2C misprint 17a - misprint Skip ~~17~~, 18, 28 Key
16 misprint 15

Statistics and Mathematical Modeling A Semester Exam Review (1)

Statistics – Final Exam Review

1. Students from a statistics class were asked to record their heights in inches. The heights were recorded as follows:

65	52	4	72	63	75	65	61
67	64	74	62	60	69	66	55
67	80	73	74	64	71	50	65

Put the data in a list on your calculator and use appropriate calculator functions to help you do the following: (Copy all graphs and show work on paper.)

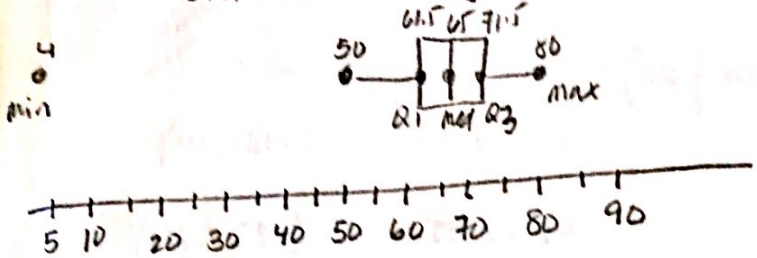
A. Make a box-and-whisker plot of the data.
B. Label the minimum, lower quartile (Q_1), median, upper quartile (Q_3), and maximum values.
C. Using the box-and-whisker plot, answer the following questions:
   1. The middle 50% of the students has heights between ___ and ___.
   2. 75% of the students have heights above _____.
   3. 25% of the students have heights below _____.
   4. What is the Interquartile Range (IQR)?
   5. Are there any data values that can be defined as outliers using the 1.5 x IQR rule? Show your work.
   6. What might be some explanations for the outlier(s)?
D. Create a stem and leaf plot for this data.
E. Omit the outlier from your list. Using the remainder of the data, create a frequency table.
F. From the frequency table, draw a histogram. Make your class width equal 6 and start the first class with 50.
G. What is the shape of the histogram? ("Approximately" normal, skewed right, skewed left?)
H. Using your table or histogram, what is the probability of a randomly selected student having a height between 65 and 79?
I. Find the mean, median, mode, standard deviation and range of the original data.
J. How would the mean, median, mode, range and standard deviation be affected if the 4 were change to a 60?

SAHM Semester Exam Review

Statistics Class Heights

1 a)

b)



c) i) 61.5" & 71.5"

ii) 61.5"

iii) 61.5"

iv) 10 (71.5 - 61.5)

v) yes 4

low = 61.5 - 1.5(10) = 46.5

low = Q1 - 1.5IQR

high = Q3 + 1.5IQR

vi) someone entered the info in incorrectly

d)

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

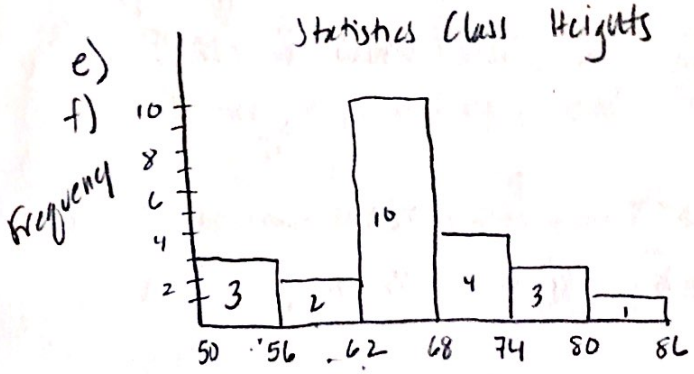
Key 5|0=50

e)

hts	freq
50-55	3
56-61	2
62-67	10
68-73	4
74-79	3
80-85	1

e)

f)



g) approx. normal

h) 13/23

i) $\bar{x} = 63.25$

median = 65

mode = 65

Stand dev = 14.5341

range = 76

j) $\bar{x}$, std. dev, range changes - others are not affected by outliers

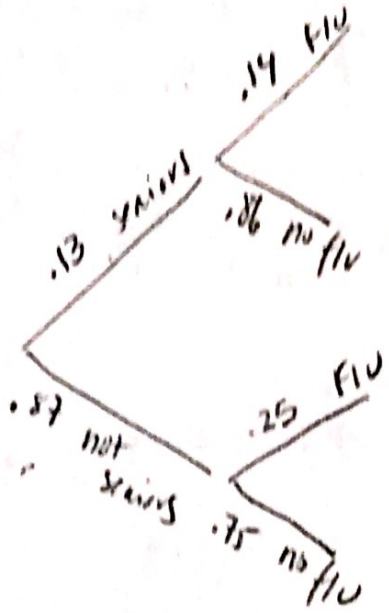
$\bar{x} = 65.5833$

Stand dev = 7.3064

range = 30

2.

