

AP STAT- Ch. 3 -- 5 Quiz Review

- 1) A survey of automobiles parked in the student and staff lots at a large university classified the brands by country of origin, as seen in the table below:

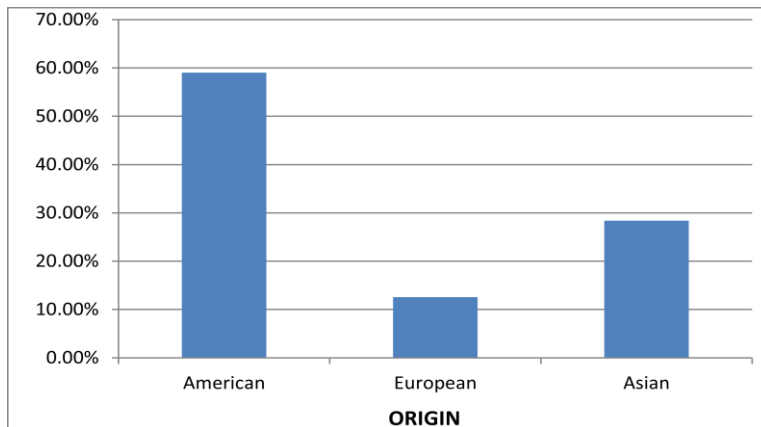
		Driver		
		Student	Staff	TOTAL
Origin	American	107	105	212
	European	33	12	45
	Asian	55	47	102
	TOTAL	195	164	359

- A) What is the marginal distribution of Origin? Make a bar graph.

American = 59.05%

European = 12.53%

Asian = 28.41%



- B) What is the marginal distribution of Driver? Do not make a bar graph.

Student = 54.317%

Staff = 45.682%

- C) What percent of Students drove Asian cars?

$P(\text{Asian}|\text{Student}) = 107/195 = 54.872\%$

- D) What percent of Asian cars are driven by staff?

$P(\text{Staff}|\text{Asian}) = 47/102 = 46.078\%$

- E) What percent of Staff drove Asian cars?

$P(\text{Asian}|\text{Staff}) = 47/164 = 28.659\%$

- F) What percent of those surveyed were Students?

$P(\text{Students}) = 54.317\%$

- G) What percent of those surveyed drove American cars or were students?

$P(\text{American} \cup \text{Students}) = (55+33+107+105)/359 = 83.565\%$

- H) What percent of those surveyed drive European cars and were staff?

$P(\text{European} \cap \text{Staff}) = 12/359 = 3.343\%$

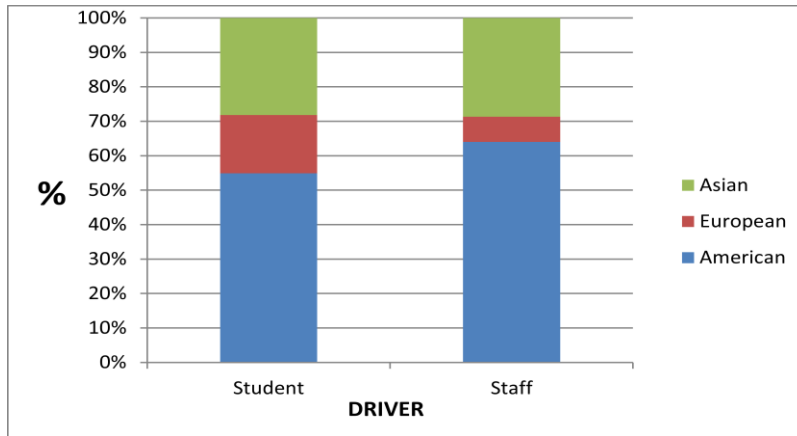
- I) What is the conditional distribution of Origin?

	American	European	Asian
Student	50.472%	73.333%	53.922%
Staff	49.528%	26.667%	46.078%

- J) What is the conditional distribution of Driver?

