

Ryan Yang
Professor Zantorian
404904494
Oct 8 2018

Question 1: What is the unit of observation in the data? How many different variables are recorded? List each variable and determine whether the variable is categorical or numerical.

Unit of observation is one child. 22 different variables are recorded.

BMonth	Categorical
Birthday	Categorical
DOW	Categorical
Gender	Categorical
Apgar5	Numerical
Premie	Categorical
LowBirth	Categorical
BirthWeight	Numerical
Gestation	Numerical
Fage	Numerical
Mage	Numerical
Feduc	Numerical
Meduc	Numerical
TotPreg	Numerical
Visits	Numerical
Marital	Categorical
Racemom	Categorical
Racedad	Categorical
Hispmom	Categorical

Hispdad	Categorical
Gained	Numerical
Smokes	Categorical

Question 2: Create a Summary Table for the variable Smokes. How many smokers are there? What percent of the sample consists of smoking mothers?

New_2009Births

	Smokes		Row Summary
	false	true	
	1744	192	1936
	0.900826	0.0991736	1

S1 = count()
S2 = rowproportion

There are 1936 smokers and roughly 10% of the table consists of smoking mothers.

Question 3: Consider the other categorical variables you listed in Question 1. Of those that record the health of the baby, which do you think will be associated with the mother's smoking and why? Make a two-way Summary Table to check your hypothesis. Do you have evidence that this variable associated with smoking? Why?

I believe low birth weight will be associated with the mother's smoking habits.

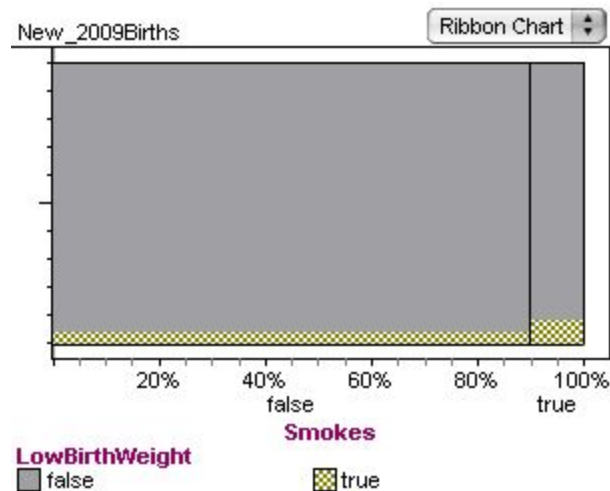
Evidence suggests a higher percentage that a baby born will have a low birth weight born from mothers who smoke in comparison to mothers who don't smoke. Babies born from mothers who do smoke have a 8% higher chance of having a low birth weight.

New_2009Births

		Smokes		Row Summary
		false	true	
LowBirthWeight	false	1673 0.904324	177 0.0956757	1850 1
	true	71 0.825581	15 0.174419	86 1
Column Summary		1744 0.900826	192 0.0991736	1936 1

S1 = count()
S2 = rowproportion

Question 4: Create a Graph that represents the variables Smokes and LowBirthWeight. Toggle between the “Bar Chart” and the “Ribbon Chart”. Which graphic best illustrates a possible association between smoking and low birth weight, and why?



I believe a ribbon chart best illustrates the association between smoking and birth weight as it displays percentages. The Bar chart only displays the absolute number. For this certain graph, it displayed a higher percentage of low birth weight babies born from mothers who smoke.

Question 5: What happens if you drop a numerical variable in the Summary Table to replace the categorical variable you chose for Question 3? Add some other summary statistics to the table that help you describe the difference between smoking and non-smoking mothers with regard to the numerical variable you chose. What are the differences between the two groups?

New_2009Births			
	Smokes		Row Summary
	false	true	
BirthWeight	1744	192	1936
	3319.58	3102.35	3298.03
	2182.91	2211.26	2182.91
	3033.4	2834.95	3005.05
	3345.24	3090.1	3316.89
	3642.91	3387.77	3628.74
	4110.68	4025.63	4110.68

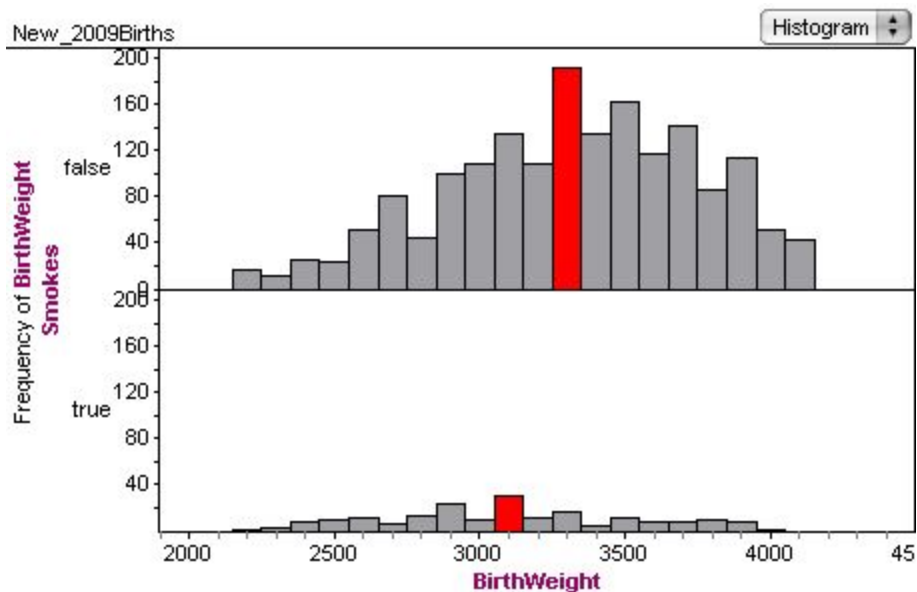
```

S1 = count( )
S2 = mean( )
S3 = min( )
S4 = Q1( )
S5 = median( )
S6 = Q3( )
S7 = max( )

```

I dropped the numerical variable BirthWeight to replace the categorical variable in question 3. Afterwards, I added a 5 number summary and found that in many ways, such as mean, median, Q1, and Q3 the birth weight for mothers who smoke are all lower than babies whose mothers don't smoke.

Question 6: Make a Graph that has the numerical variable you chose in Question 5 on one axis and the variable Smokes on the other. What do these graphs say about the differences between births of smoking mothers and non-smoking mothers? Is this surprising, or what you would expect?



The graph displays a median baby weight of roughly 3250 for mothers who don't smoke and roughly a median baby weight of 3100 for mothers who do smoke. However, I expected this as in previous summary tables, there was a correlation between babies with lower weights and

mothers who smoke. The babies from mothers who don't smoke also have a left skew meaning there is a lower amount which is expected.

Summary Question: What concepts from the textbook are covered in this lab? What concepts, if any, are not covered in the textbook? You should review earlier readings, as well as the week's current topic. Have you seen these concepts elsewhere, e.g., lecture, discussion section, previous labs, or homework problems? Be specific in your answer.

Categorical and Numerical Data as well as types of charts, such as histograms are covered in this lab. Additionally, shapes of graphs have also shown up in this lab, such as unimodal shape. I have seen these concepts in the textbook, lecture, as well as homework problems. Due to these concepts, I was able to create two way tables and charts to find correlations between data. Additionally, I was also able to interpret graphs, specifically histograms in this lab to identify a center/median as well as interpret the shape of the graph to determine that babies born from mothers who smoke have a higher percentage of having a low birth weight as well as having a lower median birth weight in general.