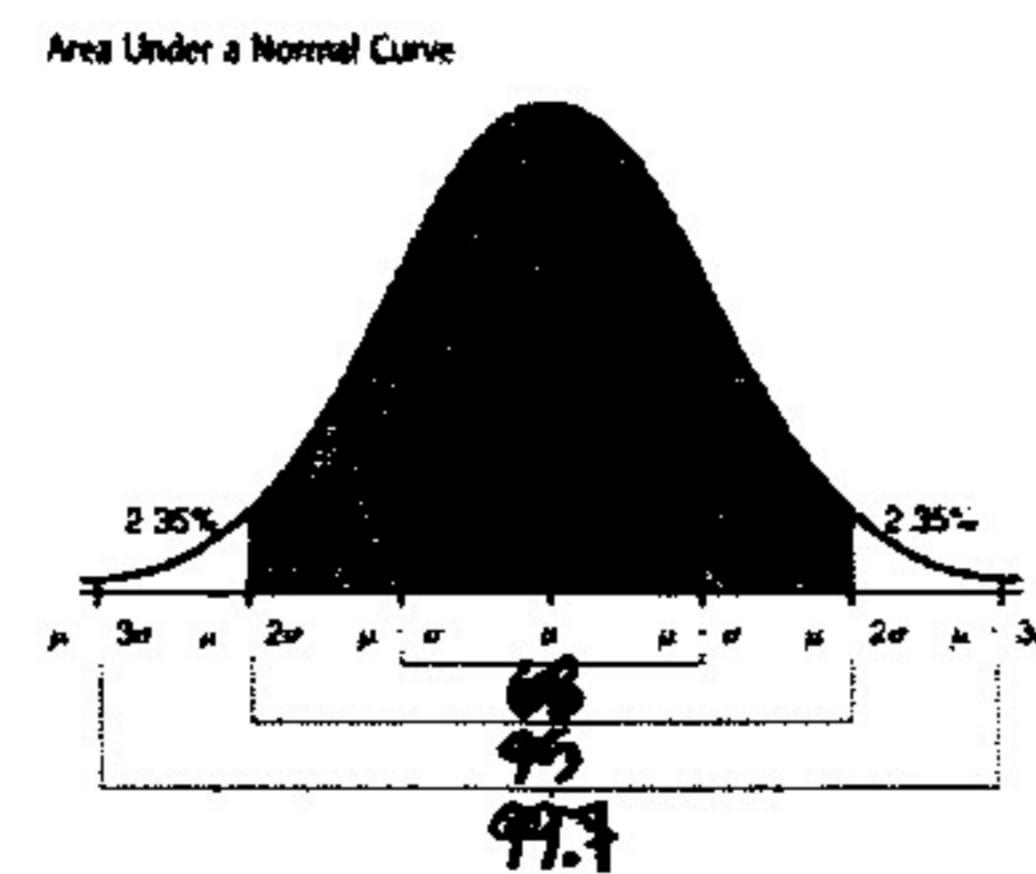
Q4

- * The incubation time for Rhode Island Red chicks is normally distributed with a mean of $\mu = 21$ days and a standard deviation of about $\sigma = 1$ day. If 1000 eggs are being incubated, how many chicks do we expect to hatch

- (a) on day 21 or after?
- (b) between 20 and 21 days?
- (c) between 22 and 23 days?
- (d) between 19 and 22 days?

Q5

- ** The Empirical Rule tells us the percentage of data values that lie within certain intervals about the mean when we have a normal distribution as follows.



- (a) According to Chebyshev's Theorem, what is the minimum percentage of data that must lie within the interval $\mu - 2\sigma$ to $\mu + 2\sigma$? How does this compare to the results for the Empirical Rule?
- (b) The diagram for the Empirical Rule only accounts for 99.7% of the data values. Where do the remaining 0.3% of the data fall?

the complement of the central 99.7%

- (c) Keeping in mind that remaining 0.3% of the data, what percentage of the data values lie to the left of $\mu - \sigma$?
- (d) Keeping in mind that remaining 0.3% of the data, what percentage of the data values lie to the left of $\mu + 2\sigma$?

computing
Z-scores

Q6

~~% of data between $\mu - 3\sigma$ and $\mu + 3\sigma$ is 99.7% (from 6.1 and 9.3 of Brase and Brase)~~, suppose the mean of a normal distribution is $\mu = 75$ with a standard deviation of $\sigma = 8$. For each of the following values of the random variable x , determine whether x is above or below the mean and then determine the number of standard deviations away from the mean that x lies. (Note that your second answer could be a fraction or decimal number.)

(a) $x = 83$

(b) $x = 51$

(c) $x = 79$

(d) $x = 65$

Q7

AA Hydraulic pressure in the main cylinder of the landing gear of a commercial jet is very important for a safe landing. In-flight landing tests show that the actual pressure in the main cylinders is variable (approximately normal) with a mean of 819 pounds per square inch and a standard deviation of 23 pounds per square inch, which are considered safe.

Two different planes were tested with 10 test-landings.

landing number for plane A	1	2	3	4	5	6	7	8	9	10
pressure	870	855	830	815	847	836	825	810	792	820
landing number for plane B	1	2	3	4	5	6	7	8	9	10
pressure	865	850	841	820	815	789	801	765	730	725

"control
signals"

- ★ (a) Is the pressure for Plane A "in control" or not? If not, describe the specific out-of-control signals present.

- ★ (b) Is the pressure for Plane B "in control" or not? If not, describe the specific out-of-control signals present.

Q8

* Let y be a random variable with a normal distribution with mean $\mu = 15$ and standard deviation $\sigma = 4$. Find the indicated probability.

(a) $P(y \leq 12)$

(b) $P(y > 10)$

(c) $P(15 \leq y \leq 20)$

(d) $P(8 < y < 22)$

Q9

- ★ * Find the z -value described.

- (a) Find z such that 5.2% of the standard normal curve lies to the left of z .

- (b) Find z such that 5% of the standard normal curve lies to the right of z .

"inverse
 z -value
problems"

"confidence
intervals"

- ★ (c) Find z such that 95% of the standard normal curve lies between $-z$ and $+z$.