	Student	Staff
American	54.872%	64.024%
European	16.923%	7.317%
Asian	28.205%	28.659%

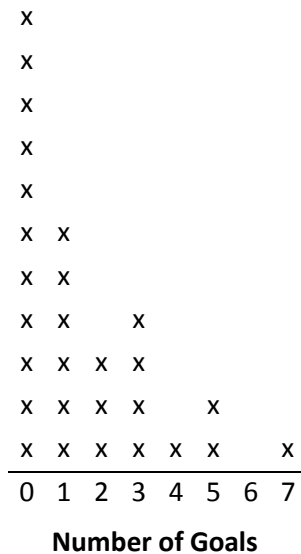
K) Create a segmented bar chart for the conditional distribution of Driver.



L) Is there an association between Origin and Driver? Provide statistical evidence to support your claim.
There DOES appear to be an association (the variables appear to be DEPENDENT). This is shown in the stacked bar graph above. There appear to be DIFFERENT percentages of car origins for the different types of drivers. Staff seems to drive more American cars and less European cars than Students do. However it seems that both Students and Staff drive the same percentage of Asian cars.

2) Create a dotplot of the number of goals scored by each team in the first round of the California high school soccer playoffs. Then briefly describe the distribution.

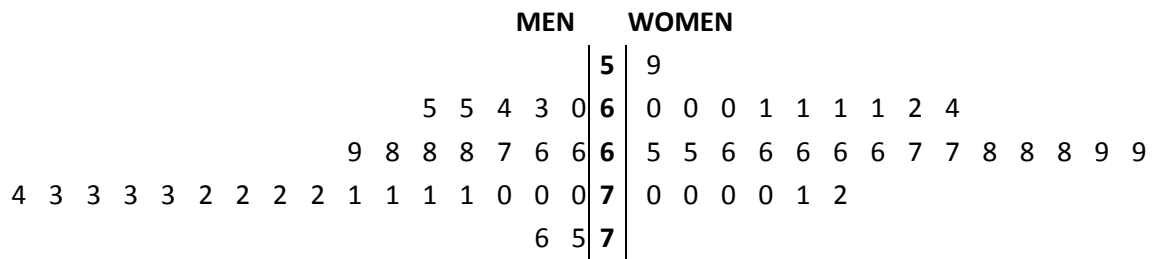
5	0	1	0	7	2	1	0	4	0	3	0	2	0
3	1	5	0	3	0	1	0	1	0	2	0	3	1



SHAPE: unimodal, right skewed
CENTER: Median of 1 goal
SPREAD: (0, 7)

3) Create back-to-back stemplots of the following male and female heights. Compare & describe both distributions

MALE	FEMALES
72 75 66 76 70 71	72 69 70 64 70 66
73 74 65 73 73 66	70 60 71 65 61 67
73 68 65 63 72 68	66 59 70 66 69 68
70 68 70 64 72 69	61 61 60 66 68 68
71 67 71 60 71 72	60 62 61 66 67 65



SHAPES: both mens and womens distributions are unimodal. Mens distribution is left skewed while women's distribution is roughly symmetric.

CENTERS: Men's center is the median of 70.5 which is higher than the women's mean is 65.6.

SPREAD: The men's spread is (60, 76) which is similar in spread to the women's spread of (59, 72).

4) Find the 5# summaries and create parallel boxplots for the heights of males and females in question #3

MEN:

Min: 60

Q1: 67

Med: 70.5

Q3: 72

Max: 76

WOMEN:

