Lab: Baby Boom

Investigative Question: What is the effect of smoking during pregnancy on the baby's health?

Objectives: Summaries of data help us detect and describe differences between groups. In this lab you'll explore some of the basic capabilities Fathom provides for exploring and analyzing data.

Data: *2009Births.ftm*

Reading: Chapters 1-3

Background: In 2009, the state of North Carolina released to the public a large data set containing information on births recorded in this state. This data set has been of interest to medical researchers who are interested in studying the relation between habits and practices of expectant mothers and the birth of their children.

In this lab, we will focus on studying how smoking affects the health of a newborn infant, although other factors that might affect the infants' health are also included. To do so, we will use various numerical and graphical features in Fathom.

Instead of providing the entire data set, we will work with a sample of 1,936 observations. These cases were chosen at random and the following variables were recorded:

Bmonth	Birth Month
Bday	Birth Day
DOW	Day of Week
Gender	Gender of baby
Apgar5	Apgar Score at 5 minute (low scores indicate a need for medical attention.)
Premie	Premature
LowBirthWeight	Low birth weight
BirthWeight	Weight of baby at birth (grams)
Gestation	Length of gestation (weeks)
Fage	Father's age (years)
Mage	Mom's age (years)
Feduc	Father's Education (years)
Meduc	Mother's Education (years)
TotPreg	Total Number of Pregnancies (Number of pregnancies including current)
Visits	Pre-delivery doctor visits
Marital	Marital Status
Racemom	Race of mom
Racedad	Race of dad
Hispmom	Hispanic mom
Hispdad	Hispanic dad
Gained	Weight gained by mom (kilograms)
Smokes	Mom's smoking habits

Lab Questions

> **Lab Instructions:** If needed, use the Fathom instructions on the following pages to answer the questions below. The instructions to copy and paste Fathom output into your solutions are also in the following pages.

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.

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

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?

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?

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?

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.

Due next week: Your answers to Questions 1-6 and the summary question are due at the start of lab next week. Include appropriate graphs and tables to support your narrative.

Fathom Instructions

Accessing and Viewing the Data

Getting familiar with the data is always the first step when trying to answer a research question. Open the Fathom file Births2009.ftm by clicking on the link to the data in CCLE. Ask your TA if you are having trouble.

The first thing you will see is a Fathom **Document** with a **Collection** called **Births2009** on it. **Collections** look like boxes and contain data. You can have several **Collections** in the same document. If you make the **Collection** box larger (by dragging on the corners) you will be able to see the contents. Each ball represents an observation or case. In this case, the balls are labeled with the weight of the baby at birth.

Along the top of the **Document Window** is the **Toolbar**. For the most part, the **Toolbar** contains objects that act upon variables or **Collections**. We will make frequent use of **Table, Graph,** and **Summary**.

I   Opening the Inspector of a Collection

Double-click on the **Collection** (specifically on the picture of the box, not the words). This will open an **Inspector**. The **Inspector** allows you to view the data, one observation at a time, along with other information about the variables.

At the top of the **Inspector** are various tabs. The first tab in the **Inspector**, called **Cases**, is like a deck of cards, and on each card is recorded the information for one observation. You can flip through the deck (i.e. each observation) by clicking on the right and left arrows at the bottom of the **Inspector**.

Note that the first column of the **Cases** tab is called **"Attribute"**. This is simply a list of all the variables included in the **Collection** of data. To see what the variables mean, click on the **Comments** tab of the **Inspector**.

II   Opening the Case Table of a Collection

Instead of viewing each observation one at a time, you can view all the observations from a **Collection** in one large table using a **Case Table**:

- Click once on the **Collection** to highlight it.

- Click on the **Table** button in your *Fathom* **Toolbar** and drag the hand that appears to an empty space on your **Document**.

You should now see the **Case Table** for your **Collection**. The rows are cases, or observations, and the columns are variables, which *Fathom* calls attributes. Scroll through a few observations.

Note: If your table is blank, then you forgot to highlight your **Collection** first. You can delete anything you create in *Fathom* by clicking on it to highlight it, and then pressing the **"Delete** button on your keyboard. Alternatively, in the **"Edit** drop-down menu, you can also select **"Undo"** from the menu for a "re-do". This is a *very* useful menu option. It lets you back up through a great number of past actions.

