Math 214 Statistics Spring 19 Midterm 1a

Name: Solutions

- I will count your best 8 of the following 10 questions.
- You may use a calculator.

1	10	
2	10	
3	10	
4	10	
5	10	
6	10	
7	10	
8	10	
9	10	
10	10	
	80	

Midterm 1	
Overall	

1. A study surveys Americans about smoking habits and produces the following data:

	sex	age	marital	income	smoke	amount
1	male	42	single	\$67,000	yes	10 cig/day
2	female	26	married	\$48,000	no	0 cig/day
⋮	⋮	⋮	⋮	⋮	⋮	⋮
1567	male	31	married	\$108,000	yes	3 cig/day

(a) How many participants does the survey have?

(b) Which variables are categorical and which are quantitative/numerical?

(c) How many observations are there? How many variables are there?

a) 1567

b) categorical : sex, marital, smoke

quantitative : age, income, amount

c) $1567 \times 6 = 9402$

6 variables

2. A teacher records the ages of everyone in a class of 15 students:

25, 18, 22, 23, 27, 26, 21, 19, 18, 29, 20, 22, 27, 33, 20

Find the:

(a) minimum:

(b) maximum:

(c) median:

(d) first and third quartiles:

(e) interquartile range:

26 29

18, 18, 19, 20, 20, 21, 22, 22, 23, 25, 27, 27, 33

M

a) 18

b) 33

c) 22

d) $Q_1 = 20$

$Q_3 = 26.5$

e) 6.5

3. Vitamin D is needed for the body to use calcium. Researchers plan to study the effects of calcium and vitamin D supplements on the bones of first-year college students. Three doses of calcium will be used: 0, 250, and 500 milligrams per day (mg/day). The doses of vitamin D will be 0, 75, and 150 international units (IU) per day. The calcium and vitamin D will be given in a single tablet. All tablets, including those with no calcium and no vitamin D, will look identical. Subjects for the study will be 45 men and 45 women, and each will be assigned at random to receive one of the calcium doses and one of the vitamin D doses. The outcome measure is the total body bone mineral content (TBBMC), a measure of bone health.

(a) Is this an experiment or an observational study?

(b) What are the explanatory variables (also known as factors)?

(c) What are the response variables?

(d) What are the treatments?

(e) Is there a placebo group? Explain your answer.

a) experimental

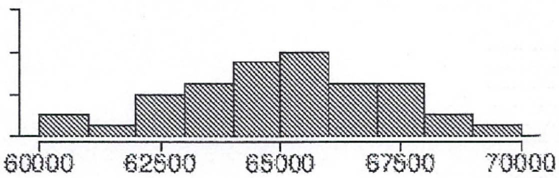
b) calcium, vitamin D, sex

c) bone mineral content

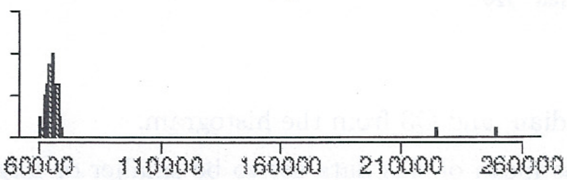
d) take dose of calcium (0, 250 or 500) and vitamin D (0, 75, or 150)

e) yes, those who receive 0 calcium and 0 vitamin D

4. The first histogram below shows the distribution of the yearly incomes of 40 patrons at a college coffee shop. Suppose two new people walk into the coffee shop: one making \$225,000 and the other \$250,000. The second histogram shows the new income distribution. Summary statistics are also provided.



(1)



(2)

	(1)	(2)
n	40	42
Min.	60,680	60,680
1st Qu.	63,620	63,710
Median	65,240	65,350
Mean	65,090	73,300
3rd Qu.	66,160	66,540
Max.	69,890	250,000
SD	2,122	37,321

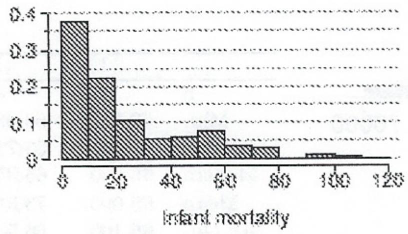
(a) Would the mean or the median best represent what we might think of as a typical income for the 42 patrons at this coffee shop? What does this say about the robustness of the two measures?

(b) Would the standard deviation or the IQR best represent the amount of variability in the incomes of the 42 patrons at this coffee shop? What does this say about the robustness of the two measures?

a) median : 65,240 , 65,350 only changes by a small amount
median is robust to outliers, mean is not robust.

b) IQR : 2,540 , 2,830 only changes by a small amount, so robust to outliers, s.d. is not robust.

