Unit 2 - Probability and distributions

LO 1. Define trial, outcome, and sample space.

LO 2. Explain why the long-run relative frequency of repeated independent events settle down to the true probability as the number of trials increases, i.e. why the law of large numbers holds.

LO 3. Distinguish disjoint (also called mutually exclusive) and independent events.

- If A and B are independent, then having information on A does not tell us anything about B.
- If A and B are disjoint, then knowing that A occurs tells us that B cannot occur.
- Disjoint (mutually exclusive) events are always dependent since if one event occurs we know the other one cannot.

LO 4. Draw Venn diagrams representing events and their probabilities.

LO 5. Describe properties of probability distributions.

LO 6. Define complementary events as two events whose probabilities add up to 1.

LO 7. Distinguish between union of events (A or B) and intersection of events (A and B).

- Calculate the probability of union of events using the (general) addition rule.
- Calculate the probability of intersection of independent events using the multiplication rule.

* *Reading: Section 2.1 of OpenIntro Statistics*
* *Videos:*
  - *Basics of probability, YouTube (1:42)*
  - *Union of events and the addition rule, YouTube (3:37)*
  - *Independent events, intersection of events, multiplication rule, and Bayes' Theorem, YouTube (3:25)*
* *Test yourself:*
  1. *What is the probability of getting a head on the 6th coin flip if in the first 5 flips the coin landed on a head each time?*
  2. *True / False: Being right handed and having blue eyes are mutually exclusive events.*
  3. *P(A) = 0.5, P(B) = 0.6, there are no other possible outcomes in the sample space. What is P(A and B)?*

LO 8. Distinguish marginal and conditional probabilities.

LO 9. Construct tree diagrams to calculate conditional probabilities and probabilities of intersection of non-independent events using Bayes' theorem.

* *Reading: Section 2.2 of OpenIntro Statistics*
* *Videos:*
  - *Probability trees, Dr.Çetinkaya-Rundel (8:23)*

- *Conditional probability, YouTube (8:59 - watch from 3:33 onwards)*
- *Bayes' Theorem worked out example, YouTube, (9:20, somewhat lengthy)*
- *Another example of conditional probabilities using Bayes' Theorem, YouTube (7:20)*

* *Test yourself: 50% of students in a class are social science majors and the rest are not. 70% of the social science students and 40% of the non-social science students are in a relationship. Create a contingency table and a tree diagram summarizing these probabilities. Calculate the percentage of students in this class who are in a relationship.*

LO 10. Calculate the standardized (Z) score of a data point given the mean and standard deviation of its distribution.

LO 11. Use the Z score to determine the percentile score of a data point if the distribution is normal (using technology or normal probability tables), or to assess whether or not the particular observation would be considered unusual (regardless of the shape of the distribution).

LO 12. Depending on the shape of the distribution determine whether the median would have a negative, positive, or 0 Z score.

LO 13. Assess whether or not a distribution is nearly normal using the 68-95-99.7% rule or graphical methods such as a normal probability plot.

* *Reading: Section 3.1 and 3.2 of OpenIntro Statistics*
* *Videos:*
  - *Normal Distribution - Finding Probabilities, Dr.Çetinkaya-Rundel (6:04)*
  - *Normal Distribution - Finding Cutoff Points, Dr.Çetinkaya-Rundel (4:25)*
  - *Normal distribution and 68-95-99.7% rule, YouTube (3:18)*
  - *Z scores - Part 1, YouTube (3:03)*
  - *Z scores - Part 2, YouTube (4:01)*
* *Test yourself: True/False: In a right skewed distribution the Z score of the median is positive.*

LO 14. Determine if a random variable is binomial using the four conditions.

LO 15. Calculate the number of possible scenarios for obtaining k successes in n trials.

LO 16. Calculate probability of a given number of successes in a given number of trials using the binomial distribution.

LO 17. When number of trials is sufficiently large, use normal approximation to calculate binomial probabilities, and explain why this approach works.

* *Reading: Section 3.4 of OpenIntro Statistics*
* *Videos:*
  - *Binomial Distribution - Finding Probabilities, Dr.Çetinkaya-Rundel (8:46)*
  - *Binomial distribution, YouTube (4:25)*
  - *Mean and standard deviation of a binomial distribution, YouTube (1:39)*
* *Test yourself:*

1. *True/False: We can use the binomial distribution to determine the probability that in 10 rolls of a die the first 6 occurs on the 8th roll.*
2. *True / False: If a family has 3 kids, there are 8 possible combinations of gender order.*
3. *True/ False: When* $n = 100$ *and* $p = 0.92$ *we can use the normal approximation to the binomial to calculate the probability of 90 or more successes.*