Useful Tools for Lab Solutions

I   Creating a Blank Table to Fill In

If you want to create a blank table that you can manually fill in with information for your lab solutions, you can follow these steps:

- Click once in the empty white space of your **Document** to ensure no **Collection** is highlighted.

- Click on the **Table** button in your *Fathom* **Toolbar** and drag the hand that appears to an empty space on your **Document**. You should see a blank table.

- To name the first column, click once on "<new>", type the heading name and press "Enter" on your keyboard.

- If you click on the cell beneath your column name, you can now start typing in your entries for that column. If you press "Enter" or the down button on your keyboard, a new cell in a new row will appear.

- Once you finish filling out the first column, you can fill additional columns the same way.

II   Using a Word Processor on a Mac

You can use any word processor of your choice for your lab solutions, but for the Macs in the lab, we suggest you use *NeoOffice*. To open *NeoOffice*, follow these instructions:

- Move your mouse to the bottom of your screen and a "Dock" will appear that contains several icons for several Mac applications

- Click on the first icon, a happy face called the "Finder".

- On the left side of the "Finder", click on "Applications". To view the applications as a list (if they are not already), click on the "List View" button at the top of the finder.

- Double-click on *NeoOffice*. Once it opens (it might take awhile), open a blank document and beginning writing your lab solutions.

- When you save your document, be sure to save it as a **".doc"** file if you want to be able to open in with other applications (such as *Word*, for example) in the future.

- Also, be sure to remember where you save your document so that you can e-mail it to yourself at the end of lab. We suggest you save your documents on the Desktop so they are easy to find. ***It is important that you e-mail yourself all work done in the lab because you may not be able to access files saved on your lab computers in the future once the computers are refreshed.***

III   Copying and Pasting Fathom Output onto Solutions

You can copy and paste any tables or graphs from *Fathom* onto your lab solutions by following these directions:

- In *Fathom*, Click once on the item you would like to copy to highlight it.

- In the **"Edit"** drop-down menu, select **"Copy As Picture"**.

- In your word processor, go to the **"Edit"** drop-down menu and select **"Paste"**.

Summarizing Categorical Data

Now we are ready to examine some variables and look more closely at their relation to the babies health. You can get an informative summary of any type of variable with the **Summary** object in the **Toolbar**.

I   Creating a Summary Table

- Click on the **Summary** button in the Toolbar and drag and drop a blank **Summary** object onto your *Fathom* **Document**.

- In either the **Inspector** or the **Case Table**, click on the *name* of the variable of interest (in this case, *Habit*, for example) and drag and drop the variable onto the **Summary Table**. Notice that you have a choice of placing *Habit* as a row or a column (it does not matter which you choose).

II   Creating a Two-Way Summary Table

- If you drag-and-drop another categorical variable onto the same **Summary Table**, and put it where the other variable is *not* (as a row or column), you will get a two-way table.

- If you place it in the same position as the other variable, it will replace that variable instead.

III   Adding Proportions to Summary Table

For categorical variables, you should see the counts of each possible outcome of that variable in the **Summary Table**. To see the breakdown of proportions or percentages, follow these steps:

- Click on the **Summary Table** to highlight it, click on the **"Summary"** drop-down menu and select **"Add Formula"**. In general, whenever you click and select a *Fathom* object (such as a **Table, Graph,** and **Summary**) the menu at the top of the screen will change to give you options for working on that object.

- In the formula editor that pops up, type *"rowproportion"* (without the quotes) to see the row proportions or *"columnproportion"* to see the column proportions. Be sure to spell the names of the formulas correctly or else *Fathom* will give you an error. (If you spell the names correctly, they should change to a purplish color in your editor.)

- You will see that each cell in the **Summary Table** now includes numbers for multiple statistics. To see which numbers correspond with which statistics, simply look at the bottom of your summary table to see the order of the statistics or formulas within each cell.

- To delete (or change) a particular statistic from the table, you can double click on its name at the bottom of the **Summary Table**. In the formula editor, press delete (or make your changes) and then click **"OK**.