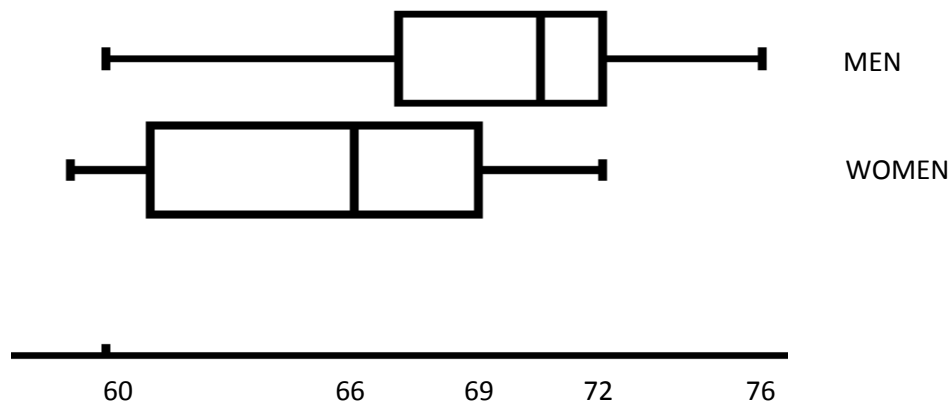
Min: 59

Q1: 61

Med: 66

Q3: 69

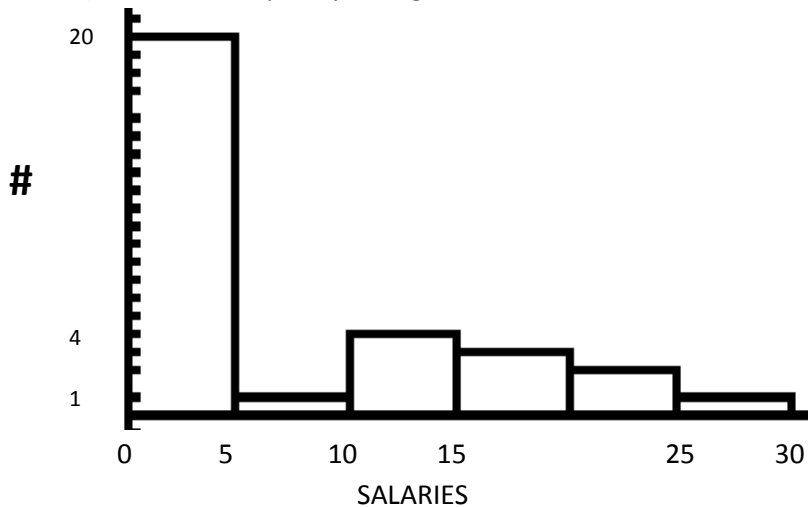
Max: 72



5) Salaries of 2008 New York Yankees (in millions of dollars):

Rodriguez	28	Giambi	23.428
Jeter	21.6	Abreu	16
Petite	16	Rivera	15
Posada	13.1	Damon	13
Matsui	13	Mussina	11.071
Pavano	11	Farnsworth	5.917
Wang	4	Hawkins	3.75
Cano	3	Molina	1.875
Ensberg	1.75	Brackman	1.185
Betemit	1.165	Bruney	0.725
Traber	0.500	Cabrera	0.461
Hughes	0.406	Duncan	0.398
Henn	0.397	Kennedy	0.394
Karstens	0.393	Albaladejo	0.393
Ohlendorf	0.391	Chamberlain	0.390
Sanchez	0.390		

A) Create a frequency histogram of the data above. Describe the distribution.



Shape: Right skewed, unimodal

Center: Median of 1.75 million dollars

Spread: range of (0.39, 28) and an IQR of 12.602

B) Based on this description, what measure of center and spread should you report?

