DSFA

Spring 2020

Lecture 12

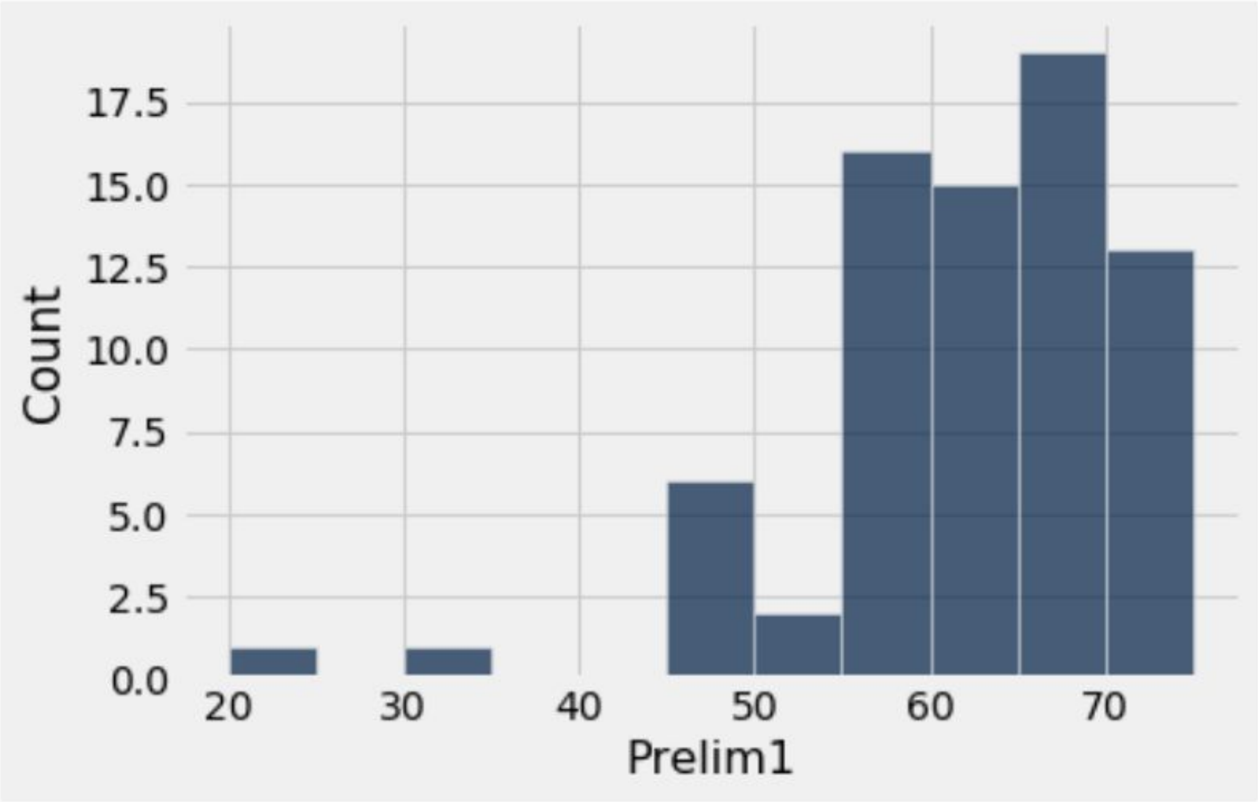
Probability and Sampling

Announcements

- Prelim 1 grades
- Lab 5 Wednesday-Thursday
- College of Engineering survey; due 3/9 midnight for 3 clicker points.
- Project 1 part 2 due Friday 3/6, 5:59PM

