Chapter 5

Joint Probability Distributions and Random Samples

5.3

Statistics and their Distributions

Statistic

A *statistic* is any quantity whose value can be calculated from sample data. Prior to obtaining data, there is uncertainty as to what value of any particular statistic will result. A statistic is a random variable denoted by an uppercase letter; a lowercase letter is used to represent the calculated or observed value of the statistic.

Random Samples

The rv's $X_1,\ldots,X_n$ are said to form a simple *random sample* of size n if

1. The X_i's are independent rv's.
2. Every X_i has the same probability distribution.

Simulation Experiments

The following characteristics must be specified:

1. The statistic of interest.
2. The population distribution.
3. The sample size n.
4. The number of replications k.

5.4

The Distribution of the Sample Mean

Using the Sample Mean

Let $X_1,\ldots, X_n$ be a random sample from a distribution with mean value μ and standard deviation σ. Then

1. $E\left(\overline{X}\right) = \mu_{\overline{X}} = \mu$
2. $V\left(\overline{X}\right) = \sigma^2_{\overline{X}} = \sigma^2/n$

In addition, with $T_o = X_1 + \ldots + X_n$,
$E(T_o) = n\mu,\ V(T_o) = n\sigma^2$, and $\sigma_{T_o} = \sqrt{n}\sigma.$

Normal Population Distribution

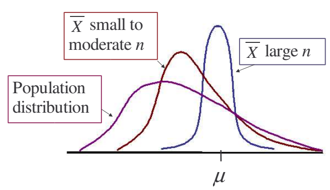
Let $X_1,\ldots, X_n$ be a random sample from a normal distribution with mean value μ and standard deviation σ. Then for any n, $\overline{X}$ is normally distributed.

μ

The Central Limit Theorem

Let $X_1,\ldots, X_n$ be a random sample from a distribution with mean value μ and variance σ^2. Then if n sufficiently large, $\overline{X}$ has approximately a normal distribution with $\mu_{\overline{X}} = \mu$ and $\sigma^2_{\overline{X}} = \sigma^2/n$, and T_o also has approximately a normal distribution with $\mu_{T_o} = n\mu,\ \sigma_{T_o} = n\sigma^2$. The larger the value of n, the better the approximation.

The Central Limit Theorem



Rule of Thumb

If $n > 30$, the Central Limit Theorem can be used.

5.5

The Distribution of a Linear Combination

Linear Combination

Given a collection of n random variables $X_1,\ldots, X_n$ and n numerical constants $a_1,\ldots,a_n$, the rv

$$Y = a_1X_1 + \ldots + a_nX_n = \sum_{i=1}^{n} a_iX_i$$

is called a *linear combination* of the X_i's.

Expected Value of a Linear Combination

Let $X_1,\ldots, X_n$ have mean values $\mu_1,\mu_2,\ldots,\mu_n$ and variances of $\sigma_1^2,\sigma_2^2,\ldots,\sigma_n^2$, respectively

Whether or not the X_i's are independent,

$$E(a_1X_1 + \ldots + a_nX_n) = a_1E(X_1) + \ldots + a_nE(X_n) = a_1\mu_1 + \ldots + a_n\mu_n$$

Variance of a Linear Combination

If $X_1,\ldots, X_n$ are independent,

$$V(a_1X_1 + \ldots + a_nX_n) = a_1^2V(X_1) + \ldots + a_n^2V(X_n) = a_1^2\sigma_1^2 + \ldots + a_n^2\sigma_n^2$$

and

$$\sigma_{a_1X_1+\ldots+a_nX_n} = \sqrt{a_1^2\sigma_1^2 + \ldots + a_n^2\sigma_n^2}$$

Variance of a Linear Combination

For any $X_1,\ldots, X_n$,

$$V(a_1X_1 + \ldots + a_nX_n) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_ia_j\text{Cov}(X_i, X_j)$$

Difference Between Two Random Variables

$E(X_1 - X_2) = E(X_1) - E(X_2)$

and, if X_1 and X_2 are independent,

$V(X_1 - X_2) = V(X_1) + V(X_2)$

Difference Between Normal Random Variables

If $X_1, X_2, \ldots X_n$ are independent, normally distributed rv's, then any linear combination of the X_i's also has a normal distribution. The difference $X_1 - X_2$ between two independent, normally distributed variables is itself normally distributed.