Since it is right skewed, we should report Median and IQR and Range

C) Find the mean, standard deviation, 5# summary, and IQR

Mean = 6.42

Min = 0.39

Std. Dev = 8.12

Q1 = 0.398

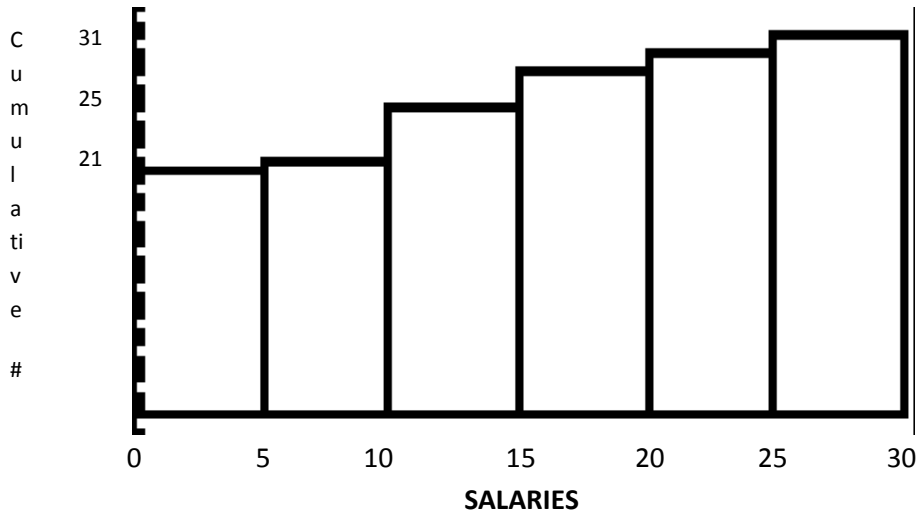
Med = 1.75

Q3 = 13

Max = 28

IQR = 12.602

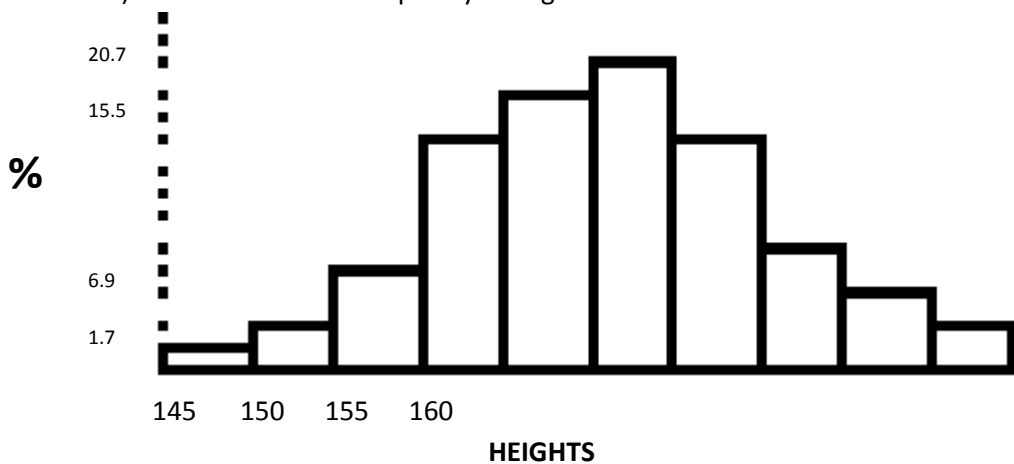
D) Create a cumulative frequency histogram.



6) Heights (in cm) of 58 randomly selected Canadian students who participated in a survey

166.5	170	178	163	150.5	169	171	166	190	183	178	161
171	170	191	168.5	178.5	173	175	160.5	166	164	163	174
173	169	160	174	182	167	166	170	170	181	171.5	160
178	157	165	187	168	157.5	145.5	156	182	168.5	177	162.5
160.5	185.5	151	159	177	171	176	177	181	186		

A) Create a relative frequency histogram of the data. Describe the distribution.



Shape: roughly symmetric, unimodal

Center: Mean of 170.21

Spread: std. deviation of 9.9 and range of (145.5, 191)

B) Based on this description, what measure of center and spread should you report?

Since the distribution is roughly symmetric, we should report the mean and standard deviation.

C) Find the mean, standard deviation, 5# summary, and IQR

Mean = 170.21

Min = 145.5

Q3 = 177

Std. Dev = 9.9

Q1 = 163

Max = 191

IQR = 14

Med = 170

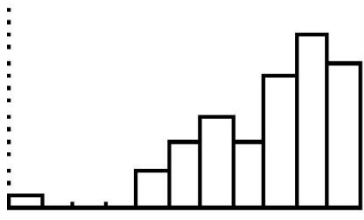
7) Use the following data. {30, 30, 30, 30, 30, 30, 30, 30, 30}. Find the mean and standard deviation. Why is the standard deviation this value?