5. The infant mortality rate is defined as the number of infant deaths per 1,000 live births. This rate is often used as an indicator of the level of health in a country. The relative frequency histogram below shows the distribution of estimated infant death rates for 224 countries for which such data were available in 2014.



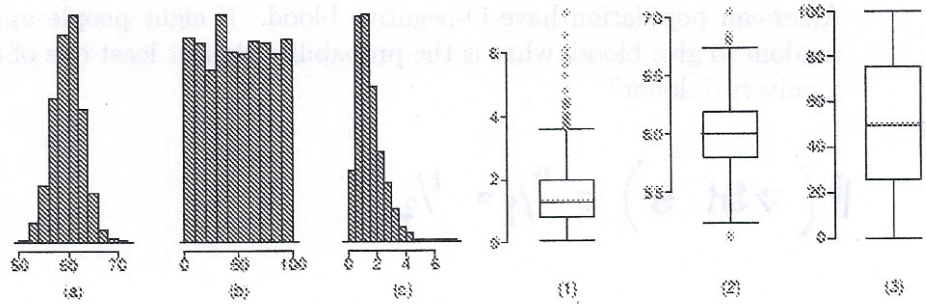
(a) Estimate Q1, the median, and Q3 from the histogram.

(b) Would you expect the mean of this data set to be smaller or larger than the median? Explain your reasoning.

a) $Q_1 = 10$ $M = 20$ $Q_3 = 40$

b) expect mean > median as data is right skewed.

6. Describe the distribution in the histograms below and match them to the box plots.



a) 2 symmetric, bell curve, centered at 60

b) 3 symmetric, uniform on 0, 100 centered at 60

c) 1 centered close to 2, right skewed / right tail

7. (a) You toss a coin three times. What is the probability that you get at least two heads?

(b) People with type O-negative blood are universal donors. That is, any patient can receive a transfusion of O-negative blood. Only 7% of the American population have O-negative blood. If eight people appear at random to give blood, what is the probability that at least one of them is a universal donor?

$\geq 2H$

a) HHH ✓
HHT ✓
HTH ✓
HTT
THH ✓
THT
TTH
TTT

$\mathbb{P}(\geq 2H \text{ 's}) = 4/8 = 1/2$

b) $\mathbb{P}(\text{at least 1 O-type}) = 1 - \mathbb{P}(\text{no O-type})$

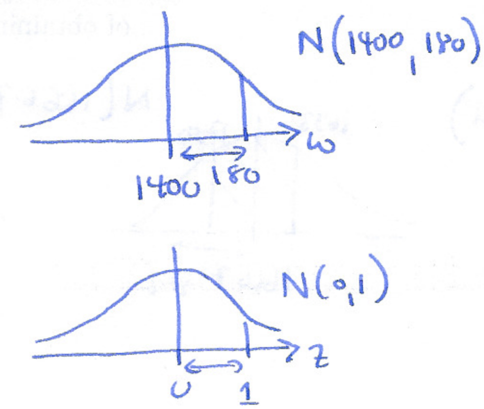
$= 1 - (0.93)^8 = 44\%$

8. Suppose that the weight of dairy cows is normally distributed with mean 1,400 pounds, and standard deviation 180 pounds.

   (a) What proportion of cows weigh more than 1,600 pounds?

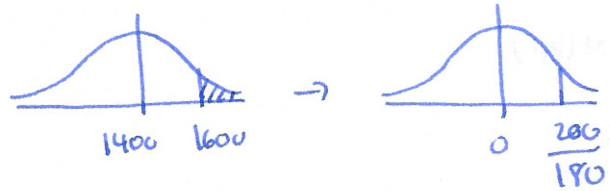
   (b) What proportion of cows weigh between 1,000 and 1,500 pounds?

   (c) What weight does a cow need to have to be in the bottom 20% of cows by weight?



