SM 2

Date: **Section:** 8.1

Objective: Answering questions using marginal & conditional distributions of data

Are you more likely to survive a boat crash if you have a first-class ticket? Do boys and girls like different colors? These are questions that can be answered by examining ***distributions*** of data.

Individuals: The objects described by a set of data. Individuals may be people, animals, or objects.

Categorical Variable: A characteristic of an individual that places the individual into one of several groups or categories. Examples: Eye color, favorite ice cream flavor, gender, whether or not the person did their homework…

Relative Frequency: The *fraction or percent* of a group who fall into a category.

Distribution of a Categorical Variable: Lists the different categories that the individuals in the data set fall into and states how many (or what percent) of individuals fall into each category.

Two-way Table: A table that is broken down into rows and columns. The values of one categorical variable go along the rows, and the values of another categorical variable go down the columns. A two-way table helps us see if there's any relationship between the two variables.

Marginal Distribution: Ignores the inside of the table, and just gives the *percent of all the individuals in the whole sample* who fall into each category. **Look at the totals in the <u>margins</u>!**

Conditional Distribution: Gives the *percent of individuals in just one sub-group* who fall into each category. (For example, the percent of <u>just the boys</u> who like each color instead of the percent of everyone in the entire sample who likes each color.)

Example: In 1912, the *Titanic* hit an iceberg on its first voyage across the Atlantic and sank. Some passengers got off the ship in lifeboats, but many died. The two-way table gives information about adult passengers who lived and who died, based on what type of ticket they had.

	Survival Status		
Type of Ticket	**Survived**	**Died**	**Total**
First Class	203	122	325
Second Class	118	167	285
Third Class	178	528	706
Crew	212	673	885
Total	711	1490	2201

a) Give the marginal distribution of survival status. (For everyone in the entire sample, what percent survived and what percent died?)

Survived: $\frac{711}{2201} = 32.3\%$

Died: $\frac{1490}{2201} = 67.7\%$

b) Give the marginal distribution of type of ticket. What does this distribution tell you about?

First class: $\frac{325}{2201} = 14.8\%$ Third class: $\frac{706}{2201} = 32.1\%$

Second class: $\frac{285}{2201} = 12.9\%$ Crew: $\frac{885}{2201} = 40.2\%$

(Tells what percent of everyone on the entire ship had each type of ticket.)

c) Give the conditional distribution of survival status for people with 1st class tickets.

Survived: $\frac{203}{325} = 62.5\%$ Died: $\frac{122}{325} = 37.5\%$

(What % of 1st-class ticket holders survived & died)

d) Give the conditional distribution of survival status for people with 2nd class tickets.

Survived: $\frac{118}{285} = 41.4\%$ Died: $\frac{167}{285} = 58.6\%$

(What % of 2nd-class ticket holders survived & died)

e) Give the conditional distribution of survival status for people with 3rd class tickets.

Survived: $\frac{178}{706} = 25.2\%$ Died: $\frac{528}{706} = 74.8\%$

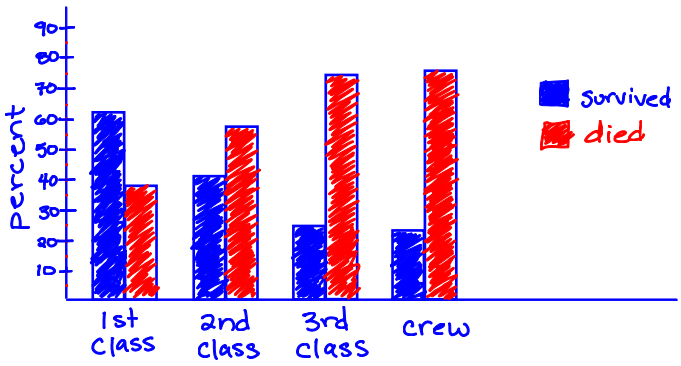
(What % of 3rd-class ticket holders survived & died)

f) Give the conditional distribution of survival status for the crew.

Survived: $\frac{212}{885} = 24.0\%$ Died: $\frac{673}{885} = 76.0\%$

(What % of crew survived & died)

g) Draw side-by-side bar graphs to compare the distributions in parts c-g. Then write a few sentences comparing and contrasting the conditional distributions.



The conditional distributions of survival for 3rd class & crew are very similar. In both of those classes, nearly twice as many people with those types of tickets died than survived. In second class, more died than survived, but the gap was not as extreme as 3rd class or crew. In 1st class, over 60% survived.

g) One of your friends tries to argue that 3rd class tickets were actually better than 2nd class tickets by saying, "A higher number of 3rd class ticket holders survived the Titanic disaster than 2nd class ticket holders." Explain what is misleading about this statement.

The problem is that there were a higher number of people with 3rd class tickets than 2nd class tickets. For the comparison to be fair, you have to compare percentages. A lower percent of 3rd class ticket holders survived than 2nd class ticket holders.

h) What percent of the passengers in 1st class survived?

$\frac{203}{325} = 62.5\%$

i) What percent of the survivors were in 1st class?

$\frac{203}{711} = 28.6\%$

Take a survey in your class of favorite colors and fill in the following table:

	Favorite Color								
	Red	**Orange**	**Yellow**	**Green**	**Blue**	**Purple**	**Pink**	**Other**	**Total**
Boys	5	0	0	2	5	1	0	1	14
Girls	2	1	2	1	4	4	1	1	16
Total	7	1	2	3	9	5	1	2	30

a) Give the marginal distribution of favorite color for your class.

Red: $7/30 = 23.3\%$ Blue: $9/30 = 30\%$

Orange: $1/30 = 3.3\%$ Purple: $5/30 = 16.7\%$

Yellow: $2/30 = 6.7\%$ Pink: $1/30 = 3.3\%$

Green: $3/30 = 10\%$ Other: $2/30 = 6.7\%$

b) Give the marginal distribution of gender for your class.

Boys: $14/30 = 46.7\%$ Girls: $16/30 = 53.3\%$

c) Give the conditional distributions of favorite color for boys and girls.

Boys: Red: $5/14 = 35.7\%$
Orange: $0/14 = 0\%$
Yellow: $0/14 = 0\%$
Green: $2/14 = 14.3\%$
Blue: $5/14 = 35.7\%$
Purple: $1/14 = 7.1\%$
Pink: $0/14 = 0\%$
Other: $1/14 = 7.1\%$

Girls: Red: $2/16 = 12.5\%$
Orange: $1/16 = 6.25\%$
Yellow: $2/16 = 12.5\%$
Green: $1/16 = 6.25\%$
Blue: $4/16 = 25\%$
Purple: $4/16 = 25\%$
Pink: $1/16 = 6.25\%$
Other: $1/16 = 6.25\%$

d) Write a few sentences comparing and contrasting the conditional distributions of favorite color for boys and girls.

A much higher percentage of boys than girls chose red and a much higher percentage of girls than boys chose yellow, purple, & pink. Both genders had high percentages who chose blue.

e) What percent <u>of the girls</u> in the class chose blue?

$\frac{4}{16} = 25\%$

f) What percent <u>of the people who chose blue</u> are girls?

$\frac{4}{9} = 44.4\%$

g) Was your answer to part f) part of a marginal distribution or a conditional distribution?

conditional - asks about just the people who chose blue

h) What percent of the people in the class chose red, orange, or yellow as a favorite color?

$\frac{7+1+2}{30} = 33.3\%$

i) Was your answer to part h) part of a marginal distribution or a conditional distribution?

marginal - asks about the whole class