

Name: _____ Date: _____

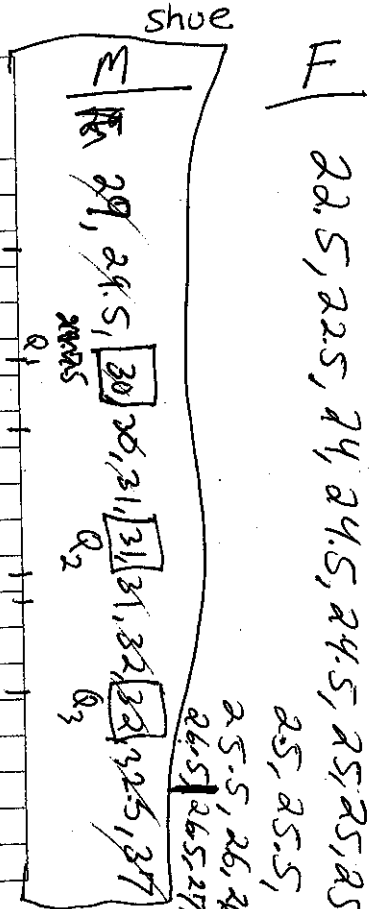
Learning Task: If the Shoe Fits!

MCC9-12.S.ID. 1 Represent data with plots on the real number line (dot plots, histograms, and box plots).
 MCC9-12.S.ID. 2 Use statistics appropriate to the shape of the data distribution to compare center (median, mean) and spread (interquartile range, mean absolute deviation) of two or more different data sets.
 MCC9-12.S.ID. 3 Interpret differences in shape, center, and spread in the context of the data sets, accounting for possible effects of extreme data points (outliers).

Welcome to CSI at School! Over the weekend, a student entered the school grounds without permission. Even though it appears that the culprit was just looking for a quiet place to study undisturbed by friends, school administrators are anxious to identify the offender and have asked for your help. The only available evidence is a suspicious footprint outside the library door.

After the incident, school administrators arranged for the data in the table below to be obtained from a random sample of this high school's students. The table shows the shoe print length (in cm), height (in inches), and gender for each individual in the sample.

Shoe Print Length	Height	Gender	Shoe Print Length	Height	Gender
24	71	F	24.5	68.5	F
32 ✓	74	M	22.5	59	F
27	65	F	29 ✓	74	M
26	64	F	24.5	61	F
25.5	64	F	25	66	F
30 ✓	65	M	37 ✓	72	M
31 ✓	71	M	27	67	F
29.5 ✓	67	M	32.5 ✓	70	M
29	72	F	27	66	F
25	63	F	27.5	65	F
27.5	72	F	25	62	F
25.5	64	F	31 ✓	69	M
27	67	F	32 ✓	72	M
31 ✓	69	M	27.4	67	F
26	64	F	30 ✓	71	M
27	67	F	25	67	F
28	67	F	26.5	65.5	F
26.5	64	F	30	70	F
22.5	61	F	31	66	F
			27.25	67	F



- Explain why this study was an observational study and not an experiment.
 • Didn't set anything up, just observed characteristics of random students
- Why do you think the school's administrators chose to collect data on a random sample of students from the school? What benefit might a random sample offer?
 Random would give a good idea about entire school
- Suggest a graph that might be used to compare the shoe print length data distributions for females and males.
 Double histogram, Double box & whisker
- Describe one advantage of using comparative box plots instead of comparative dot plots to display these data.
 Box plot allows you to see ~~the~~ middle values (Q1, Q2, Q3)

group

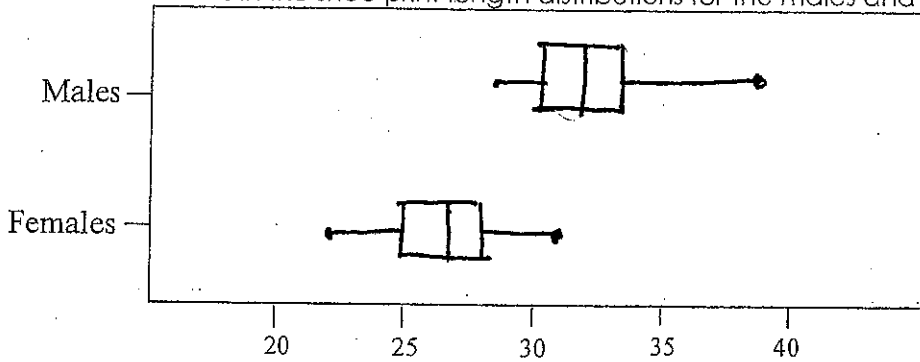
5. For each gender calculate the five-number summary for the shoe print lengths. Additionally, for each gender, determine if there are any outlying shoe print length values.