"percentiles
in normal
distributions"

Q10

Suppose that the scores for a Chemistry midterm had a normal distribution with mean $\mu = 71.4$ and a standard deviation of $\sigma = 8.3$.

- (a) Find the exam score for a student that scored in the 80th percentile.
- ★ (b) Find the exam score for a student that scored in the 25th percentile.
- (c) Find the exam score for a student who score lower than 60% of the class.

Q11

Suppose the distribution of heights of American men (20 years of age and older) is approximately normal with a mean of 69.4 inches and a standard deviation of 3 inches.

- (a) What is the z value for a height of 6 feet?
- (b) What percentage of American men (20 years of age and older) are shorter than 6 feet?
- (c) What is the z value for a height of 6 feet 4 inches?
- (d) What percentage of American men (20 years of age and older) are taller than 6 feet 4 inches?

Q12

Suppose a population has a distribution with $\mu = 72$ and $\sigma = 8$.

- ★ ★ (a) If we know nothing about the original population distribution and random samples of size $n = 16$ are selected, why can't we say anything about the \bar{x} distribution of sample means?
"nothing relevant to be concluded"
- ★ (b) If we know that the original population distribution is normal, then what can we say about the \bar{x} distribution of random samples of $n = 16$. In this case, find $P(68 \leq \bar{x} \leq 73)$.

Q13

The heights of 18-year-old men are approximately normally distributed, with a mean of 68 inches and a standard deviation of 3 inches.

- (a) What is the probability that an 18-year-old man selected at random is between 67 and 69 inches tall?
- ★ (b) If a random sample of nine 18-year-old men are selected at random, what is the probability that the mean height of the sample \bar{x} is between 67 and 69 inches?
- ★ (c) Explain why the probability in part (b) is so much higher than the probability in part (a) even though they both refer to the same interval of heights (67 to 69 inches).

Q14

- Let x be a random variable that represents white blood cell count per cubic milliliter of whole blood. Assume that x has a distribution that is approximately normal with a mean of $\mu = 7500$ and estimated standard deviation of $\sigma = 1750$. A test result of $x < 3500$ is an indication of leukopenia. This indicates bone marrow repression that may be the result of a viral infection.

- (a) What is the probability that, on a single test, x is less than 3500?
- ★ (b) Suppose a doctor uses the average \bar{x} for two tests taken about a week apart. What can we say about the probability distribution of \bar{x} ? What is the probability that $\bar{x} < 3500$?
- ★ (c) Repeat part (b) but with $n = 3$ tests.
- ★★ (d) How did the probabilities change as n increased? What do such results imply about a patient that has $\bar{x} < 3500$ for 3 those tests?

Q15

- ★★ 23. Assume that IQ scores are normally distributed, with a standard deviation of 15 points and a mean of 100 points. If 10 people are chosen at random, what is the probability that the sample mean of their IQ scores will not differ from the population mean by more than 2 points?

Q16

- ★ 23. A large tank of fish from a hatchery is being delivered to a lake. The hatchery claims that the mean length of fish in the tank is 15 inches, and the standard deviation is 2 inches. A random sample of 36 fish is taken from the tank. Let \bar{x} be the mean length of the sample. What is the probability that \bar{x} is within 0.5 inches of the claimed population mean?

MATH 2510
Mini-Midterm 1

Q17

A health and fitness club surveys 40 randomly selected members and found that the average weight of those questioned is 157 lb. Is this average an example of a statistic or a parameter?

A) Parameter

B) Statistic

Q18

in Fahrenheit ($^{\circ}\text{F}$)

** Which of the four levels of measurement is most appropriate for temperatures of the ocean at various depths?

A) Nominal

B) Ordinal

C) Ratio

D) Interval

Q19

An archer is able to hit the bull's-eye 55% of the time. If she shoots 8 arrows, what is the probability that gets exactly 4 bull's-eyes? Assume each shot is independent of the others.

A) 0.00375

B) 0.263

C) 0.172

D) 0.0915

Q20

** An education researcher randomly selects 48 middle schools and interviews all the teachers at each school. Identify which of these types of sampling was used.

A) Systematic

B) Cluster

C) Random

D) Stratified

E) Convenience

Q21

** The frequency distribution below summarizes employee years of service for Alpha Corporation. Determine the width of each class.

Years of service	Frequency
1-5	5
6-10	20
11-15	25
16-20	10
21-25	5
26-30	3

A) 10

B) 5

C) 4

D) 6

Q22

* A sample of 100 wood and 100 graphite tennis rackets are taken from the warehouse. If 7 wood and 14 graphite are defective and one racket is randomly selected from the sample, find the probability that the racket is wood or defective.

A) 0.105

B) 0.57

C) 0.535

D) There is insufficient information to answer the question.

"make a
table of
possible
values"

☆☆

Q24

The ages of the members of a gym have a mean of 44 years and a standard deviation of 12 years. What can you conclude from Chebyshev's theorem about the percentage of gym members aged between 26 and 62?

- A) The percentage is at least 55.6%
 - B) The percentage is at least 33.3%
 - C) The percentage is approximately 33.3%
 - D) The percentage is at most 55.6%

Q25

A test consists of 10 true/false questions. To pass the test a student must answer at least 6 questions correct. If a student guesses on each question, what is the probability that the student will pass the test?

~~10~~ A police department reports that the probabilities that 0, 1, 2, and 3 burglaries will be reported in a given month are 0.45, 0.37, 0.17, and 0.01, respectively. Determine the mean of this probability distribution. $k \quad P(X=k)$

Compute the expectation of this r.r. X with PMF given -

$$k \quad P(X=k)$$

o | o.45

1 | 033

2 0.97
3 0.47

2 | 0.1 +
3 | -

Q27

Q26

20 For the stem-and-leaf plot below, find the minimum and maximum data values.

Key: ~~graph~~ $2 \mid 7 = 27$

1	0	5						
2	6	6	6	7	8	9		
2	7	7	7	8	8	9	9	9
3	0	1	1	2	3	4	4	5
3	6	6	6	7	8	8	9	
4	0	2						

- A) min = 20; max = 22
 C) min = 10; max = 42

- D) min = 10; max = 12

11) The random variable X is the number of houses sold by a realtor in a single month at the Sendsom's Real Estate office. Its probability distribution is as follows.

