L- 3: Descriptive Statistics

Today…..

- Recall the past for a while_ Simple tools
- Visualization of data
- Basics of probability
- Discussion & Problems on probability
- Conditional probability
- Box plot

Visualization

- Summary gives an idea about the data

  summary(income)

Min	*1st QU.*	*Median*	*Mean*	*3rd Qu*	*Max*
- 7.8	*12.5*	*32.0*	*52.03*	*67.2*	*585*

- Visualization – why

3

A Survey conducted by a bank revealed that 40% of the accounts are savings accounts and 35% of the accounts are current accounts and the balance are loan accounts.

- What is the probability that an account taken at random is a loan account ?
- What is the probability that an account taken at random is NOT savings account ?
- What is the probability that an account taken at random is NOT a current account
- What is the probability that an account taken at random is a current account or a loan account?

4

From a Hospital data it is found that 45% of the patients are having high B.P. Also it was found that 35% of these patients having high B P is also having diabetes.

What is the probability that a patient having high BP is also diabetic

Conditional Probability

The probability of event B given that event A has occurred $P(B|A)$ or, the probability of event A given that event B has occurred $P(A|B)$

Conditional Probability

Definition

The conditional probability of an event B given an event A, denoted as $P(B \mid A)$, is

$$P(B \mid A) = P(A \cap B)/P(A) \tag{2-9}$$

for $P(A) > 0$.

Multiplication and Total Probability Rules

innovate achieve lead

Multiplication Rule

$$P(A \cap B) = P(B \mid A)P(A) = P(A \mid B)P(B) \qquad (2\text{-}10)$$

innovate achieve lead

Total Probability Rule (two events)

For any events A and B,

$$P(B) = P(B \cap A) + P(B \cap A') = P(B \mid A)P(A) + P(B \mid A')P(A') \quad (2\text{-}11)$$

Independence

Definition (two events)

Two events are **independent** if any one of the following equivalent statements is true:

(1) $P(A|B) = P(A)$

(2) $P(B|A) = P(B)$

(3) $P(A \cap B) = P(A)P(B)$ (2-13)

Bayes' Theorem

Definition

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad \text{for} \quad P(B) > 0 \qquad (2\text{-}15)$$

$$P(B) = P(E_1 \cap B) + P(E_2 \cap B) + \cdots + P(E_n \cap B)$$

For each

$$P(E_i \cap B) = P(B \mid E_i)P(E_i)$$

$$\begin{aligned} P(B) &= P(E_1 \cap B) + P(E_2 \cap B) + \cdots + P(E_n \cap B) \\ &= P(B \mid E_1)P(E_1) + P(B \mid E_2)P(E_2) + \cdots + P(B \mid E_n)P(E_n) \\ &= \sum_{i=1}^{n} P(B \mid E_i)P(E_i) \end{aligned}$$

Bayes' Theorem

Bayes' Theorem

If $E_1, E_2, \ldots, E_k$ are k mutually exclusive and exhaustive events and B is any event,

$$P(E_1|B) = \frac{P(B|E_1)P(E_1)}{P(B|E_1)P(E_1) + P(B|E_2)P(E_2) + \cdots + P(B|E_k)P(E_k)} \qquad (2\text{-}16)$$

for $P(B) > 0$

Applications

- Diagnostic tests in medicine
- Telecommunication
- Customer service
- Trouble shooting in engineering processes & systems

Random Variables

We now introduce a new term

Instead of saying that the possible outcomes are 1,2,3,4,5 or 6, we say that ***random variable*** X can take values {1,2,3,4,5,6}. **A random variable is an expression whose value is the outcome of a particular experiment.**

The random variables can be either *discrete* or *continuous*.

It's a convention to use the upper case letters (X, Y) for the names of the random variables and the lower case letters (x, y) for their possible particular values.

Random Variables

Definition

A **discrete** random variable is a random variable with a finite (or countably infinite) range.

A **continuous** random variable is a random variable with an interval (either finite or infinite) of real numbers for its range.

Random Variables

Examples of Random Variables

Examples of **continuous** random variables:

electrical current, length, pressure, temperature, time, voltage, weight

Examples of **discrete** random variables:

number of scratches on a surface, proportion of defective parts among 1000 tested, number of transmitted bits received in error

The Probability Function for discrete random variables

We assigned a probability 1/6 to each face of the dice. In the same manner, we should assign a probability 1/2 to the sides of a coin.

What we did could be described as *distributing the values of probability* between different elementary events:

$$P(X=x_k)=p(x_k), k=1,2,\dots$$

It is convenient to introduce the *probability function p(x)* :

$$P(X=x)=p(x)$$

Continuous distribution and the probability density function

A random variable X is said to have a ***continuous distribution*** with ***density function*** *f(x)* if for all $a \leq b$ *we have*

$$P(a \leq X \leq b) = \int_a^b f(x)dx \qquad (1.15)$$

$$\int_\Omega f(x) = 1 \qquad (1.16)$$

$$P(E) = \int_E f(x)dx \qquad (1.17)$$

Expected Value

$$E(X) = \sum_{i=1}^{n} X_i P(X_i)$$

Variance

$$\sigma^2 = \sum_{i=1}^{n} [X_i - E(X)]^2 P(X_i)$$

Thanks