Mean = 30

Std. Deviatin = 0

The standard deviation is 0 because all the values are the same. The data does not “deviate” from the mean at all. So the average deviation = 0.

8) Describe the following distributions using the terms we learned in class. Scale on x-axis: (1, 12), bins = 1

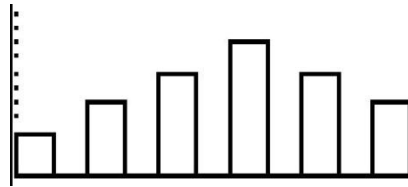


Shape: unimodal, left skew

Center: approx. 8

Spread: (5, 11)

Outlier @ 1

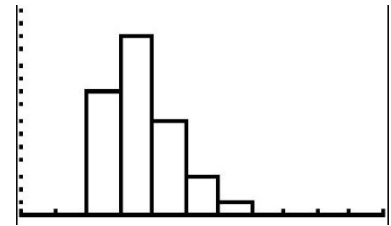


Shape: unimodal, symmetric

center: approx. 7

spread: (1, 11)

granularity

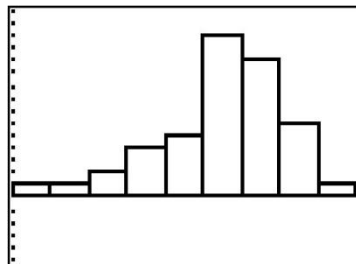


shape: unimodal, right skewed

center: approx. 4

spread: (3, 7)

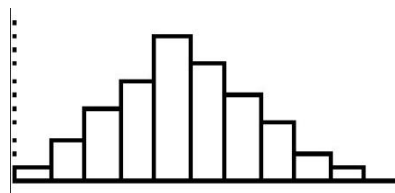
clustered



Shape: left skewed, unimodal

Center: approx. 5

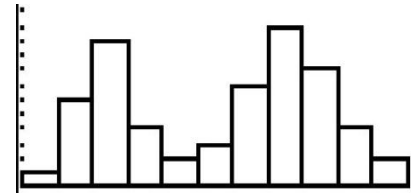
Spread: (1, 10)



shape: symmetric, unimodal

center: approx. 5

spread: (1, 10)



shape: bimodal, symmetric

center: approx. 6

spread: (1, 11)

9) Use the following data: {20, 23, 24, 27, 29, 31, 30, 33, 36, 37, 35, 40}

A) Calculate the following statistics:

Mean	30.42
Median	30.5
Range	(20, 40) = 20 units
IQR	10
Std. Dev.	6.127

B) Suppose we now add a new point to the data set: 60. Indicate whether adding the new point to the rest of the data made each of the summary statistics in part (a) increase, decrease, or stay about the same

Increase = mean, std. deviation, range

Same = median, IQR

10) A random sample of the heights of 24-34 year old women was taken (in inches). The following summary statistics were calculated.

Statistic	mean	st. dev.	min	Q ₁	med	Q ₃	max
Heights of 24-34 year old women	69.5	2.65	58	62	64	68	78

A) Based on the summary statistics would you describe the distribution as symmetric or skewed? Explain.

I would say the data are skewed because the mean is significantly greater than the median

B) Are there any outliers present? Show all work.

$$\text{IQR} = 68 - 62 = 6$$

$$1.5 \times \text{IQR} = 9$$

$$\text{UF} = \text{Q}_3 + 9 = 77$$

$$\text{LF} = \text{Q}_1 - 9 = 53$$

Anything outside (53, 77) is considered an outlier, so 78 is an outlier.

OR

$\text{Mean} \pm 2s = 69.5 \pm (2 \times 2.65) = (64.2, 74.8)$ Anything outside this range is an outlier, so 78 is an outlier.