Houses Sold  Probabiliti

B)	Probability
0	0.21
1	0.01
2	0.12
3	0.16
4	0.01
5	0.14
6	0.11
7	0.21

- A) $I = 2.50$ ✓ B) $I = 3.60$ ✓ C) $I = 3.40$ ✓ D) $I = 3.35$

"conditional probability" [Q28]

list:
draw
a tree
diagram

	Number of flights arrived on time	Number of flights arrived late	Totals
Podunk Airlines	33	6	39
Upstate Airlines	43	5	48

- a.) If a flight is selected at random, what is the probability that it was an Upstate Airline flight?
- b.) If a flight is selected at random, what is the probability that it was an Upstate Airline flight or that it arrived on time?
- c.) If a flight is selected at random, what is the probability that it was an Upstate Airline flight given that it arrived on time?
- d.) If a flight is selected at random, what is the probability that it was an Upstate Airline flight and that it arrived on time?

[Q29]

** A car insurance company has determined that 8% of all drivers were involved in a car accident last year. Among the 14 drivers living on one particular street, 3 were involved in a car accident last year. If 14 drivers are randomly selected, what is the probability of getting 3 or more who were involved in a car accident last year?

A) 0.074

B) 0.096

C) 0.926

D) 0.407

either
binomial
or use normal
approximation
... which?

Q30

- j-var
stats.*
- * A medical research team studied the ages of patients who had strokes caused by stress. The ages of 34 patients who suffered stress strokes were as follows.

29	30	36	41	45	50	57	61	28	50	36	58
60	38	36	47	40	32	58	46	61	40	55	32
61	56	45	46	62	36	38	40	50	27		

- a.) Compute the mean and standard deviation of this sample.

- A** b.) Construct a frequency distribution for these ages. Use 8 classes beginning with a lower class limit of 2

Age	Midpoint	Frequency	Relative Frequency	Cumulative Frequency

- b.) Draw a histogram representing this data.

Q31

- * Find the indicated probability. Round to the nearest thousandth.

15) In a batch of 8,000 clock radios 7% are defective. A sample of 5 clock radios is randomly selected without replacement from the 8,000 and tested. The entire batch will be rejected if at least one of those tested is defective. What is the probability that the entire batch will be rejected?

- A) 0.304 B) 0.200 C) 0.696 D) 0.0700

Midterm Exam 1

Use this table for questions 1 through 3. This table represents the first 8 observations from a sample of 200 individuals, who reported their age, race, income, and job satisfaction score on a scale from 0 to 100.

Age	Race	Income	Score
21	W	less than \$10,000	29
33	B	\$20,000-23,000	32
41	B	more than \$100,000	84
26	A	\$30,000-40,000	78
22	O	\$10,000-20,000	87
19	A	\$40,000-50,000	42
34	W	\$50,000-60,000	21
26	W	less than \$10,000	91
:	:	:	:

Q32

★ ★ Which of the following best describes the Income variable?

- a. categorical, dichotomous nominal
- b. categorical, ordinal
- c. quantitative, "continuous", ratio
- d. quantitative, "discrete", interval
- e. categorical, boolean

Q33

★ ★ Which type of plot would be most useful for visualizing the relationship between Age and job satisfaction Score?

- a. histogram
- b. single box plot
- c. side by side box plot
- d. scatter plot
- e. dot plot

Q34

★ ★ Below are some summary statistics from the score variable. Which of the following is true?

min		Q1		median		Q3		max		mean		sd		n		missing	
30		57		69.5		77		99		65.075		16.09361		200		0	

- a. the standard deviation estimate is not possible because score is a whole number
- b. there is evidence that the distribution of score is right-skewed
- c. the minimum value of 30 would be identified as out outlier in a box plot
- d. there were more survey respondents who reported job satisfaction scores less than 57 than survey respondents reported job satisfaction scores greater than 77
- e. none of the above are true

Q35

* * Which of the following is true regarding the Central Limit Theorem (CLT)?

- a. If your sample size is $n = 30$ exactly, then you are guaranteed to have an approximately normal sampling distribution of the sample mean.
- b. As the sample size n increases, the data distribution should become approximately normal.
- c. The Central Limit Theorem states that the sampling distribution of the sample mean should always have the same shape as the population distribution.
- d. none of the above

Q36

* A researcher is interested in seeing if there is an association between whether or not an individual uses a smart phone right before bed and how well an individual sleeps. Participants in the study report whether or not they used a smart phone before bed and then rate their quality of sleep as either “very poor,” “poor,” “average,” “good,” or “very good.” The researcher concludes at the end of the study that there is an association between the two variables. Which of the following statements are true?

- a. The explanatory variable is whether or not a participant uses a smart phone and the response variable is sleep quality. An association was present, so the researcher can say the response and explanatory variables are independent.
- b. The explanatory variable is sleep quality and the response variable is whether or not a participant uses a smart phone. An association was present, so the researcher can say the response and explanatory variables are not independent.
- c. The explanatory variable is whether or not a participant uses a smart phone and the response variable is sleep quality. An association was present, so the researcher can say the response and explanatory variables are not independent.
- d. The explanatory variable is whether or not a participant uses a smart phone and the response variable is sleep quality. An association was present, so the researcher can determine that smart phone use right before bed causes change in quality of sleep.
- e. The explanatory variable is sleep quality and the response variable is whether or not a participant uses a smart phone. An association was present, so the researcher can determine that smart phone use right before bed causes change in quality of sleep.

Q37

* A comprehensive survey released by a college reports that the true proportion of all students at the college who use drugs is 0.3. You survey 100 students in your dorm and record that the proportion of students who use drugs is 0.15. The proportion of all students at this college who use drugs is a _____ and the proportion of students who use drugs in your dorm is a _____.

