Dis 3: Sampling (Cont'd); Probability

Related textbook chapters: 5 and 6
Ch 6 handout and solution offered by Dr. Pac can be accessed here: Handout Solution
This handout incorporates reviews with all exercises from the handout given by Dr. Pac.

1 Sampling

- Recall from discussion 1 that a sample is a subset of data taken from the population. The action of taking this subset of data to construct your sample is called **sampling**.
- Eventually, our goal is to use the sample to draw conclusion about the population (inferential statistics), so it is important that our sample resembles the population.
- Different **sampling plans** are proposed to construct the sample, weighing the benefits of the plan against the costs:
  - **Simple random sampling**: every possible sample entry has equal chance of being selected.
  - **Stratified random sampling**: separate the population into mutually exclusive sets (i.e. strata), and then draw simple random samples from each stratum.
  - **Cluster sampling**: population is first divided into groups, and then one uses simple random sampling to select groups; all observations within the selected groups thus enter the sample.

Exercise. Which sampling plan is used in each of the following examples?

1. Categorize all Econ 310 students based on their class standing (freshman, sophomore, junior, senior, above senior), and then randomly selects 30 students from each class.
2. Categorize all Econ 310 students based on their class standing (freshman, sophomore, junior, senior, above senior), and then randomly select 2 out of the 5 possible groups. The groups corresponding with the class standing selected are chosen as the sample.
3. Number Econ 310 students sequentially from 1 to N. Draw 50 non-repeat random positive integers that are less than or equal to N. Select the students with the same numbers.

- As you can already see from the exercise, factoring in the specific steps taken when sampling, some sampling plan is expected to construct a sample that more closely resembles the population than the others.

  To formally examine how far the samples are from the population, we look at two types of errors that occur:

  1. **Sampling error**: difference between the sample and the population that exists only because the observations that happen to be included in the sample.
     $\Rightarrow$ increasing the sample size reduces this error

2. **Nonsampling errors**: more serious type of error due to samples being selected improperly.
   $\Rightarrow$ increasing the sample size will NOT reduce this type of error
   Nonsampling errors can be divided into three categories:
   (a) **Errors in data acquisition**: the data is recorded wrong (due to incorrect measurement, mistake made during transcription, human errors)
   (b) **Nonresponse errors**: responses are not obtained from certain people.
   (c) **Selection bias**: some members from the target population cannot possibly be selected to be within the sample.

Exercise. Which type of error arises from the following examples?

1. You sent out a survey to all Econ 310 students via email, but some people quickly archived your email without filling out the survey.

2. You sent out a survey to all Econ 310 students via email, but some freshmen has yet to activate their UW email account, so the survey was not delivered to them.

3. You randomly selected 30 Econ 310 students to have them answer your survey questions. All 30 of them responded, and you did not make any mistake in recording the data. However, your result derived from the sample is still quite different from the parameter in the population.

4. You randomly selected 30 Econ 310 students to have them answer your survey questions. All 30 of them responded, but you messed up the order of items in two columns of the data recorded.

2 Probability

2.1 Sample Space, Outcome, and Event

- Say that we roll a six-sided die:
  - The roll could be 1, 2, 3, 4, 5, or 6
  - The set containing all possible rolls is defined as the **sample space**.
  - Each possible roll is an **outcome**.
    * Outcomes in a sample space should be **exhaustive** – all possible outcomes must be included in the sample space.
    * Outcomes in a sample space should be **mutually exclusive** – no two outcomes can occur at the same time.
  - An **event** is a collection of one or more simple events (i.e. outcomes) in a sample space.

Exercise. Is the following true or false? Why?

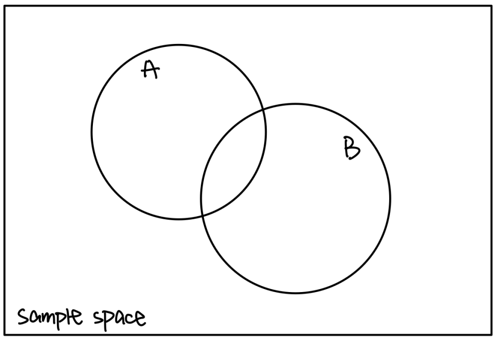
1. The set $\{$rolling a 1, rolling a 2, rolling a 3, rolling a 4$\}$ describes the sample space from a fair, six-sided die roll.

2. "Rolling a 3" is an event.

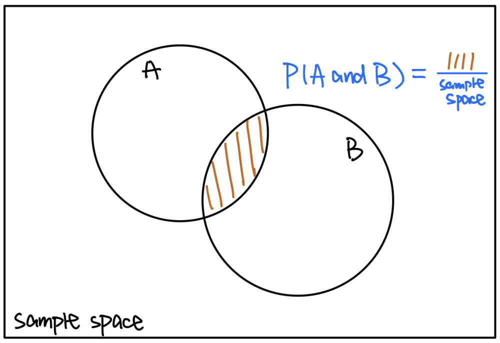
3. "Rolling an even number" is an event.