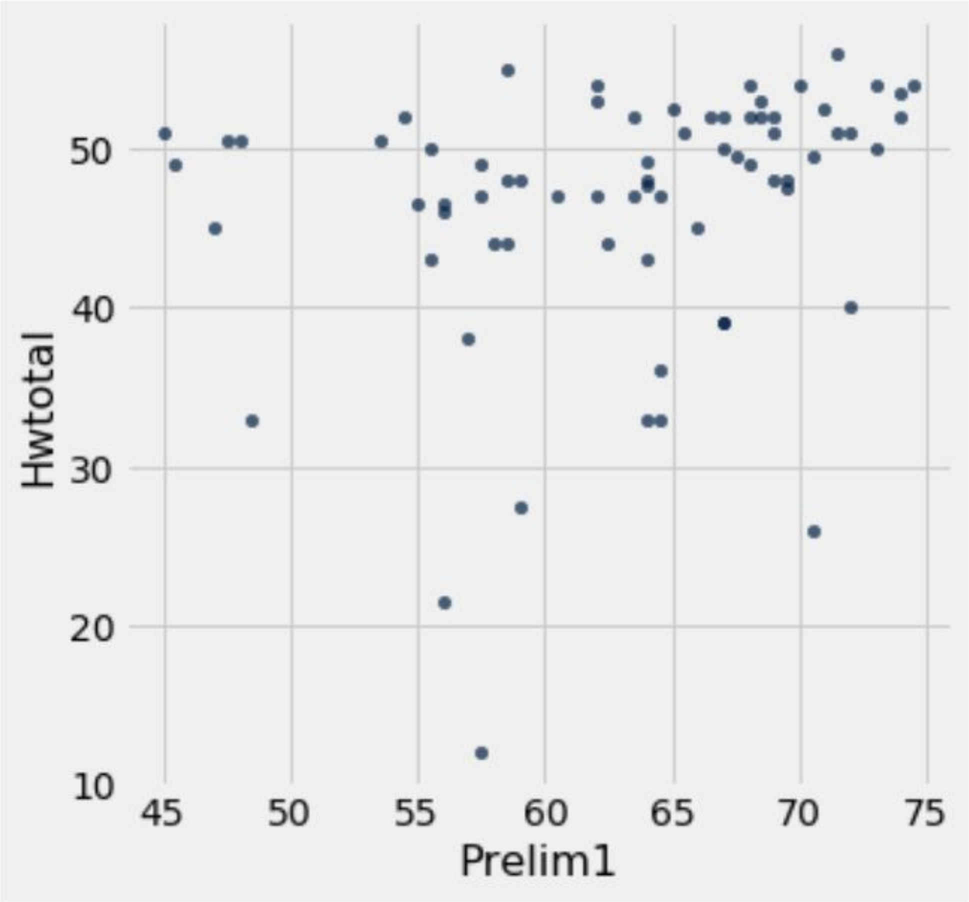
In [17]:

```
grades.hist("Prelim1", bins=np.arange(20,76,5),normed=False)
```