- a. statistic; parameter
- b. parameter; statistic
- c. population; sample
- d. measure of central tendency, measure of variability
- e. none of the above

Midterm Exam 1

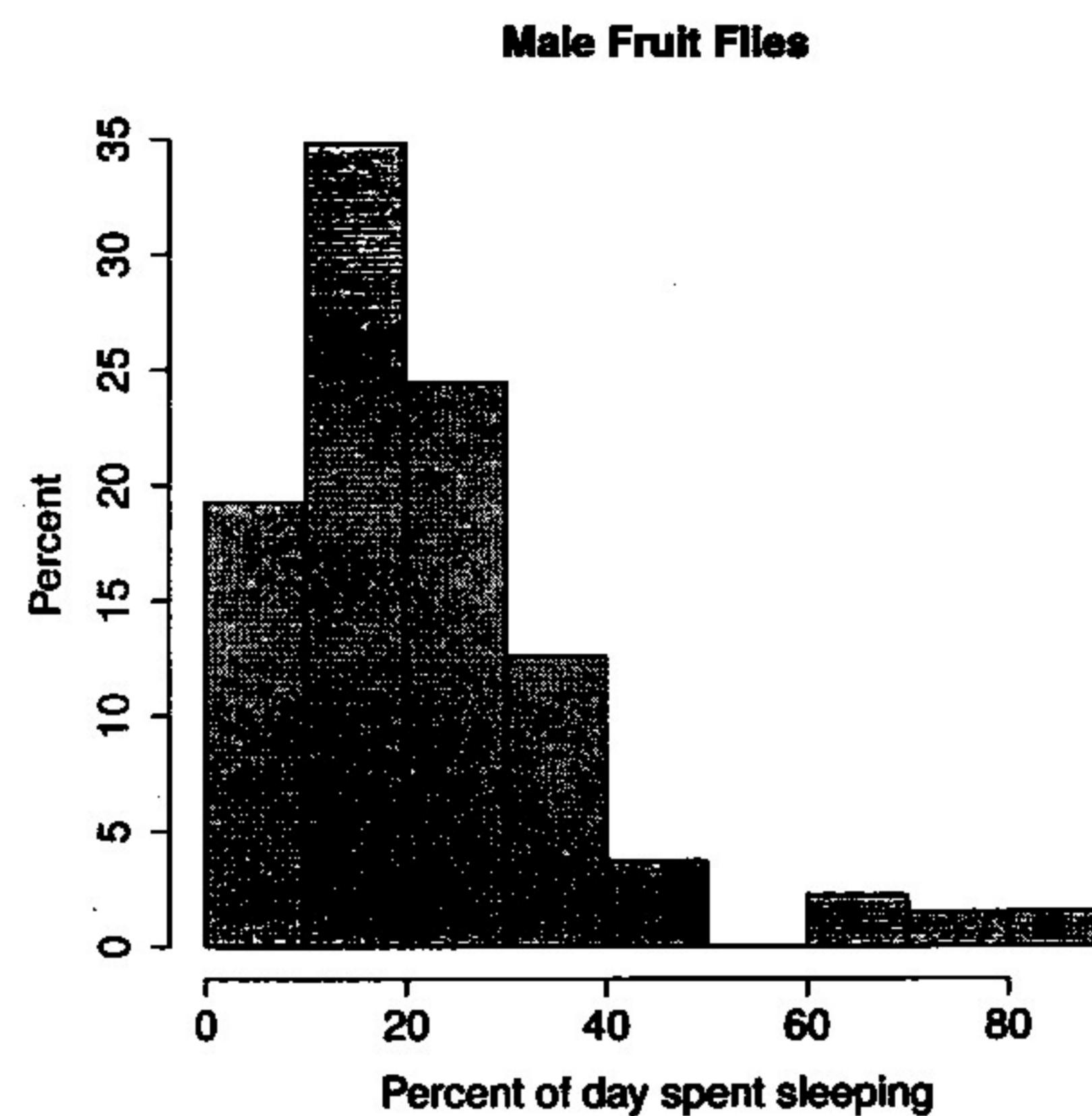
Q38

★ When a variable follows a normal distribution, what percent of observations are contained within 1.75 standard deviations of the mean?

- a. 4.01%
- b. 9.02%
- c. 68.26%
- d. 91.88%
- e. 95.99%

Q39

★ The plot below displays the distribution of the percent of days spent sleeping by male fruit flies. Which of the following are valid estimates of the mean and median of this distribution?



- a. mean = 24, median = 18
- b. mean = 18, median = 24
- c. mean = 18, median = 18
- d. mean = 20, median = 40
- e. mean = 25, median = 35

Q40

★ Which of the following statements about z-scores is/are true?

- a. larger z-scores are always better
- b. the z-score for an observation that is equal to the mean is 1
- c. if a z-score is 2 that means that the observation is two times the value of the mean
- d. if a z-score is negative that means that the observation is less than mean
- e. none of the above are true

Q41

- ★ A telephone survey is done to determine the average number of pets in the typical American family. Past experience has shown that 30% of those telephoned will refuse to respond to the survey. Which of the following scenarios satisfy the conditions for the binomial distribution?
- Number of families responding in 100 calls
 - Number of individuals in the family
 - Number of pets in a family that responds
 - Number of phone calls made until we first identify a family with a pet
 - All scenarios satisfy the conditions for the binomial distribution.

Q42

- ★ The distribution of rhesus monkey tail lengths is bell-shaped, unimodal, and approximately symmetric. The average tail length is 6.8 cm and the standard deviation is 0.44 cm. Roscoe has a tail that is 10.2 cm long. What conclusion can we make based on the information given?
- We can apply the empirical rule to conclude that Roscoe is a potential outlier because he falls more than three standard deviations away from the mean.
 - We can apply the empirical rule to conclude that Roscoe is not a potential outlier because he falls within three standard deviations away from the mean.
 - We cannot apply the empirical rule because the distribution does not fit the criteria for the empirical rule.
 - There is not enough information given to make any conclusions about potential outliers.

