Agenda

1. Randomness
2. Probability

Randomness

- Key Idea: if a process is random, then individual outcomes cant be predicted
- BUT: the distribution of outcomes, in the long run, is often quite regular and predictable.
  - Equally likely outcomes (e.g., drawing any card, rolling a die)
  - Unequally likely outcomes (e.g., drawing a face card vs. non-face card, mens heigths)
- The Law of Large Numbers guarantees that the sample proportion (p) will converge towards the population proportion (p) as $n \to \infty$.

Example: It's Unfair Of 50 students in a class, 15 are SDS majors and 35 are non-SDS majors. Four competitors will represent Smith at DataFest and are (allegedly) chosen by chance. The 4-person team turns out to be 3 SDS majors and 1 non-SDS major. The non-SDS majors cry foul! What is the chance that three or more of the four drawn will be SDS majors (assuming students are picked with replacement)?

- Possible outcomes: 0 SDS (none), 1 SDS, 2 SDS, 3 SDS, 4 SDS (all)
- Event A: 3 or more SDS majors end up on the team (combination or two outcomes: 3 SDS + 4 SDS)
- How would we simulate this random event (Event A) with cards?

```
library(mosaic)
major <- c(rep("SDS", 15), rep("non-SDS", 35))
the_class <- data.frame(major)
sim <- do(100) * filter(sample_n(the_class, 4, replace = TRUE), major == "SDS") %>%
  summarise(n = n()) %>%
mutate(event_yes = ifelse(n > 2, 1, 0))
sim %>% summarise(mean(event_yes))

##   mean(event_yes)
## 1            0.08

true_prob <- dbinom(3, 4, .3) + dbinom(4, 4, .3)
```

Probability

- Basic probability notation
- Disjoint (Mutually exclusive) events
- Independence of events
  1. Events that are disjoint are definitely NOT independent
  2. If events are not disjoint may or may not be independent
- Probability rules

In-Class Problems

1. For each pair of events A and B, say whether or not you think they are disjoint, and whether or not you think they are independent.

   (a) A: Washing your hands, B: getting sick

   (b) Consider rolling two dice, A: The sum greater than 2, B: rolling snake eyes

   (c) A: Dealt a face card, B: dealt a red card

   (d) For a randomly selected Smithie, A: She is a democrat, B: She thinks global warming is a myth

   (e) For a randomly selected Smithie, A: She is a democrat, B: She is a republican

   (f) For a randomly selected Smithie, A: She is a democrat, B: She is a cat person

2. 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(female full-time) = 0.44, Pr(female part-time) = 0.56, Pr(male full-time) = 0.46, Pr(male part-time) = 0.54.

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

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

3. 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?

4. 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.

5. 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:

   RGRRR RGRRRG GRRRRR

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

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.