a)

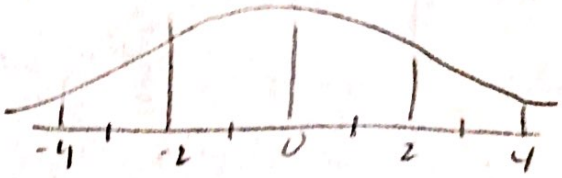


b) .0182 (.13 · .14)

c) .2175 (.87 · .25)

3. a)

b)



4. central tendency
   - mean
   - median
   - mode

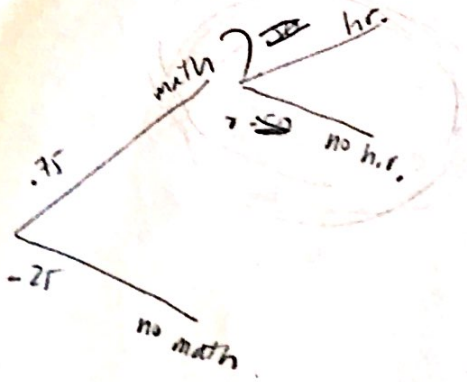
   spread
   - Stand. dev
   - IQR
   - r^2
   - range

5. ±4% means that pollsters are confident that the true percentage who favor this type of punishment is between 64% and 72%. The pollsters calculated a confidence interval to get this (don't know the level).

6. population - entire group/collection μ σ

   sample - subset of population $\bar{x}$ s_x

7.



$$P(HR \mid math) = \frac{P(HR \cap math)}{P(math)}$$

$$= \frac{.50}{.75} = .67 \quad 2/3$$

8 a) P our parameter of interest is the actual percentage of kids that got 1200 on SAT

A SRS ✓ $np \geq 10$ $n(1-p) \geq 10$ Population $\geq 10n$ all kids take SAT $\geq 10(125)$ ✓

$125\left(\frac{100}{125}\right) \geq 10$ ✓

$125\left(\frac{25}{125}\right) \geq 10$ ✓

N one prop. z interval

I $\hat{p} \pm z^* \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$ $.8 \pm 1.96\sqrt{\frac{.8(.2)}{125}} = [.7299, .8701]$

C we are 95% confident that the actual % of kids who get 1200 on SAT lies between .7299 and .8701.

b) H_0: $p = .85$ H_a: $p < .85$

c) Since .85 lies in the 95% confidence interval we will fail to reject Ho at the 2.5% significance level. So the actual percentage is 85%.

d) P we want to test claim about actual percentage who get 1200 on SAT

H. H_0 $p = .85$ H_a $p < .85$

A ✓

N 1 proportion z test

T $z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} = -1.1656$

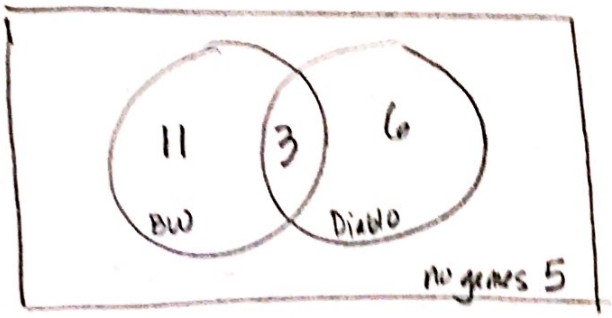
O. pvalue = .1175

M since .1175 > .05 we fail to reject

So the percentage is 85%.

e) no!

9.

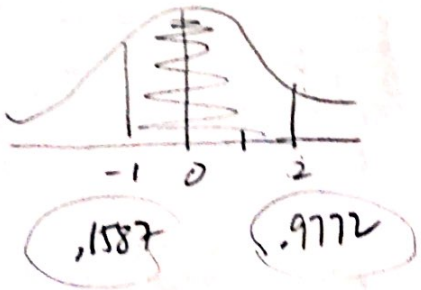


a) $\frac{5}{25} = .2$

b) $\frac{17}{25} = .68$

10. wider - as confidence level increases, margin of error increases ∴ it is wider (z^* is larger)

11.



.9772 - .1587 =

.8185

normalcdf(-1, 2, 0, 1)

12. P our parameter of interest is the true proportion of Americans who feel too much violence on TV

A SRS✓ $320\left(\frac{243}{320}\right) \geq 10$✓ $320\left(\frac{77}{320}\right) \geq 10$✓ Population of TV watchers $\geq 10(320)$✓

N one proportion z interval

I $\frac{243}{320} \pm 1.96\sqrt{\frac{\frac{243}{320}\left(\frac{77}{320}\right)}{320}} = [.7125, .8062]$

C we are 95% confident that the true proportion of Americans who feel there is too much violence on TV lies between .7125 and .8062.