Q43

- ★ A political scientist is interested in the effect of government type on economic development. She wants to use a sample of 30 countries evenly represented among the Americas, Europe, Asia, and Africa to conduct her analysis. What type of study should she use to ensure that countries are selected from each region of the world?
- Observational - simple random sample
 - Observational - cluster
 - Observational - stratified
 - Experimental

Q44

- ★ Heights of boys in a high school are approximately normally distributed with mean of 175 cm standard deviation of 5 cm. What is the first quartile of heights?
- 25.0 cm
 - 165.8 cm
 - 171.7 cm
 - 173.5 cm
 - 181.0 cm

Q45

A (5 marks - 1 mark each)

Identify each variable as categorical, quantitative discrete, or quantitative continuous:

(also as ordinal, nominal, interval, or ratio)

each of these 3

and also all 4
of these

(a) Rating of the workload for a ~~Math 12510~~ course. (1=not enough, 2= could be more, 3=just right, 4=too much, 5=far too much)

in Celsius (°C)

(b) Temperature of water in a container 30 minutes after filling it with boiling water.

* (c) Distance a certain machine can throw a ball.

(d) Letter grade in BIO 101.

** (e) Number of people whose blood pressure is reduced by at least 20% taking a new treatment.

Q46

Heating and combustion analyses were performed in order to study the composition of moon rocks.

Find below the recordings of hydrogen and carbon in parts per million (ppm) for 11 specimens.

Hydrogen (ppm)	120	82	90	8	38	20	2.8	66	2.0	20	85
Carbon (ppm)	105	110	99	22	50	50	7.3	77	7.7	45	51

Summary statistics for " $x = \text{hydrogen}$ " and " $y = \text{carbon}$ " are:

$$n = 11 \quad \sum x_i = 533.8 \quad \sum y_i = 554.4 \quad \sum x_i^2 = 43124.8$$

$$\sum y_i^2 = 43203.3 \quad \sum x_i y_i = 39365.2 \quad s_x = 41.5 \quad s_y = 39.1$$

A (5 marks)

Give the 5-number-summary for the hydrogen contents in the rocks from the moon.

B

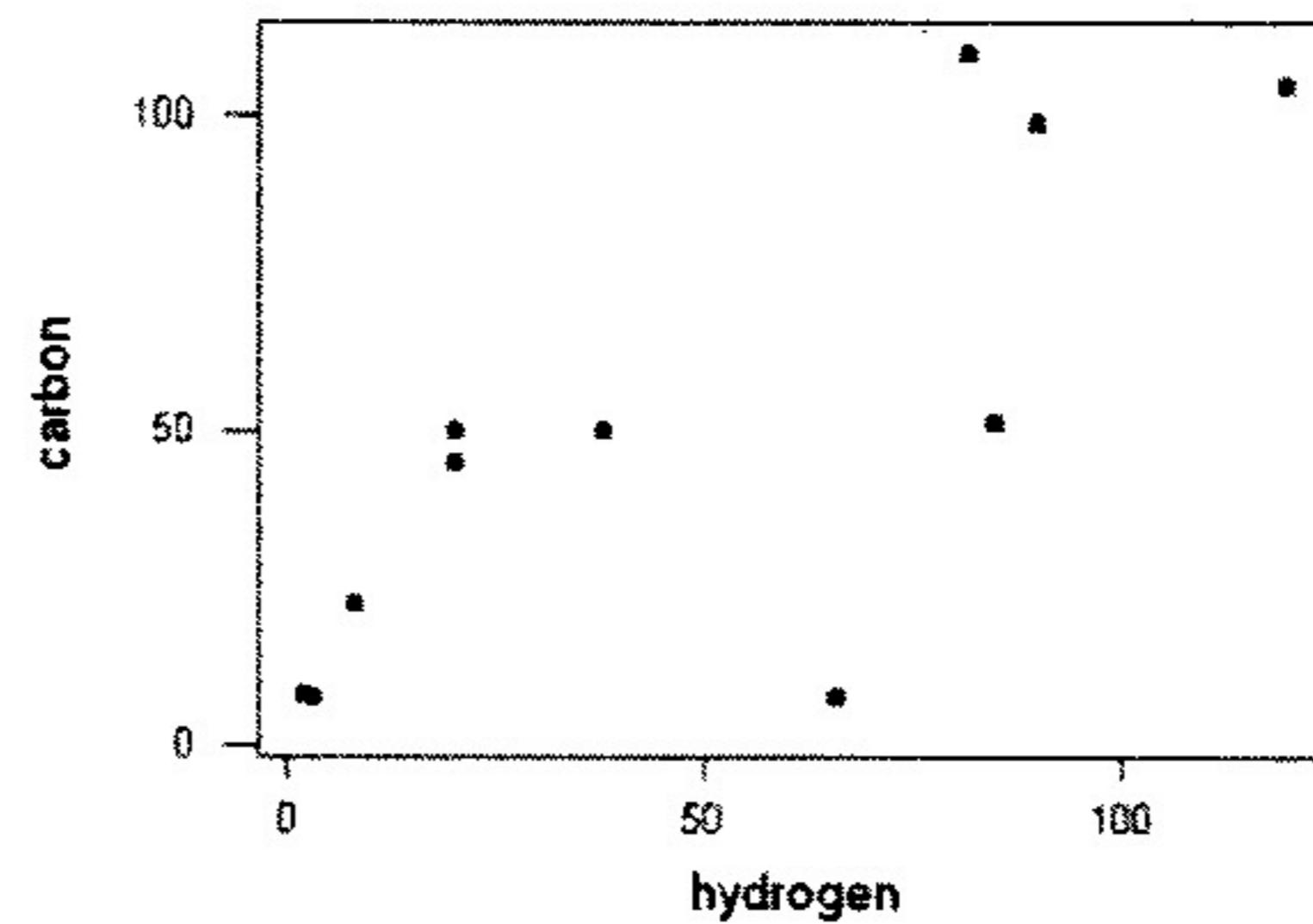
(5 marks)

"box-and-whiskers"

Obtain a boxplot for the hydrogen level in the rocks from the moon.

Q4(6 ctd)

(3marks) This scatterplot visualizes the relationship between the hydrogen and the carbon level of the 11 rocks from the moon.



★ Comment on the relationship.

★ (5 marks)

Calculate Pearson's Correlation Coefficient for the hydrogen and the carbon level in the 11 moon rocks.