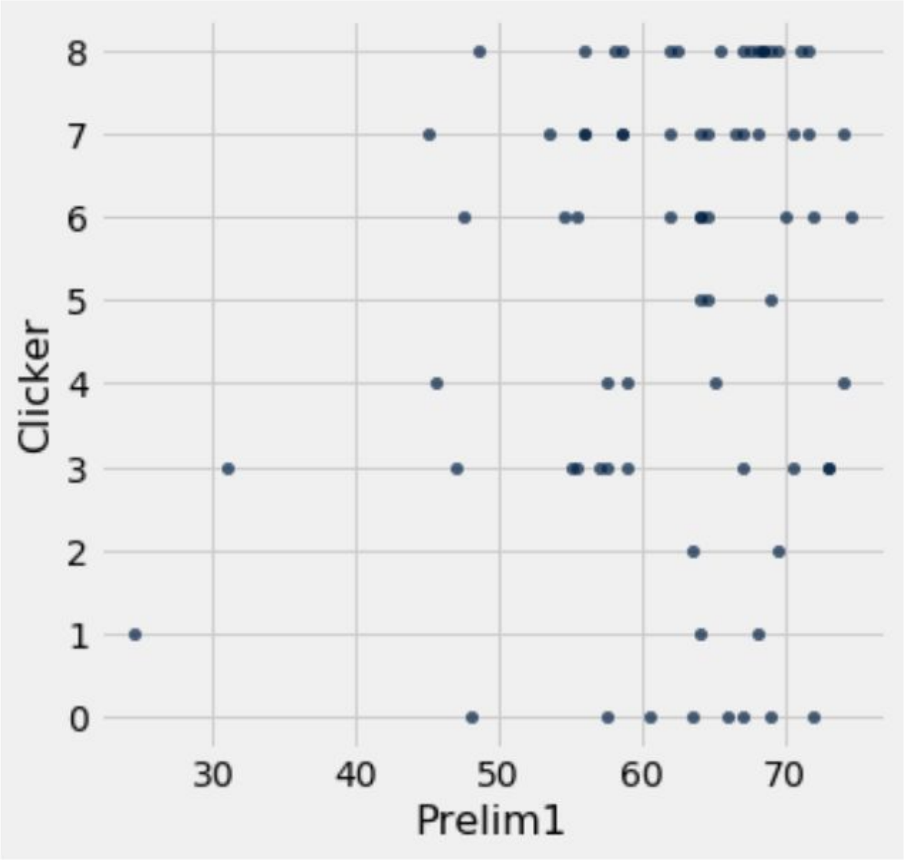


In [20]:

```
grades.where("Prelim1",are.above(40)).scatter("Prelim1","Hwtotal")
```



In [45]:
```
grades.scatter("Prelim1", "Clicker")
```



Regrade procedure

- Either professor or section TA will fix arithmetic errors on request.
- If you think a question was graded erroneously, fill out a sheet of paper explaining the error, staple on top of exam, bring to lecture by Thursday 3/12.
- We reserve the right to regrade all of the exam (not just the part you want regraded), so your final score after regrading could be higher or lower.

Probability

Probability

- Lowest value: 0
  - Chance of event that is impossible
- Highest value: 1 (or 100%)
  - Chance of event that is certain

- If an event has chance 70%, then the chance that it doesn’t happen is
  - 100% - 70% = 30%
  - 1 - 0.7 = 0.3

Equally Likely Outcomes

Assuming all outcomes are equally likely, the chance of an event A is:

$$P(A) = \frac{\text{number of outcomes that make A happen}}{\text{total number of outcomes}}$$

(Demo)

Multiplication Rule

Chance that two events *A* and *B* both happen

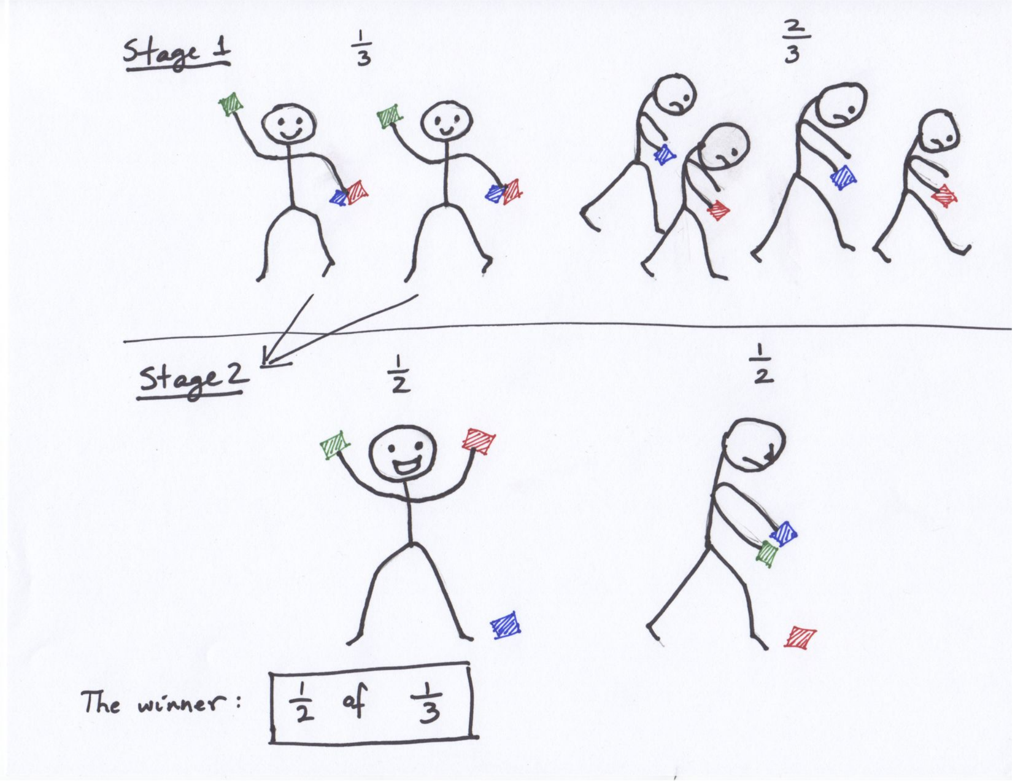
= P(*A* happens)