13. I SRS - least bias
- each person is equally likely to be chosen
- independent

II Systematic random sampling - order participants & then choose every 5th person.

Stratified random sampling - divide population into subgroups (male/female, seniors/juniors) with a similar characteristic & take a random sample from each subgroup.

Cluster sampling - pick entire cluster
- Choose entire class; all houses on one street; all the remotes in 1 box.

III Convience - most bias

volunteer - self selected. call in radio/tv show

response bias - the testers for the produce department selects good looking oranges from a bushel to determine if truckload is good or not

14. I can be done - impossible to name all shoppers in the mall

II systematic - stand at entrance and ask every 5th shopper

Stratified - randomly ask only females under 25

Cluster - randomly ask shoppers in Macy's shoe department

III volunteer - set up a booth & ask for volunteers to answer survey

response bias - ask 100 shoppers that interviewer feels will "represent" population

15. ✗

16. $\frac{730}{2000} = .365$

17. a) $\frac{4}{52} + \frac{13}{52} - \frac{1}{52} = \frac{16}{52} = .3077$

b) no

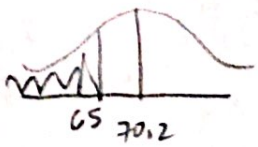
18. a) $\frac{1}{52} \cdot \frac{1}{52} = \frac{1}{2704}$

b) no

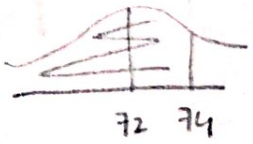
19. $\frac{1}{8}$ (1:7)

20. a) .1078

normalcdf(-1e99, 65, 70.2, 4.2)



b)



$z = \frac{74 - 70.2}{4.2} = .9048$

21. $\sigma = s_x/\sqrt{20}$ $\mu = \bar{x}$

22. P our parameter of interest is the true percentage of voters who favor extending the school yea

A SRS ✓ $np \geq 10$, $n(1-p) \geq 10$ Population voters $\geq 10(1235)$

N one proportion z interval

I $\frac{935}{1235} \pm 1.96\sqrt{\frac{.7571(.2429)}{1235}} = [.7332, .781]$

C We are 95% confident that the true percentage of voters who favor extending the school year lies between

23. As the sample size increases, the sampling distribution becomes normal. ~~The mean~~

$\mu = \bar{x}$ $\sigma = \frac{s}{\sqrt{n}}$

24. 1) easy to see skewness
2) easy to compare boxplots
3) outliers are obvious
4) 5 number summary

25. 1) mode
2) shape of distribution is easy to see
3) outliers are easy to see

26. $6! = 720$ ways

27. 68% - 95% - 99.7%

28. ${}_{12}C_5 = 792$

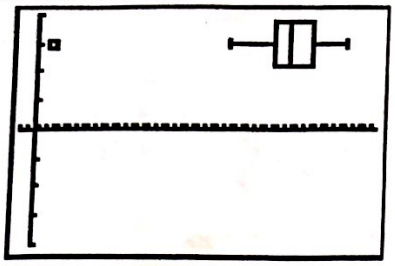
29.

card	1	2	3	4	5
P(card)	.2	.1	.1	.1	.5

expected value = 3.6

$1(.2) + 2(.1) + 3(.1) + 4(.1) + 5(.5)$

1. A) Modified Box Plot



B) minimum = 50, Q1 = 61.5, median = 65, Q3 = 71.5, Max = 80

C) 1. 61.5 and 71.5
2. 61.5
3. 61.5
4. 10
5. Outliers should be less than 61.5-1.5(10)= 46.5 or greater than 71.5+1.5(10) = 71.5. Therefore 4 is an outlier
6. Student might have recorded a height in feet instead of inches.

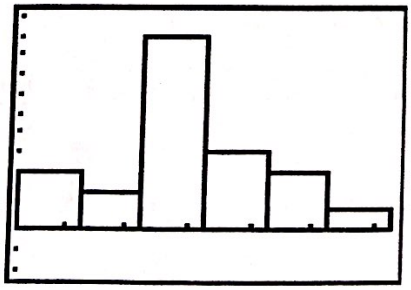
D) 5|2 represents 52

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

E)