For this lab, we want to compare the smoker and the non-smoker groups with respect to the other variables chosen. If you represented *Habit* by rows and *Premie* by columns, for example, then you will use row proportions to see the proportion of smokers who had premature (or full-term) babies versus the proportion of non-smokers who had premature (or full-term) babies. If you placed the variables in the opposite configuration, then you would use column proportions to find the above proportions.

IV   Creating Graphs of Categorical Variables

It is always useful and important to visualize the data you are trying to analyze. To do this in Fathom, follow these steps:

- Click on the **Graph** button in the **Toolbar** and drag and drop a blank **Graph** object onto the **Document**.

- Drag and drop the *name* of the variable of interest (*Habit*, for example) onto the x-axis (horizontal axis) of this graph.

- Notice that a selector appears in the upper-right side of the graph. You can choose from **"Bar Chart"** or **"Ribbon Chart"** for categorical variables.

- To add another categorical variable, drag and drop that variable (for example, *LowBirthWeight*) onto the graph. You have two choices: drop it onto the center of the graph or drop it onto the y-axis (vertical axis). Do both so that you can see what the effect is. Remember that you can **"Undo"** any actions you make.

When you create graphs, you want to use the one that best tells the story you want to convey about the data. After playing around with the different options in *Fathom*, you can choose which graph **you** think is most appropriate to use to answer the questions in your labs.

Take a moment to click on part of the bar chart or ribbon chart you created above. (It doesn't matter which one.) Notice that the observations associated with the part of the **Graph** that you clicked are highlighted in your **Inspector, Case Table** and any **Summary Tables** you have open corresponding to that graph.

Summarizing Numeric Data

I   Creating a Summary Table

As you did with categorical variables, you can create **Summary Tables** using a *numerical* variable instead. For this lab, we will still keep the categorical *Habit* variable on the **Summary Table** since we want to compare smokers and non-smokers, but we can replace the other categorical variable with a numerical variable just by dragging and dropping the name of the numerical variable on top of the existing variable on the **Summary Table**.

You will see that you can examine different statistics for the numerical variables than you can with the categorical variables, such as the mean. To add additional statistics, such as the five-number summary and the standard deviation, follow these steps:

- Once you had created a **Summary Table** with a numerical variable, highlight it and in the **"Summary"** drop-down menu, select **"Add Five-Number Summary"** to add the five-number summary. Again, you will see each statistic listed at the bottom of your **Summary Table** so you know what each number represents.

- To add the standard deviation, highlight the **Summary Table**, and in the **"Summary"** drop-down menu, select **"Add Formula"**. In the formula editor, type **"stdDev()"** (without the quotes) and press OK. (Remember to spell this correctly and dont forget the parentheses at the end.)

- In general, you can add any of the individual formulas as they appear at the bottom of your **Summary Table** using the **"Add Formula"** method. You can delete any unnecessary statistics as described above (in the proportions section).

When summarizing numerical data, you do not need to use every statistic available, but you should use whichever ones you think best allow you to describe the data. Remember that when you compare the distributions of data for two groups of observations, it is important to compare the shape, center, spread and any unusual features, such as outliers. You can summarize center and spread using the numerical statistics above. You can summarize shape and unusual features using graphs, which will be described below.

II Creating Graphs of Numerical Variables

You can create **Graphs** for numerical data in *Fathom*, but you will see that there are different options to visualize numerical data than there are for categorical data.

- Click on the **Graph** button in the **Toolbar** and drag and drop a blank **Graph** object onto the **Document**.

- Drag and drop the name of a *numerical* variable of interest onto the x-axis (horizontal axis) of this graph. You will see a **"Dot Plot"**, as specified by the selector in the upper-right part of the **Graph**. You can change this selector to some of the other options (specifically, **"Histogram"** or **"Box Plot"**).

- If you want to look compare the distribution of this numerical variable between two groups, drag the categorical grouping variable of interest (in this case, *habit*) to the y-axis (vertical axis) of the **Graph**.

- Because the two groups you are comparing might be different sizes, for the **"Histogram** option, it might be useful to look at a ***relative frequency histogram*** so that the two graphs are more comparable. Highlight the **Graph** and under the **"Graph"** drop-down menu, select **"Scale"** and then **"Relative Frequency"**.

Again, it doesnt matter what type of graph you use as long as you are able to tell the story about your data that you want to tell in conjunction with your summary statistics.