Interpret your result. In which way is your result reflecting what can be seen in the scatterplot?

$$CV := 100 \times \left(\frac{\sigma}{\mu}\right)$$

~~(c) (4 marks)~~

The least-squares regression line is given through the equation
carbon = $15.3 + 0.724$ hydrogen (or $\hat{y} = 15.3 + 0.724x$).

Interpret the values of the intercept and the slope of the least-squares-regression line.

we haven't covered regression yet

~~(f) (4 marks)~~ Predict the carbon level in a rock from the moon with a hydrogen level of 75ppm using the regression line.

Is this a good prediction? Explain.

Q47

certain

{ $7\heartsuit, 7\spadesuit, 7\clubsuit, 7\diamondsuit$, etc. ...}

A stack of cards include **4** Sevens, **2** Jacks, **1** Queen, **3** Kings, and **2** Aces.

Consider the experiment of drawing one card from this deck and observing the category of the card.



A (2 marks) Give the sample space for this experiment.

Q47 (c,d)

- (3 marks) Let the event
 $A = \{ \text{Seven, Queen, Ace} \}$, to draw a Seven, a Queen, or a King,
and the event $B = \{ \text{Queen, King, Ace} \}$.

★

Sketch a Venn diagram visualizing the events A and B .

- (2 marks) Find the probability of event A , $P(A)$.

- (2 marks) Find $P(\text{Not } A) = P(A^c)$

★★

- (3 marks) Find ~~$P(A \cap B)$~~ . $P(A \cap B) = P(\text{"A and B"})$.

$$\frac{P(A \cap B)}{P(B)}$$

★★

- (4 marks) Find $P(A|B)$. Give an interpretation of this probability.

does $P(A) = P(A|B)$?

★★

- (4 marks) Are the events A and B independent? Justify your answer.

[Q48]

4. (4 marks) In my office I have a box with 5 red felt pens, of which 2 ran out of ink, and a box with 10 black felt pens, of which also 2 ran out of ink.

Before I go to class I randomly take one red and one black pen out of those boxes.

★ What is the probability that both pens have no ink left?

counting
independent
events... ★

use the
multiplication
rule

$$P(A)P(B) = P(A \cap B)$$

[Q49]

5. According to the National Vital Statistics Report, Vol.48, No.10, the probability that a randomly selected 40-year-old male will live to be 41 years old is 0.99718.

★ **A** (4 marks)

What is the probability that five randomly selected 40-year-old males will live to be 41 years old?

★

B

(4 marks)

What is the probability that at least one of five randomly selected 40-year-old males will not live to be 41 years old? Interpret your result.

Q50

- Q50. A) (4 marks) The tools to be used in data description depend on the type of the variable. Give the different possible types of (data) variables. Explain for each type, which parts an appropriate data description should include. (As many as you feel like.)

hint: You'll
be tested on
nominal/ordinal
and ratio/interval
though these are
very similar terms.

- * B) (2 marks) Explain the difference between a sample and a population.

- C) (3 marks) Give interpretations of probabilities 0.01, 0.5, and 1.2.

- (d) (3 marks) What is the purpose of a Regression Analysis?

- * D) (3 marks) Draw a Venn diagram illustrating the event ((not A) and B). = $A^c \cap B$

Q51

(4 marks - 1 mark each)

also as ordinal, ratio, interval, nominal, etc.

Identify each variable as categorical, numerical discrete, or numerical continuous:

* (a) Distance a person runs in one minute.

(b) Number of apples on a tree.

? (c) Rating of the performance of a candidate on a prelection debate.

* (d) Breed of dog.

he's a scientist.

AA Q52

The American Michelson set up many different trials to measure the velocity of light in air. For this work he later received the Nobel prize. One of his trials resulted in the following 18 sorted measurements in km/sec (299,000 km/s was subtracted from the measurements).

650 740 760 810 850 850 880 900 930
950 960 960 980 980 980 1000 1000 1070

You may use the following information on the data to answer the questions:

$$\sum_{i=1}^n x_i = 16250 \text{ and } \sum_{i=1}^n x_i^2 = 14869900.$$

A (4 marks)

Find the sample mean \bar{x} and the sample standard deviation s for those measurements.

B

(6 marks)

Give the five number summary for the measurements by Michelson.

C

(6 marks)

Draw a graph summarizing the measurements. ~~Comment. (Use the reverse of the paper for your answer.)~~

which graph is most appropriate?

Q53

The following table gives the probability distribution for a random variable X

$P(X=k)$	1	1	2	4	8	16
		0.1	0.3	0.25	0.1	?

(a) (1 marks) Find ~~$P(X=16)$~~ $P(X=16)$.

(b) (3 marks) Find the mean of the distribution. $E(X)$.

(c) (2 marks) Find $P(3 < X < 16) = P(X < 16) - P(X > 3)$. (why?)

Q54

$$\mu = 48.5$$

4. (4 marks) The average number of students in a first year course at Grant MacEwan College is 48.5 with a standard deviation of 6.3. $\sigma = 6.3$

Locate an interval that includes the number of students in approximately 95% of first year courses. Explain. Consider $Norm(\mu = 48.5, \sigma = 6.3)$.

Q55

In 2005, the average marathon time in the U.S. was 4.5 hours for men with a standard deviation of 1.1 hours. Assume that marathon times are normally distributed.

(a) (3 marks) Find the probability that a randomly chosen marathon runner in the U.S. (in 2008) had a time below 5 hours.

"normal
percentiles"

(b) (4 marks) Give the 20% fastest marathon times of men in the U.S. in 2008 (according to this model).

~~Q56~~

- ★ ★ 6. Psychologists tend to believe that there is a relationship between aggressiveness and order of birth.

The following table gives the outcome of a study of 500 students that were classified for their behaviour and their birth order

	Firstborn	Not Firstborn
Aggressive	75	75
Not Aggressive	125	225

Consider the experiment of choosing one student at random.

→ A (2 marks) What is the probability that the student is aggressive?

→ B (3 marks) What is the probability that the student is aggressive given he/she is a firstborn?

★ ★ → C (4 marks) Are the two events "aggressive" and "firstborn" independent? Justify your answer. Please! What's the definition of independent events?

