Mathematics and Biostatistics

Probability

1st Semester 2021

Lecture 6

Review

- Things we look for in a distribution
  1. Central Tendency
     a. MEAN -- average
     b. MEDIAN -- middle value
     c. MODE -- most frequently observed value.
  2. Variability of values and their spread
     a. Quartiles -- Median
     b. Standard deviation -- Mean
  3. Shape of the distribution.
     a. Normal Curve
     b. Normal distribution (Gaussian distribution)
     c. The 68–95–99.7 rule
  4. Gaps and clumping of values

Probability

- Probability can be defined as a measure of how likely an event was to occur.
- A probability of 0 indicates the event would not occur
- A probability of 1 indicates the event was sure to occur.
- Values between 0 and 1 indicated occurrence somewhere in between the extremes.
- We will denote events with capital letters: A, B, C etc. These can be viewed as generic forms of an outcome in a statistical analysis.
- If we want to write the probability of an event A we will use the notation P(A).
- Read as the probability of A.

Probability

- Determine a single event with a single outcome.
- Identify the total number of outcomes that can occur.
- Divide the number of events by the number of possible outcomes.

- Example When a coin is tossed, there are two possible outcomes: heads (H) or tails (T)
- We say that the probability of the coin landing H is

$$P(H)=\frac{\text{the number of heads (H)}}{\text{the number of possible outcomes}}=\frac{1}{2}$$

- And the probability of the coin landing T is

$$P(T)=\frac{\text{the number of tails (T)}}{\text{the number of possible outcomes}}=\frac{1}{2}$$

Probability

- The term mutually exclusive is used to describe events which cannot simultaneously occur.
- The occurrence of one rules out the occurrence of the other.
- A clear example is the set of outcomes of a single coin toss, which can result in either heads or tails, but not both.

Probability

- A joint event is the simultaneous occurrence of two or more events.
- Example: the occurrence of both a specific exposure and the occurrence of a specific disease.
- The probability of a specific disease exposure has an impact on the probability of a specific disease occurrence.
- It is the examination of joint events that leads to the determination of association between two variables..

Probability

- Conditional events are those that occur in a specified sequential order, restricting the focus in some manner.
- The occurrence of disease after an exposure. Or as we say the occurrence of disease given exposure.

Probability

- Conditional events are those that occur in a specified sequential order, restricting the focus in some manner.
- The occurrence of disease after an exposure. Or as we say the occurrence of disease given exposure.

Probability

- The probabilities over the entire sample space must sum to 1.
- The first of these is that if we have an event A in some sample space. Then

$$P(A) + P(\text{not } A) = 1$$

- A simple example is a sample space of two events: A and B.

Then

$$P(A) + P(B) = 1, \quad P(B) = 1 - P(A) \text{ and likewise, } P(A) = 1 - P(B).$$

Probability

- This rule becomes useful in situations where you cannot directly determine P(A) but you can compute the probability of not A events.

  Then

  $$P(A)= 1-P(\text{Not } A)$$

- Note (Not A) $\leftrightarrow$ ($\neg$A), read as not A.

Probability

- If two events A and B are mutually exclusive,

  Then

  P(A or B) = P(A) + P(B).

- If the events are not mutually exclusive, then this rule does not hold.
- Two events A and B are said to be independent or not associated if and only if:

  P(A and B) = P(A)*P(B).

Probability

- Example of independence.

The data Yield:

- P(D) = 13/100
- P(E) = 50/100
- P(D and E) = 10/100

- from these we can determine if disease and exposure are independent.
- If they are independent, then 0.10 should be close to the produce of 0.13 and 0.50 0.13*0.50 = 0.065

Disease▶ Exposure▼	Yes	NO	Total
Yes	10	40	50
No	3	47	50
Total	13	87	100

Sensitivity and Specificity

- SENSITIVITY measures the proportion of positives that are correctly identified
- is the probability of a test to indicate the presence of the disease when the person is truly diseased.
- a correct behavior for a diagnostic test, and a conditional probability.
- This is a True Positive: TP

Disease ▶ Test ▼	Yes	NO
Yes	TP	FP
No	TN	FN

Sensitivity and Specificity

- SPECIFICITY measures the proportion of negatives that are correctly identified

- is the probability of the test to be negative when the person does not have the disease.

- Again, this is a correct behavior for the test.

- This is a True Negative: TN

Disease ▶ Test ▼	Yes	NO
Yes	TP	FP
No	TN	FN

Sensitivity and Specificity

- There are two errors that can occur:
- A false positive-- The probability that the test is positive when in fact the person is truly negative for the disease. (FP)
- A false negative– The probability that the test is negative when in fact the person does in fact have the disease. (FN)

Disease ▶ Test ▼	Yes	NO
Yes	TP	FP
No	FN	TN

Sensitivity and Specificity

- Sensitivity = TP/(TP+FN) x 100
- Specificity = TN/(FP+TN) x 100
- **SENSITIVITY** =(10/13) x 100 = 77%, The test correctly found 77% of the diseased individuals.
- **SPECIFICITY** = (47/87) x 100 = 54%, the test correctly identified 54% of those without disease.

Disease ▶ Exposure ▼	Yes	NO	Total
Yes	10	40	50
No	3	47	50
Total	13	87	100

The end of lecture