Group 2

	Minimum	Quartile 1 (Q1)	Median (Q2)	Quartile 3 (Q3)	Maximum
Male	29	30	31	32	37
Female	22.5	25	26.5	27.5	31

6. Construct comparative box plots for the shoe print lengths of males and females. Discuss the similarities and differences in the shoe print length distributions for the males and females in this sample.

Group 3



Group 4

7. For each gender calculate the mean shoe print length. What information does the mean shoe print length provide?

The average of girls & boys are provided by calculating the mean.
 girls $\bar{x} = 26.32$ boys $\bar{x} = 31.36$

Group 5

8. The mean will give us an indication of a typical shoe print length. In addition to knowing a typical length we would also like to know how much variability to expect around this length. For each gender calculate the **Range**; **Interquartile Range**; and **Mean Absolute Deviation** of the shoe print lengths. Interpret each of the calculated values.

	Range	IQR	M.A.D.
Male	$37 - 29 = 8$	$32 - 30 = 2$	≈ 1.46
Female	$31 - 22.5 = 8.5$	$27.4 - 25 = 2.4$	≈ 1.47

Group 6

9. If the length of a student's shoe print was 32 cm...

A. Would you think that the print was made by a male or a female?

Male

B. How sure are you that you are correct? Explain your reasoning. Use results from Questions 5 through 8 in your explanation.

No females had a shoe size of 32 in our random sample

10. How would you answer Question 9 if the suspect's shoe print length was 27 cm?

If 27 was the shoe size I would assume female based on our random sample

Name: _____ Date: _____

Two-way Frequency Charts

MCC9-12.S.ID.5 Summarize categorical data for two categories in two-way frequency tables. Interpret relative frequencies in the context of the data (including joint, marginal, and conditional relative frequencies). Recognize possible associations and trends in the data.

There are essentially two types of data: **quantitative** and **categorical**.

- Examples of categorical data: color, type of pet, gender, ethnic group, religious affiliation, etc.
- Examples of quantitative data: age, years of schooling, height, weight, test score, etc.

Researchers use both types of data but in different ways. Bar graphs and pie charts are frequently associated with categorical data. Box plots, dot plots, and histograms are used with quantitative data. The measures of central tendency (mean, median, and mode) apply to quantitative data. Frequencies can apply to both categorical and quantitative.

Bivariate data consists of pairs of linked numerical observations, or frequencies of things in categories. Numerical bivariate data can be presented as ordered pairs and in any way that ordered pairs can be presented: as a set of ordered pairs, as a table of values, or as a graph on the coordinate plane.

- An example would be the number of people that play certain sports or are in certain clubs at your school broken down by gender.

A bivariate or **two-way frequency chart** is often used with data from two categories. Each category is considered a variable, and the categories serve as labels in the chart. Two-way frequency charts are made of cells. The number in each cell is the frequency of things that fit both the row and column categories for the cell. From the two-way chart below, we see that there are 12 males in the band and 3 females in the chess club.

School Club	Gender		Totals
	Male	Female	
Band	12	21	33
Chorus	15	17	32
Chess	16	3	19
Latin	7	9	16
Yearbook	28	7	35
Totals	78	57	135

If no person or thing can be in more than one category per scale, the entries in each cell are called **joint frequencies**. The frequencies in the cells and the totals tell us about the percentages of students engaged in different activities based on gender. For example, we can determine that if we picked at random from the students, we are least likely to find a female in the chess club because only 3 of 135 students are females in the chess club. These frequencies are converted to percents in the chart below.

School Club	Gender		Totals
	Male	Female	
Band	8.9%	15.6%	24.5%
Chorus	11.1%	12.6%	23.7%
Chess	11.9%	2.2%	14.1%
Latin	5.2%	6.7%	11.9%
Yearbook	20.7%	5.2%	25.9%
Totals	57.8%	42.3%	100%

There is also what we call **marginal frequencies** in the bottom and right margins (grayed cells). These frequencies lack one of the categories. For our example, the frequencies at the bottom represent percents of males and females in the school population. The marginal frequencies on the right represent percents of club membership.

Lastly, associated with two-way frequency charts are **conditional frequencies**. These are not usually in the body of the chart, but can be readily calculated from the cell contents. One conditional frequency would be the percent of females that are in the chorus out of the total number of females in some type of club. 17 of the 57 females are in the chorus, so 29.8%. This could also be stated as "Given that a female in a club is selected, what is the probability that she is in the chorus?"

Practice #1:

Elizabeth surveys 9th graders, 10th graders, and 11th graders in her school. She asks each student how many hours they spend doing homework each night. She records the responses in the table below.