~~already done~~ (a) (6 marks) In data description we have to distinguish between different types of variables/data and then choose the appropriate descriptive methods for each type. Give the different types of variables/data and for each type list which parts a proper data description should include.

(b) ~~(2 marks)~~ Sketch a Venn diagram showing two events A and B . Show $A \cap B^c$ in the diagram.

Q57

- ★ ★ (3 marks) Sketch a histogram that shows a data distribution for which the mean and the standard deviation are NOT appropriate descriptive measures.

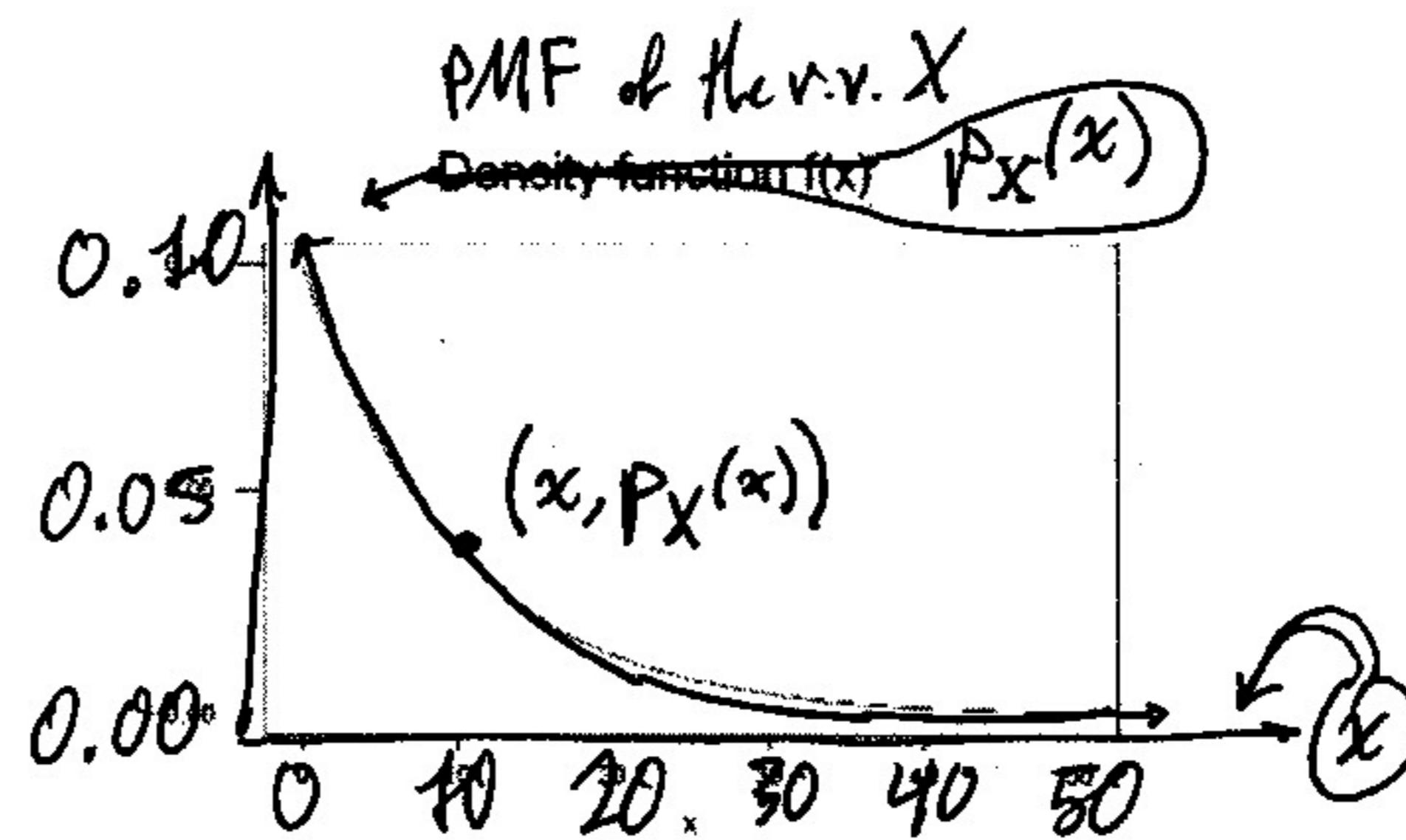
hint:
when does
the Empirical Rule fail?

- ★ (3 marks) In probability theory, what is a distribution? what's the PMF of a r.v.?

"P.M.F."

- ★ (3 marks) The following graph shows a probability density curve for a random variable X . Include $P(X \geq 20)$ in the diagram.

Draw it!



*challenge problems
(ask Cotton for solutions)*

6 Stat 110 Midterm from 2008

Q58

1. The gambler de Méré asked Pascal whether it is more likely to get at least one six in 4 rolls of a die, or to get at least one double-six in 24 rolls of a pair of dice. Continuing this pattern, suppose that a group of n fair dice is rolled $4 \cdot 6^{n-1}$ times.

(a) Find the expected number of times that "all sixes" is achieved (i.e., how often among the $4 \cdot 6^{n-1}$ rolls it happens that all n dice land 6 simultaneously). (Simplify.)

B)

Give a simple but accurate approximation of the probability of having at least one occurrence of "all sixes", for n large (in terms of e but not n).

C)

de Méré finds it tedious to re-roll so many dice. So after one normal roll of the n dice, in going from one roll to the next, with probability $6/7$ he leaves the dice in the same configuration and with probability $1/7$ he re-rolls. For example, if $n = 3$ and the 7th roll is $(3, 1, 4)$, then $6/7$ of the time the 8th roll remains $(3, 1, 4)$ and $1/7$ of the time the 8th roll is a new random outcome. Does the expected number of times that "all sixes" is achieved stay the same, increase, or decrease (compared with (a))? Give a short but clear explanation.

