Agenda

1. Randomness

2. Probability

Simulation   Sally and Joan plan to meet to study in the Thims College campus center. They are both impatient people who will only wait 10 minutes for the other before leaving. Rather than pick a specific time to meet, they agree to head over to the campus center sometime between 7:00 and 8:00pm. Let both arrival times be uniformly distributed over the hour, and assume that they are independent of each other. What is the probability that they actually meet?

   Rather than figure out a mathematical solution to the problem, we can use R to simulate the situation and then derive an estimate of the true value.

```
require(mosaic)
friends <- data.frame(
  sally = runif(100000, min = 0, max = 60),
  joan = runif(100000, min = 0, max = 60)) %>%
  mutate(meet = abs(sally - joan) <= 10)
tally(~ meet, data = friends, format = "percent")
qplot(data = friends, x = joan, y = sally, color = meet, alpha = 0.1) +
  geom_abline(intercept = c(-10, 10), slope = 1)
```

   Extra Credit: Provide a mathematical solution to the above problem.

Randomness

- Key Idea: if a process is *random*, then individual outcomes can't be predicted

- BUT: the distribution of outcomes, in the long run, is often quite regular and predictable.

- The Law of Large Numbers guarantees that the sample proportion ($\hat{p}$) will converge towards the population proportion (p) as $n \to \infty$.

In-class Activity   Many things in life that are called "random," are anything but (e.g. when you had that song stuck in your head last week and then overheard someone at the Campus Center singing it).

1. Click on the Google spreadsheet link on Moodle under today.

2. Find an empty cell in the first column, and type in a "random" string of 100 T's and H's

3. I'll use `googlesheets` to crowd-source your work into one long random string

   If I also generate a random string of T's and H's using R's pseudorandom number generator, do you think I will be able to tell which string was generated by human minds and which was generated by the computer?

   Given the length of the string n, it turns out the number of runs of length k, for any integer $k \geq 1$ is a *random variable* that follows a known distribution.

Probability

In-Class Problems

1. Are the probabilities legitimate? In each of the following situations, state whether or not the given assignment of probabilities to individual outcomes is legitimate, that is, satisfies the rules of probability. If not, give specific reasons for your answer.

   (a) Choose a college student at random and record gender and enrollment status: $\Pr(\text{female full-time}) = 0.44$, $\Pr(\text{female part-time}) = 0.56$, $\Pr(\text{male full-time}) = 0.46$, $\Pr(\text{male part-time}) = 0.54$.

   (b) Deal a card from a shuffled deck: $\Pr(clubs) = 16/52$, $\Pr(diamonds) = 12/52$, $\Pr(hearts) = 12/52$, $\Pr(spades) = 12/52$.

   (c) Roll a die and record the count of spots on the up-face: $\Pr(1) = 1/3$, $\Pr(2) = 0$, $\Pr(3) = 1/6$, $\Pr(4) = 1/3$, $\Pr(5) = 1/6$, $\Pr(6) = 0$.

2. Loaded dice. There are many ways to produce crooked dice. To *load* a die so that 6 comes up too often and 1 (which is opposite 6) comes up too seldom, add a bit of lead to the filling of the spot on the 1 face. Because the spot is solid plastic, this works even with transparent dice. If a die is loaded so that 6 comes up with probability 0.21 and the probabilities of the 2,3,4, and 5 faces are not affected, what is the assignment of probabilities to faces?

3. Is this calculation correct? Government data show that 6% of the American population are at least 75 years of age and that about 51% are women. Explain why it is wrong to conclude that because $(0.06)(0.51) = 0.0306$ about 3% of the population are women aged 75 or over.

4. Colored dice. Here's more evidence that our intuition about chance behavior is not very accurate. A six-sided die has four green and two red faces, all equally probable. Psychologists asked students to say which of these color sequences is most likely to come up at the beginning of a long set of rolls of this die:

$$\text{RGRRR} \qquad \text{RGRRRG} \qquad \text{GRRRRR}$$

More than 60% chose the second sequence. What is the correct probability of each sequence?