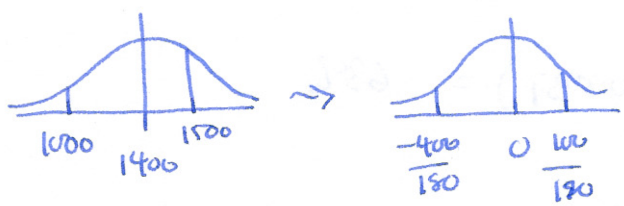
$z = \dfrac{w - 1400}{180}$

a)



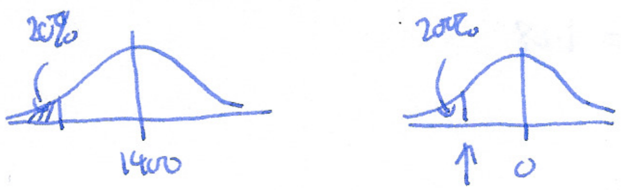
proportion = $1 - \text{pnorm}\left(\frac{200}{180}\right) = 13\%$

b)



proportion = $\text{pnorm}\left(\frac{100}{180}\right) - \text{pnorm}\left(\frac{-400}{180}\right) = 70\%$

c)



$\text{qnorm}(0.2) = -0.8416$

$w = 180 \cdot (-0.8416) + 1400$

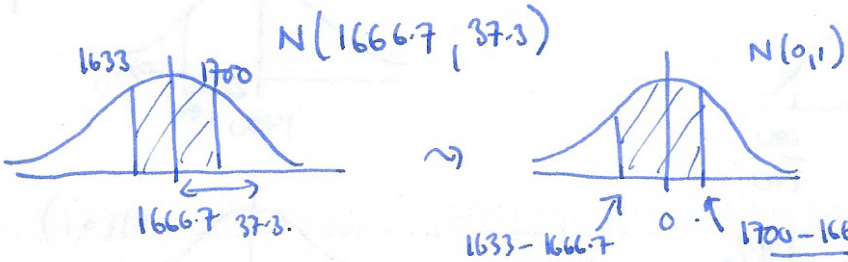
$= 1250$ pounds or less

9. A sample of 10,000 dice rolls is taken from the records of a casino. Assume that the number of sixes rolled is distributed by the binomial distribution Bin(10, 000, 1/6), which in this case can be closely approximated by the normal distribution $N(1, 666.7, 37.3)$.

(a) What is the probability that the number of sixes lies within 2% of 1,666?

(b) How many sixes would there need to be for there to be only a 10% chance of obtaining this many sixes?

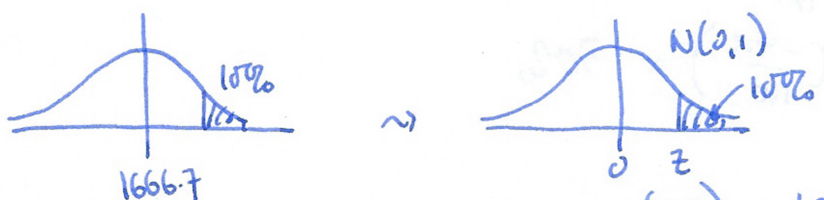
a)



$\frac{1633-1666.7}{37.3} = -0.8937$ $\qquad$ $\frac{1700-1666.7}{37.3} = 0.8937$

probability = pnorm(0.8937) − pnorm(−0.8937) = 63%

b)



qnorm(0.9) = 1.28

w = z * 37.3 + 1666.7 = ≥ 1715 sixes

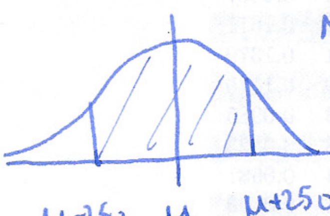
10. A neighborhood has 98,000 people. The annual income for people in this neighborhood has a standard deviation of \$5,000.

You choose a simple random sample of 1,200 people and measure the mean income $\bar{x}$ for the sample. You can assume that this sample is large enough to make the sampling distribution of $\bar{x}$ nearly normal. What is the chance that $\bar{x}$ deviates from the true mean income in the population by more than \$250?

population $N(\mu, 5000)$ sample size $n = 1200$

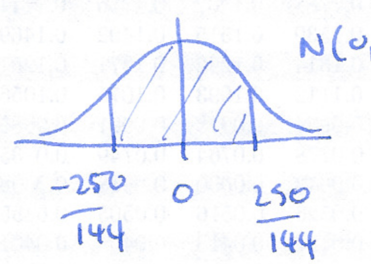
$\bar{x}$ sample mean distribution $N\left(\mu, \frac{5000}{\sqrt{1200}}\right)$, $\frac{5000}{\sqrt{1200}} = 144$

$N(\mu, 144)$



$\sim$

$N(0,1)$



$\text{pnorm}\left(\frac{250}{144}\right) - \text{pnorm}\left(\frac{-250}{144}\right) = 92\%$

so probability difference ≥ 250 is 8%.