2.2 Joint vs. Conditional Probability

- Say that there are two events, denoted by A and B.



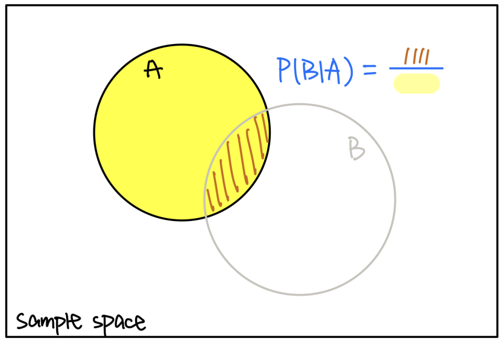
- If one is interested in the probability that event A and B occurs together, then



$$P(A \text{ and } B) = \frac{\text{Area of the line shaded space}}{\text{Area of the entire sample space (rectangle)}}$$

This is because **joint** probability looks at case where event A and B occur together with respect to all the possible outcomes (i.e. the sample space).

- Sometimes though, one is interested in the probability of B occuring conditional on A already occurs. This is depicted as



$$P(B|A) = \frac{\text{Area of the line shaded space}}{\text{Area of event } A} = \frac{P(A \text{ and } B)}{P(A)}$$

This is because **conditional** probability of $P(B|A)$ looks at case where event B occurs with respect to A already occuring.

- One way to visualize the relationship between joint and conditional probability is through the tree diagram (will see this in exercise later).
- Conditional probability gives rise to **Bayes Law**:

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)} = \frac{P(B|A)P(A)}{P(B)} = \frac{P(B|A)P(A)}{\underbrace{P(B|A)P(A) + P(B|A^C)P(A^C)}_{P(A \text{ and } B)+P(A^C \text{ and } B)}}$$

Go to Exercise Q3

2.3 Mutually Exclusive vs. Independence

- Two events are **mutually exclusive** if they cannot occur at the same time.

  e.g. When rolling a fair six-sided die, "Rolling a 3" and "Rolling a 4" are mutually exclusive, since you cannot roll both a 3 and a 4 at the same time.
- Two events are **independent** if one event is not impacted by the other.

  e.g. When rolling a fair six-sided die, "Rolling a 3 in the first round" and "Rolling a 3 in the second round" are independent, since the first and second roll does not influence each other.
  - This is why if A and B are independent, $P(B|A) = P(B)$, and $P(A|B) = P(A)$, since conditional on event A does not give us any new information regarding event B, and conditional on event B does not give us any new information regarding event A.
- When looking at joint probability between A and B,
  - If A and B are mutually exclusive, $P(A \text{ and } B) =$
  - If A and B are independent, $P(A \text{ and } B) =$

Exercise. Are the following two events mutually exclusive, independent, both, or neither?

1. “Weather today in Madison at 11:05am is sunny” vs. “Weather today in Madison at 11:05am is rainy”
2. “Roll a (fair six-sided) die and get an even number” vs. “Roll a (fair six-sided) die and get 4”
3. “Flip a coin and get tail” vs. “Roll a (fair six-sided) die and get 4”
4. “Flip a coin and get tail” vs. “Roll a (fair six-sided) die and get 10”

Go to Exercise Q2

3 Exercises

1. You go to Vegas to play craps; luckily, Econ 310 has prepared you to solve the following problems:

   (a) Assuming you rolled a single die and got an even number, what’s the probability the number on that die is a two?

   (b) Rolling two dice, what’s the probability of two sixes? What’s the probability of no sixes?

(c) Rolling two dice, what's the probability both dice are even? What's the probability either die is even?

(d) Rolling two dice, what's the probability they sum to seven?

2. Consider the following joint probability table:

	A	A^C
B	.15	.25
B^C	.40	.20

(a) What is $P(A \text{ and } B)$? What is $P(A)$? What is $P(B)$?

	A	A^C
B	.15	.25
B^C	.40	.20

(b) What is $P(A \text{ or } B)$?

(c) What is $P(A|B)$? What is $P(B|A)$?

(d) Are A and B independent?

3. Suppose exactly half the population is male. 5% of males and only 0.25% of females are color blind.

   (a) Draw a tree diagram to illustrate the sample space.

(b) A person is chosen at random and that person is color blind. What's the probability that the person is a male?

4. Marie is getting married tomorrow at an outdoor ceremony in the desert. In recent years, it has rained only 5 days each year. Unfortunately, the weatherman has predicted rain for tomorrow. When it actually rains, the weatherman correctly forecasts rain 90% of the time. When it doesn't rain, he incorrectly forecasts rain 20% of the time. Assume there are 365 days in a year.

   (a) Draw a tree diagram to illustrate the sample space.

   (b) What is the probability it will rain tomorrow for Marie's wedding?