Q59

2. To battle against spam, Bob installs two anti-spam programs. An email arrives, which is either legitimate (event L) or spam (event L^c), and which program j marks as legitimate (event M_j) or marks as spam (event M_j^c) for $j \in \{1, 2\}$. Assume that 10% of Bob's email is legitimate and that the two programs are each "90% accurate" in the sense that $P(M_j|L) = P(M_j^c|L^c) = 9/10$. Also assume that given whether an email is spam, the two programs' outputs are conditionally independent.

A) Find the probability that the email is legitimate, given that the 1st program marks it as legitimate (simplify).

B) Find the probability that the email is legitimate, given that both programs mark it as legitimate (simplify).

C) Bob runs the 1st program and M_1 occurs. He updates his probabilities and then runs the 2nd program. Let $\tilde{P}(A) = P(A|M_1)$ be the updated probability function after running the 1st program. Explain briefly in words whether or not $\tilde{P}(L|M_2) = P(L|M_1 \cap M_2)$: is conditioning on $M_1 \cap M_2$ in one step equivalent to first conditioning on M_1 , then updating probabilities, and then conditioning on M_2 ?

Q60

2. Consider four nonstandard dice (the *Efron dice*), whose sides are labeled as follows (the 6 sides on each die are equally likely).

- A: 4, 4, 4, 4, 0, 0
- B: 3, 3, 3, 3, 3, 3
- C: 6, 6, 2, 2, 2, 2
- D: 5, 5, 5, 1, 1, 1

These four dice are each rolled once. Let A be the result for die A, B be the result for die B, etc.

A) Find $P(A > B)$, $P(B > C)$, $P(C > D)$, and $P(D > A)$.

B)

Is the event $A > B$ independent of the event $B > C$? Is the event $B > C$ independent of the event $C > D$? Explain.

Q61

4. The book *Red State, Blue State, Rich State, Poor State* (by Andrew Gelman) discusses the following election phenomenon: within any U.S. state, a wealthy voter is more likely to vote for a Republican than a poor voter; yet the wealthier states tend to favor Democratic candidates! In short: rich individuals (in any state) tend to vote for Republicans, while states with a higher percentage of rich people tend to favor Democrats.

A) Assume for simplicity that there are only 2 states (called Red and Blue), each of which has 100 people, and that each person is either rich or poor, and either a Democrat or a Republican. Make up numbers consistent with the above, showing how this phenomenon is possible, by giving a 2 by 2 table for each state (listing how many people in each state are rich Democrats, etc.).

B) In the setup of (a) (not necessarily with the numbers you made up there), let D be the event that a randomly chosen person is a Democrat (with all 200 people equally likely), and B be the event that the person lives in the Blue State. Suppose that 10 people move from the Blue State to the Red State. Write P_{old} and P_{new} for probabilities before and after they move. Assume that people do not change parties; so we have $P_{\text{new}}(D) = P_{\text{old}}(D)$. Is it possible that both $P_{\text{new}}(D|B) > P_{\text{old}}(D|B)$ and $P_{\text{new}}(D|B^c) > P_{\text{old}}(D|B^c)$ are true? If so, explain how it is possible and why it does not contradict the law of total probability $P(D) = P(D|B)P(B) + P(D|B^c)P(B^c)$; if not, show that it is impossible.

8 Stat 110 Midterm from 2010

Q62 A family has two children. The genders of the first-born and second-born are independent (with boy and girl equally likely), and which seasons the children were born in are independent, with all 4 seasons equally likely.

A) Find the probability that both children are girls, given that a randomly chosen one of the two is a girl who was born in winter (simplify).

B) Find the probability that both children are girls, given that at least one of the two is a girl who was born in winter (simplify).

Q63

4. Let $X_1, X_2, X_3, \dots, X_{10}$ be the total number of inches of rain in Boston in October of 2011, 2012, 2013, ..., 2020, with these r.v.s independent $\mathcal{N}(\mu, \sigma^2)$. (Of course, rainfall can't be negative, but μ and σ are such that it is extremely likely that all the X_j 's are positive.) We say that a *record value* is set in a certain year if the rainfall is greater than all the previous years (going back to 2011; so by definition, a record is always set in the first year, 2011).

A On average, how many of these 10 years will set record values? (Your answer can be a sum but the terms should be simple.)

B

Is the indicator of whether the year 2013 sets a record independent of the indicator of whether the year 2014 sets a record? (Justify briefly.)

C

Later in the course, we will show that if $Y_1 \sim \mathcal{N}(\mu_1, \sigma_1^2)$ and $Y_2 \sim \mathcal{N}(\mu_2, \sigma_2^2)$ are independent, then $Y_1 - Y_2 \sim \mathcal{N}(\mu_1 - \mu_2, \sigma_1^2 + \sigma_2^2)$. Using this fact, find the probability that the October 2014 rainfall will be more than double the October 2013 rainfall in Boston, in terms of Φ .

Q64

2. Two coins are in a hat. The coins look alike, but one coin is fair (with probability $1/2$ of Heads), while the other coin is biased, with probability $1/4$ of Heads. One of the coins is randomly pulled from the hat, without knowing which of the two it is. Call the chosen coin "Coin C".

(A) Coin C is tossed twice, showing Heads both times. Given this information, what is the probability that Coin C is the fair coin? (Simplify.)

(B) Are the events "first toss of Coin C is Heads" and "second toss of Coin C is Heads" independent? Explain briefly.

(C) Find the probability that in 10 flips of Coin C, there will be exactly 3 Heads. (The coin is equally likely to be either of the 2 coins; do not assume it already landed Heads twice as in (a). Do not simplify.)

Q65 5. Alice flips a fair coin n times and Bob flips another fair coin $n + 1$ times, resulting in independent $X \sim \text{Bin}(n, \frac{1}{2})$ and $Y \sim \text{Bin}(n + 1, \frac{1}{2})$.

(A) Let $V = \min(X, Y)$ be the smaller of X and Y , and let $W = \max(X, Y)$ be the larger of X and Y . (If $X = Y$, then $V = W = X = Y$.) Find $E(V) + E(W)$ in terms of n (simplify).

(B) Is it true that $P(X < Y) = P(n - X < n + 1 - Y)$? Explain why or why not.

(C) Compute $P(X < Y)$ (simplify). Hint: use (b) and that X and Y are integers.