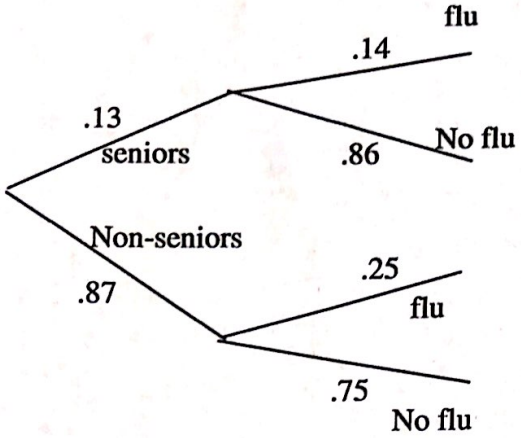
Heights	Frequencies
50-55	3
56-61	2
62-67	10
68-73	4
74-79	3
80-85	1

F)



G) Approximately normal
H) 17/23
I) mean:63.25, median: 65, mode: 65, standard deviation: 14.53, range: 76
J) mean: 65.58, median 65, mode: 65, standard deviation: 7.3, range: 30

2. A)



B) P(seniors and flu) = .13(.14) = .0182
C) P(no seniors and flu) = .87(.25) = .2175
D) P(Flu) = .0182/.2175 = .0772

3.

a) b)

8. A) $.7299 < p < .8701$

B) $Ho: p = .85$

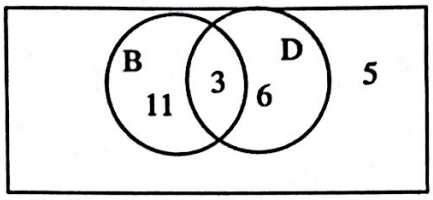
$Ha: p \neq .85$

C) There is a 95% chance that the confidence interval will contain the true population proportion. Since .85 is within the confidence interval above, , there is not sufficient evidence to reject the null hypothesis.

D) Test statistic $z = -1.57$. For a 5% confidence interval, to reject the null hypothesis $z > 1.96$ or $z < -1.96$. Therefore we fail to reject the null hypothesis. Or by p-analysis $p = .1175 > .05$

E) We conclude there is insufficient evidence that the board is exaggerating.

9.



A) 5/25 = 1/5 = .2

B) 17/25 = .68

10. The 93% interval is wider. It produces a larger z value on which the width of the confidence interval is proportional.

11. .8185

12. $.712 < p < .806$

13. I Simple Random Sample – Least Bias

a) Each member of the population is equally likely to be chosen.

b) The members are chosen independently.

II Probability Sampling Methods

1) Systematic random sampling: Order the members of the population and choose every nth member from then on. Example: Go down the rows of your class and select every third student.
2) Stratified Random Sampling: Divide the population into strata (subgroups with a similar characteristic) and take a random sample from each subgroup. Example: strata divisions could be boy/girl, juniors/seniors, professional/support staff.
3) Cluster Sampling: Choose members from clusters (not individually). Example: Choose members from 3rd period class, house on one block, or all workers on the third floor.

III Convenience Sampling: Most Bias
a) Self selected: when members participate in a survey voluntarily. Example: mail a questionnaire and request members to complete it.
b) Judgement: An expert selects a sample to be representative of the whole. Example: Select oranges from one bushel to determine the quality of the complete truckload of oranges.

14. I Cannot be done. It is impossible to name every shopper in the mall.
II 1. Stand at one entrance and ask every 5th shopper.
2. Ask only female shoppers under the age of 25.
3. Randomly select shoppers in the shoe department at Hecht's.
III. 1. Set up a booth with a sign and ask shopper to voluntarily come up and answer the question.
2. Ask 100 shoppers that he believes would fairly represent the population and use that for his decision.

15. Vertical or horizontal scale is too large or too small to accurately display data
Parts of the scale on the axes may be omitted.
The scale may be written in reverse order to disguise a decline.

16. .365

17. A) 16/52 = .31 B) No

18. A) 1/2704 B) Yes

19. 1:7

20. A) .108 B) .905

21. The mean of the sampling distribution will be the mean of the population. The standard deviation of the sampling distribution will be the standard deviation of the population divided by the square root of the sample size.

22. $.733 < p < .781$

23. As the sample size increases, the sampling distribution becomes more and more normal, with mean equal to the population mean and standard deviation equal to the population standard deviation divided by the square root of the sample size.

24. Skewness will be evident, easy to compare multiple boxplots, the 5-number summary comes straight from the box-plot, outliers are evident.

25. Shape of the distribution is evident, outliers are evident, individual data points are seen.

26. $6! = 6x5x4x3x2x1 = 720$ ways

27. The empirical rules states that in a normal distribution, about 68% of the data lie with 1 standard deviation of the mean, about 95% with 2 s.d. and about 99.7% within 3 s.d.

28. 792 ways

29. $.2(1) + .1(2+3+4) + .5(5) = 3.6$