Biostatistics

Lecture 2

Probability

Basic Probability Concepts

- Foundation of statistics because of the concept of sampling and the concept of variation or dispersion and how likely an observed difference is due to chance
- Probability statements used frequently in statistics
  - e.g., we say that we are 90% sure that an observed treatment effect in a study is real

Characteristics of Probabilities

- Probabilities are expressed as fractions between 0.0 and 1.0
  - e.g., 0.01, 0.05, 0.10, 0.50, 0.80
  - Probability of a certain event = 1.0
  - Probability of an impossible event = 0.0
- Application to biomedical research
  - e.g., ask if results of study or experiment could be due to chance alone
  - e.g., significance level and power
  - e.g., sensitivity, specificity, predictive values

Definition of Probabilities

- If some process is repeated a large number of times, n, and if some resulting event with the characteristic of E occurs m times, the relative frequency of occurrence of E, m/n, will be approximately equal to the probability of E: $P(E)=m/n$
- Also known as relative frequency

Elementary Properties of Probabilities - I

- Probability of an event is a non-negative number
  - Given some process (or experiment) with *n* mutually exclusive outcomes (events), E_1, E_2, …, E_n, the probability of any event E_i is assigned a nonnegative number
  - $P(E_i) \geq 0$
  - key concept is mutually exclusive outcomes - cannot occur simultaneously
  - Given previous definition, not clear how to construct a negative probability

Elementary Properties of Probabilities - II

- Sum of the probabilities of mutually exclusive outcomes is equal to 1
  - Property of exhaustiveness
    - refers to the fact that the observer of the process must allow for all possible outcomes
  - $P(E_1) + P(E_2) + \ldots + P(E_n) = 1$
  - key concept is still mutually exclusive outcomes

Elementary Properties of Probabilities - III

- Probability of occurrence of either of two mutually exclusive events is equal to the sum of their individual probabilities
  - Given two mutually exclusive events A and B
  - P(A or B) = P(A) + P(B)
  - If not mutually exclusive, then problem becomes more complex

Elementary Properties of Probabilities - IV

- For two independent events, A and B, occurrence of event A has no effect on probability of event B
  - $P(A \cup B) = P(B) + P(A)$
  - $P(A \mid B) = P(A)$
  - $P(B \mid A) = P(B)$
  - $P(A \cap B) = P(A) \times P(B)$*
    - * Key concept in contingency table analysis

Elementary Properties of Probabilities - V

- Conditional probability
  - Conditional probability of B given A is given by:
  - $P(B \mid A) = P(A \cap B) / P(A)$
  - Probability of the occurrence of event B given that event A has already occurred.
  - Ex. given that a test for bladder cancer is positive, what is the probability that the patient has bladder cancer

Elementary Properties of Probabilities - VI

- Given some variable that can be broken down into m categories designated A_1, A_2, …, A_m and another jointly occuring variable that is broken down into n categories designated by B_1, B_2, …, B_n, the marginal probability of A_i, $P(A_i)$, is equal to the sum of the joint probabilities of A_i with all the categories of B. That is,

$$P(A_i) = \sum P(A_i \cap B_j)$$

Elementary Properties of Probabilities - VII

- For two events A and B, where P(A) + P(B) = 1, then

$$P(\overline{A}) = 1 - P(A)$$

- Important concept in contingency table analysis

Elementary Properties of Probabilities - VIII

- Multiplicative Law
  - For any two events A and B,
  - $P(A \cap B) = P(A)\, P(B \mid A)$
    - Joint probability of A and B = Probability of B times Probability of A given B
- Addition Law
  - For any two events A and B
  - $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
    - Probability of A or B = Probability of A plus Probability of B minus the joint Probability of A and B

Calculating the Probability of an Event

- Example 1 - TV watching by Income
  - Marginal probabilities
  - Joint probabilities
  - Conditional probabilities
  - Conditional probabilities with multiplicative law
- Example 2 - Physical Appearance by BMI
  - Marginal probabilities
  - Joint probabilities
  - Conditional probabilities
  - Conditional probabilities with multiplicative law

Screening Tests

- False Positives
  - Test indicates a positive status when the true status is negative
- False Negatives
  - Test indicates a negative status when the true status is positive

	Disease		
Test Result	Present	Absent	Total
Positive	a	b	$a + b$
Negative	c	d	$c + d$
Total	$a + c$	$b + d$	n

Questions about Screening Tests

- Given that a patient has the disease, what is the probability of a positive test results?
- Given that a patient does not have the disease, what is the probability of a negative test result?
- Given a positive screening test, what is the probability that the patient has the disease?
- Given a negative screening test, what is the probability that the patient does not have the disease?

Sensitivity and Specificity

- Sensitivity of a test is the probability of a positive test result given the presence of the disease
  - $a / (a + c)$
- Specificity of a test is the probability of a negative test result given the absence of the disease
  - $d / (b + d)$

Predictive Values

- Predictive value positive of a test is the probability that the subject has the disease given that the subject has a positive screening test
  - P(D | T)
- Predictive value negative of a test is the probability that a subject does not have the disease, given that the subject has a negative screening test
  - P(D- | T-)

Bayes' Theorem

- Predictive value positive

$$P(D \mid T) = \frac{P(T \mid D)P(D)}{P(T \mid D)P(D) + P(T \mid \overline{D})P(\overline{D})}$$

$$P(D \mid T) = \frac{sensitivity\ x\ P(D)}{sensitivity\ x\ P(D) + (1 - specificity)\ x\ (1 - P(D))}$$

- Predictive value negative

$$P(\overline{D} \mid \overline{T}) = \frac{P(\overline{T} \mid \overline{D})P(\overline{D})}{P(\overline{T} \mid \overline{D})P(\overline{D}) + P(\overline{T} \mid D)P(D)}$$

$$P(\overline{D} \mid \overline{T}) = \frac{specificity\ x\ (1 - P(D))}{specificity\ x\ (1 - P(D)) + (1 - sensitivity)\ x\ P(D)}$$

ROC Curves

- Receiver Operator Characteristic (ROC) plot sensitivity vs. 1-specificity of a screening test over the full range of cutpoints for declaring the test positive for the disease
- Extremely convenient to identify an appropriate cutpoint for declaring the screening test positive
- Typically calculated as part of a logistic regression

Prevalence and Incidence

- Prevalence is the probability of having the disease or condition at a given point in time regardless of the duration
- Incidence is the probability that someone without the disease or condition will contract it during a specified period of time