LARSON—INFO 790–CLASSROOM WORKSHEET 09
Precision and Lift

We will modify the Titanic example and Steve Essi's data example to add a calculation of *precision* and lift.

1. **Log in to VCU's Athena cluster**.

   The following directions assume you have an Athena account, that you have set up Sage, and that you have set up (using `make`) the CONJECTURING program.

   (a) Start the Chrome browser.

   (b) If you are off-campus, you'll need to connect to the VPN first.

   (c) Then go to `https://athena3.hprc.vcu.edu`

   (d) Login using your VCU EID as your username, and your corresponding VCU password.

   (e) Click the Apps button and a `Sage` session. The default options are fine. This will take a couple of minutes.

   (f) Click the Apps button and start an "athena shell access" session (this will give you a terminal window, where we can issue commands).

   (g) Your Sage session will first say "Queued", then "Starting". When it is ready you will see a button that says, "Connect to Sage". Click that.

   (h) You should then get an "untitled" interactive-Python notebook (ipynb), or the last file you had open the previous time you used Athena.

   (i) When your notebook opens look on the upper-right to make sure the SageMath kernel is running (if it isn't you can change the *kernel*).

2. **Reminders for setting up Expressions and Conjecturing**. In each case, for each experiment, we will make a folder in your root directory; we will need a copy of the "expressions" compiled executable in that folder; and we will use an .ipynb located in that notebook. When we call the CONJECTURING program we will use the version in the ∼/`conjecturing` folder downloaded from github (what you did with the github command; if there are ever new files on github, using the command `git pull` will update your files).

   Precision and Lift

   (From Wikipedia:) In a classification task, the *precision* for a class is the number of true positives (i.e. the number of items correctly labelled as belonging to the positive class) divided by the total number of elements labelled as belonging to the positive class (i.e. the sum of true positives and false positives, which are items incorrectly labelled as belonging to the class)

   Suppose we are investigating sufficient conditions for survival of a Titanic passenger. The conjecture at hand is: "If the passenger is female (F) then the passenger survived". Here we will only look at the (reserved) *test* objects.

3. The **precision** is:

$$\frac{\text{Number of test objects with property F and property S}}{\text{Number of test objects with Property S}}$$

4. The **lift** is:

$$\text{Precision} \div \frac{\text{Number of test objects with Property S}}{\text{Total number of (test) objects}}$$

Titanic with Precision and Lift

5. Getting the updated script:
   (a) Go to `https://math1um.github.io/Teaching/`
   (b) Scroll down and find the INFO 790 files. Download the Titanic example script with Precision and Lift (.ipynb) file.
   (c) Put this file in your Titanic directory on Athena (where the Titanic data file is). On your Athena Jupyter notebook, there is a button for *uploading* files.

Steve Essi's Multi-level example with Precision and Lift

6. Getting the updated script:
   (a) Go to `https://math1um.github.io/Teaching/`
   (b) Scroll down and find the INFO 790 files. Download the Essi example script with Precision and Lift (.ipynb) file.
   (c) Put this file in your C08_experiment directory on Athena (where Essi's data file is). On your Athena Jupyter notebook, there is a button for *uploading* files.

Final Note

Dr Brooks wrote this file so that it could be easily imitated for a wide-variety of tabular data-files. You should read each cell carefully and **ask questions** about what the commands do. **You will be doing this with your own data files**.