Grade	Hours spent on homework			
	0-2	2-4	More than 4	
9	38	12	2	
10	21	25	9	
11	14	18	20	
Total	73	55	31	159

- a) How many 9th graders spend 0-2 hours on homework each night? What frequency is that?
 38 ; $\frac{38}{159} \approx .24$
- b) How many 10th graders spend 2-4 hours on homework each night? What frequency is that?
 25 ; $\frac{25}{159} \approx .16$
- c) Which response was the most popular among 11th graders?
 0-2 hours, 2-4 hours, or more than 4 hours?

Practice #2:

Cameron surveys students in his school who play sports, and asks them which sport they prefer. He records the responses in the table below.

Gender	Preferred sport			
	Baseball	Soccer	Basketball	
Male	49	52	16	
Female	23	64	33	
	72	116	49	237

- a) What is the joint frequency of male students who prefer soccer? (How many male students prefer soccer?)
 $\frac{52}{237} \approx .22$
- b) What is the marginal frequency of each type of sport? (Total the number of males and females who played baseball. Then give total for the other two sports.)
 $\frac{72}{237}$, $\frac{116}{237}$, $\frac{49}{237}$

Practice #3:

Abigail surveys students in different grades, and asks each student which pet they prefer. The responses are in the table below.

Grade	Preferred pet				
	Bird	Cat	Dog	Fish	
9	3	49	53	22	
10	7	36	64	10	
	10	85	117	32	244

- a) What is the joint frequency of 10th graders who prefer having fish or a cat as a pet?
 $\frac{46}{244}$
- b) What is the marginal frequency of each type of preferred pet?

$\frac{10}{244}$, $\frac{85}{244}$, $\frac{117}{244}$, $\frac{32}{244}$

Name: _____ Date: _____

Task – Public Opinions

MCC9-12.S.ID.5 Summarize categorical data for two categories in two-way frequency tables. Interpret relative frequencies in the context of the data (including joint, marginal, and conditional relative frequencies). Recognize possible associations and trends in the data.

A public opinion survey explored the relationship between age and support for increasing the minimum wage. The results are found in the following two-way frequency table.

	For	Against	No Opinion	TOTAL
Ages 21-40	25	20	5	50
Ages 41-60	30	30	15	75
Over 60	50	20	5	75
TOTAL	105	70	25	200

Frequency Count

- In the 41 to 60 age group, what percentage supports increasing the minimum wage? Explain how you arrived at your percentage. What type of probability is this? Joint, marginal, or conditional?

$\frac{30}{75} = 40\%$ conditional

- Out of the people that have no opinion, what percentage is over 60 years old?

$\frac{5}{25} = .20 \rightarrow 20\%$

- What are the marginal frequencies?

$\frac{105}{200}$ $\frac{70}{200}$ $\frac{25}{200}$ $\frac{50}{200}$ $\frac{75}{200}$ $\frac{15}{200}$

- What are the joint frequencies?

$\frac{25}{200}$ $\frac{30}{200}$ $\frac{50}{200}$ $\frac{20}{200}$ $\frac{30}{200}$ $\frac{20}{200}$ $\frac{5}{200}$ $\frac{15}{200}$ $\frac{5}{200}$ $\frac{189}{200}$

For Against No opinion

- Why are joint and marginal frequencies important when describing trends or associations in data? Do you see any significant trends when looking at the frequencies?

Joint look at 2 variables
marginal look at 1 variable

Task – Leisure Time

1. Using the table below, construct a table displaying the joint and marginal frequencies.

	Dance	Sports	Movies	TOTAL
Women	16	6	8	30
Men	2	10	8	20
TOTAL	18	16	16	50

	Dance	Sports	Movies	TOTAL
Women	.32	.12	.16	.60
Men	.04	.20	.16	.40
TOTAL	.36	.32	.32	1

2. After the basketball game, the statistician did not have time to compute Jana's relative frequency. Complete the table determining the relative frequency for Jana. Discuss any trends or associations from the table below concerning points scored by two basketball players.

Point Value	Frequency for Jana	Relative Frequency for Jana	Frequency for Jill	Relative Frequency for Jill
0	0	$\frac{0}{40} = 0$	1	0.025
1	0	$\frac{0}{40} = 0$	1	0.025
2	0	0	2	0.05
3	0	0	2	0.05
4	0	0	3	0.075
5	1	$\frac{1}{40} = .025$	3	0.075
6	0	0	5	0.125
7	3	$\frac{3}{40} = .075$	4	0.1
8	6	$\frac{6}{40} = .15$	5	0.125
9	5	.125	1	0.025
10	7	$\frac{7}{40} = .175$	4	0.1
11	5	.125	5	0.125
12	4	$\frac{4}{40} = .1$	3	0.075
13	4	.1	0	0
14	3	.075	0	0
15	2	.05	1	0.025
TOTALS	40	1	40	1