Basic Probability

→ Why is probability important for statistics?

→ What does the long term relative frequency of an event mean?

→ How does this relate to probability?

→ What are the three basic properties of probability?

1)

2)

3)

Example: If we roll a "fair" die and it lands on an even number then we win $10. Since each side has a 1/6 probability of occurring and they cannot occur at the same time we say:

\[ P(2, 4, \text{ or } 6) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2} \]

The probability that we do not win $10 is: \(1 - P(2, 4, \text{ or } 6) = 1 - \frac{1}{2} = \frac{1}{2}\)

Independence

Definition: If the occurrence of one outcome changes the probability that another outcome occurs, the outcomes are said to be dependent.

Example: A standard deck of cards contains 26 red and 26 black. Suppose you draw two cards.

→ What is the probability that the first card drawn is red?

→ What is the probability that the second card drawn is red?

→ Why are these considered dependent events?
**Example:** Suppose you complete a paper maze and record the time it takes. Then you repeat this experiment again.

→ Why are these outcomes considered dependent?

**Definition:** If the chance that one outcome occurs is not affected by knowledge of whether another outcome has occurred, these outcomes are said to be independent.

**Example:** Consider tossing two fair coins.

→ If the first coin lands on heads, does this impact the probability that the second coin will land on heads?

**Example:** Suppose you go to the doctor and have your pulse measured twice.

→ Why are these outcomes considered independent?

Multiplication rule for independent events: If two outcomes are independent, the probability that both outcomes occur is the product of their individual probabilities

\[
P(A \text{ and } B) = P(A)P(B)
\]

**Example:** Consider tossing two coins, the outcomes for each are independent.

→ So what is the probability that both coins come up heads?

→ What is the probability that neither are heads?

→ What is the probability that exactly one is heads?

→ What is the probability that at least one is heads?
Example: One in every 500 African-Americans in the US is reported to have sickle cell anemia. One in ten is a carrier of the trait, and carriers have a 1 in 4 chance of passing the disease to their children. Suppose a carrier has three children.

→ Is the event of passing the disease to a child dependent or independent?

→ What is the probability that none of the children acquires the disease?

→ What is the probability that all three acquire the disease?

→ What is the probability that at least one acquires the disease?

→ What is the probability that exactly one child acquires the disease?

Probability with Data

Example: Psychologists tend to believe that there is a relationship between aggressiveness and order of birth. To test their belief they chose 500 children in elementary school at random and measured their aggressiveness via a test. Results are shown in the table below

<table>
<thead>
<tr>
<th></th>
<th>Firstborn</th>
<th>Not Firstborn</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aggressive</td>
<td>75</td>
<td>75</td>
</tr>
<tr>
<td>Not Aggressive</td>
<td>125</td>
<td>225</td>
</tr>
</tbody>
</table>

→ If one student is chosen at random, what is the probability that the student is first born?

→ If one student is chosen at random, what is the probability that the student is aggressive?

→ Do you think that firstborn children tend to be more aggressive than not firstborn children?
Distributions

→ What is a population distribution?

→ What is the difference between a histogram based on sample data and a population distribution?

*Definition:* (page 293) A continuous probability distribution is a smooth curve, called a density curve, that serves as a model for the population distribution of a continuous variable

→ What does the total area under the curve equal?

→ The area under the curve and above any particular interval is calculated as what?

→ Sketch a picture of a normal population distribution and mark the area between a point a and b

Each normal curve is characterized by its $\mu$ and $\sigma$

→ *Notation:*

$x \sim$

→ What does the notation mean in English
What is the formula below for?

\[ f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2} \left( \frac{x - \mu}{\sigma} \right)^2} \]

If normal curves all use the same formula, how can they look so different?

Areas Under the Normal Curve

What is the Standard Scale?

What is the special name of the distribution for \( Z \sim N(0,1) \)?

What does standardizing mean?

\[ Z = \frac{x - \mu}{\sigma} \]

What does a z-score measure?
Table 2 - gives areas under the standard normal curve.

→ If $z < 2.0$ the corresponding area is

→ If $z > 2.0$ the corresponding area is

→ Is there any useful fact regarding the normal distribution that you could use with $z < 2.0$ and $z > 2.0$?

→ If $1.0 < z < 2.0$ the corresponding area is

→ Is there a difference between $1.0 < z < 2.0$ and $1.0 \leq z \leq 2.0$? Why or why not?

→ If $z < 2.56$ the corresponding area is

Application to Data

Example: Suppose that the average systolic blood pressure (SBP) for a Los Angeles freeway commuter follows a normal distribution with mean 130 mmHg and standard deviation 20 mmHg.

→ Find the percentage of LA freeway commuters that have a SBP less than 100.

→ Does a negative z-score mean that it is bad (and a positive z-score is good)?
What percentage of LA freeway commuters have SBP greater than 155 mmHg?

Between 120 and 175?

What is the probability that a randomly selected freeway commuter will have a SBP less than 100?

**Percentiles**

Suppose we want to find the value $z$ that cuts off the top 2.5% of the distribution.

What is $Z_{0.025}$ (ie. What $Z$ is the cut point for the 97.5th percentile)?

*Example:* LA freeway commuter SBP (continued)

Find the SBP that is the 90$^{th}$ percentile.
Find the SBP that is the 10\textsuperscript{th} percentile (HINT: Use the symmetry of the normal distribution and part of your previous answer).

Assessing Normality

- There are several methods for assessing normality

Why do we care if the data is normally distributed?

Normal Probability Plots

- What are we looking for on a normal probability plot?
Example: Suppose we have the height for 11 women and we want to check to see if this data is normally distributed.

<table>
<thead>
<tr>
<th>height (in)</th>
<th>Nscore</th>
</tr>
</thead>
<tbody>
<tr>
<td>61.0</td>
<td>-1.59322</td>
</tr>
<tr>
<td>62.5</td>
<td>-1.06056</td>
</tr>
<tr>
<td>63.0</td>
<td>-0.72791</td>
</tr>
<tr>
<td>64.0</td>
<td>-0.46149</td>
</tr>
<tr>
<td>64.5</td>
<td>-0.22469</td>
</tr>
<tr>
<td>65.0</td>
<td>0.00000</td>
</tr>
<tr>
<td>66.5</td>
<td>0.22469</td>
</tr>
<tr>
<td>67.0</td>
<td>0.46149</td>
</tr>
<tr>
<td>68.0</td>
<td>0.72791</td>
</tr>
<tr>
<td>68.5</td>
<td>1.06056</td>
</tr>
<tr>
<td>70.5</td>
<td>1.59322</td>
</tr>
</tbody>
</table>

→ Does this data appear to be approximately normally distributed? Why?

→ What is the RULE that we will use on the MTB normality plots to tell if the distribution is approximately normal?
**Diagnostics**

→ The two plots to the right and histogram below are for the same data set. What do they tell you about the shape of the distribution? Why?

![Scatterplot of height (in) vs Nscore](image)

![Histogram of height (in)](image)

→ The plots below would also not be considered normally distributed. What do they tell you about the shape of the distribution? Why?

![Probability Plot of height (in)](image)

![Histogram of height (in)](image)