x P(*B* happens **given that** *A* has happened)

- The answer is *less than or equal to* each of the two chances being multiplied

(Demo)

Fraction of a Fraction



Addition Rule

If event *A* can happen in *exactly one* of two ways, then

$$P(A) = P(\text{first way}) + P(\text{second way})$$

- The answer is *greater than or equal to* the chance of each individual way

Example: At Least One Head

- In 3 tosses:
  - Any outcome *except* TTT
  - $P(TTT) = (\frac{1}{2}) \times (\frac{1}{2}) \times (\frac{1}{2}) = (\frac{1}{2})^3 = \frac{1}{8}$
  - $P(\text{at least one head}) = 1 - P(TTT) = \frac{7}{8} = 87.5\%$

- In 10 tosses:
  - $P(TTTTTTTTTT) = (\frac{1}{2})^{10}$
  - $P(\text{at least one head}) = 1 - (\frac{1}{2})^{10} = 99.90\%$

(Demo)

Example: At least one six

Let’s try two ways to compute the probability of rolling at least one six from a six-sided die in twenty rolls.

- Computation: Compute the probability mathematically
- Simulation: Use Python to try it many times and see how often we get at least one six.

Sampling

Sampling

Observe some *individuals* from a *population*

a. Examine 10 rolls of a d6 (six-sided die)
b. Coat color of the first 20 people who walk through door
c. Survey 1000 students living in campus dorms, where every student living on campus is equally likely to be chosen, and ask them about who they plan to vote for.

Sampling

- Deterministic sample:
  - Sampling scheme doesn’t involve chance

- Probability (random) sample:
  - Before the sample is drawn, you have to know the selection probability of every group of people in the population
  - Not all individuals have to have equal chance of being selected

(Demo)

Sample of Convenience

- Example: sample consists of whoever walks by
- Just because you think you're sampling "at random", doesn't mean you are. If you can't figure out ahead of time
  - what's the population
  - what's the chance of selection, for each group in the population

  then you don't have a random sample

Does sample look like population?

(Demo)

Large Random Samples

If the sample size is large,

then the empirical distribution of a simple random sample

resembles the population distribution,

with high probability.

Distribution

- A distribution is a description of the likelihood of *events or outcomes*

- Empirical distribution:
  - Experimental: made from observations
  - Proportion of each event in sample

vs.

- Probability distribution:
  - Theoretical: made from mathematics
  - Probability of each event

Law of Large Numbers

If an experiment is repeated many times,
independently and under the same conditions,
then the proportion of times that an event occurs
gets closer to the theoretical probability of the event

Sometimes called *Law of Averages*