Statistical Methods for the Biological, Environmental, and Health Sciences

STAT 007

Discrete Probability Distributions

Chapter 5

Probability Distributions

Section 5-1

- In this section we will:
  - Define random variables and their probability distributions.
  - Introduce parameters, such as mean, variance and standard deviations, of probability distributions.

Parameters of a Probability Distribution

Example

Consider a procedure, in which it is of interest to describe the number of girls that are born in two births. Consider a random variable that describes the number of girls in two births, this is

$$x = \text{number of girls in two births}$$

and the following probability distribution $P(x) = \frac{1}{2(2-x)!x!}$.

- Find the mean and the variance of the number of girls in two births.
- Is the value 2 a significantly high value?

Practice

Look at the exercises at the end of Section 5-1 in page 710.

Specially, look at exercises:
1, 2, 3, 5, 6, 7-12, 13, 14, 15, 16, 17, 18.

Binomial Probability Distributions

Sections 5-2

- In this section we will:
  - Introduce the Binomial probability distribution to measure uncertainty in procedures involving number of successes in a finite number of trial.
  - Identify parameters such as mean, variance, and standard deviation.

Binomial Distribution

Definition

A binomial probability distribution can be used when

1. *The procedure has a fixed number of trials. (A trial is a single success/failure observation.)*
2. *The trials are independent.*
3. *Each trial must have all outcomes classified into exactly two categories.*
4. *The probability of success remains the same in all trials.*

*The probability distribution formula of the Binomial distribution with parameters **n** and **p** is given by*

$$P(x) = \frac{n!}{(n-x)!x!} p^x (1-p)^{n-x}, \quad x = 0, 1, 2, \ldots, n,$$

$$\begin{aligned} n &= \textit{total number of trials}, \\ x &= \textit{number of successes in n trials}, \\ p &= \textit{probability of succes in ech trial}. \end{aligned}$$

***Parameters**:* $\mu = np, \quad \sigma^2 = np(1-p), \quad \sigma = \sqrt{np(1-p)}.$

Example

When Gregor Mendel conducted his famous hybridization experiments, he used peas with green pods and peas with yellow pods. Because green is dominant and yellow is recessive, when crossing two parents with the green/yellow pair of genes, we expect that 3/4 of the offspring peas should have green pods. That is, $P(\text{green pod}) = 3/4 = 0.75$. Assume that all parents have the green/yellow combination of genes.

a) Would the Binomial probability distribution be a fit if we want to find the probability that exactly three out of five offspring peas have green pods?

b) Find the probability of getting exactly 3 peas with green pods when 5 offspring peas are generated.

c) In an actual experiment, Mendel generated 580 offspring peas. He claimed that 75%, of them would have green pods. The actual experiment resulted in 428 peas with green pods.
   - c1) Find the mean and standard deviation for the numbers of peas with green pods.
   - c2) Would you conclude that Mendels actual result of 428 peas with green pods is significantly low or significantly high? Does this suggest that Mendels value of 75% is wrong?

 Discrete Probability Distributions

Practice

Look at the exercises at the end of Section 5-2 in page 201.

Specially, look at exercises:
1, 2, 10, 11, 12, 15-20, 21-24, 25, 26.

Normal Probability Distributions

Chapter 6

Real Applications of the Normal Distribution and the Standard Normal Distribution

Sections 6-1 and 6-2

- In this sections we will:
  - Introduce the normal distribution and the standard normal distribution.
  - Identify their parameters: mean, variance, and standard deviation.

- First we introduce the normal (or bell-shaped) distribution that depends on two parameters: the population mean, μ, and population standard deviation, σ.
- Then we introduce the standard normal distribution, whose parameters are: population mean, $\mu = 0$, and population standard deviation, $\sigma = 1$.
- The normal distribution is used for continuous random variables that can take negative, positive, or both values.
- For continuous random variables, probabilities are computed as *areas under a curve*.
- The curve used to compute probabilities is called **density curve**.
- Density curves satisfy two requirements:
  1. are always are positive.
  2. the total area under the curve is exactly one.

Normal Distribution

Definition

The density curve of a random variable that follows a normal distribution with mean $\boldsymbol{\mu}$ and standard deviation $\boldsymbol{\sigma}$ is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad -\infty < x < \infty.$$

where

$$\begin{aligned} x &= \textit{value on the real line}, \\ \mu &= \textit{mean of } x, \\ \sigma &= \textit{standard deviation of } x, \\ \pi &= 3.14159 \quad e = 2.711828. \end{aligned}$$

Parameters: $\mu = \mu, \quad \sigma^2 = \sigma^2, \quad \sigma = \sigma.$

 Normal Probability Distributions

Standard Normal Distribution

- When computing probabilities for normal distributions, we usually transform the values of x to values from a standard normal distribution. This is, we find the z score of the value x.
- z scores are found as $\boldsymbol{z = \frac{x-\mu}{\sigma}}$.

Definition

The density curve of a random variable that follows a standard normal distribution with mean 0 *and standard deviation* 1 *is given by*

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z)^2}, \quad -\infty < z < \infty.$$

where

$$z = \textit{value on the real line},$$
$$\pi = 3.14159 \quad e = 2.711828.$$

Parameters: $\mu = 0, \quad \sigma^2 = 1, \quad \sigma = 1.$

Normal Distribution

Example

Assume that pulse rates of adult males are normally distributed with a mean of 69.6 bpm and a standard deviation of 11.3 bpm.

a) Sketch the density curve of the distribution of the pulse rates of adult males.

b) Males with pulse rates greater than 100 bpm are considered to be at a high risk of stroke, heart disease, or cardiac death. Find the proportion of men that are at risk of stroke, heart disease, or cardiac death.

c) Normal pulse rates are generally considered to be between 60 bpm and 100 bpm. Find the proportion of men that have normal pulse rate.

d) Find the pulse rate that separates the highest 1% from the lowest 99%.

 Normal Probability Distributions

Practice

Look at the exercises at the end of Section 6-1 in page 229.

Specially, look at exercises:
1, 3, 9-12, 13-16, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37-40.

Look at the exercises at the end of Section 6-2 in page 238.

Specially, look at exercises:
1, 2, 3, 5-8, 9-12, 13-20, 24.