Section 5.5 - The Distribution of a Linear Combination

DEFINITION

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

$$Y = a_1X_1 + \cdots + a_nX_n = \sum_{i=1}^{n} a_iX_i \tag{5.7}$$

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

$E[aX + b] = aE[X] + b$

PROPOSITION

Let $X_1, X_2, \ldots, X_n$ have mean values $\mu_1, \ldots, \mu_n$, respectively, and variances $\sigma_1^2, \ldots, \sigma_n^2$, respectively.

1. Whether or not the X_i's are independent,

$$\begin{aligned} E(a_1X_1 + a_2X_2 + \cdots + a_nX_n) &= a_1E(X_1) + a_2E(X_2) + \cdots + a_nE(X_n) \\ &= a_1\mu_1 + \cdots + a_n\mu_n \end{aligned} \tag{5.8}$$

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

$$\begin{aligned} V(a_1X_1 + a_2X_2 + \cdots + a_nX_n) &= a_1^2V(X_1) + a_2^2V(X_2) + \cdots + a_n^2V(X_n) \\ &= a_1^2\sigma_1^2 + \cdots + a_n^2\sigma_n^2 \end{aligned} \tag{5.9}$$

and

$$\sigma_{a_1X_1 + \cdots + a_nX_n} = \sqrt{a_1^2\sigma_1^2 + \cdots + a_n^2\sigma_n^2} \tag{5.10}$$

3. For any $X_1, \ldots, X_n$,

$$V(a_1X_1 + \cdots + a_nX_n) = \sum_{i=1}^{n}\sum_{j=1}^{n} a_ia_j\text{Cov}(X_i, X_j) \tag{5.11}$$

Ex: A gas station sells three grades of gasoline: regular, extra, and super. These are priced at \$3.00, \$3.20, and \$3.40 per gallon. Let

$X_1 =$ number of gallons of regular gas sold on a given day

$X_2 =$ number of gallons of extra gas sold on a given day

$X_3 =$ number of gallons of super gas sold on a given day

Suppose these are independent rv's with

$\langle \mu_1 = 1000,\ \mu_2 = 500,\ \mu_3 = 300,\ \sigma_1 = 100,\ \sigma_2 = 80,\ \sigma_3 = 50 \rangle$

Determine the expected revenue from sales on this day.

$$Y = 3X_1 + 3.2X_2 + 3.4X_3$$

$$E[Y] = 3E[X_1] + 3.2E[X_2] + 3.4E[X_3] = 3000 + 1600 + 1020$$

$$= \$5620.$$

What is the standard deviation of sales revenue on this day?

$$V[Y] = 3^2 V[X_1] + 3.2^2 V[X_2] + 3.4^2 V[X_3]$$

$$= 9(100)^2 + 10.24(80)^2 + 11.56(50)^2 = 184436$$

$$\sigma_Y = \sqrt{184436} = 429.46$$

PROPOSITION

If $X_1, X_2, \ldots, X_n$ are independent, normally distributed rv's (with possibly different means and/or variances), then any linear combination of the X_i's also has a normal distribution. In particular, the difference $X_1 - X_2$ between two independent, normally distributed variables is itself normally distributed.

$$X_1 - X_2 \sim N\left(\mu_1 - \mu_2,\ \underbrace{\sigma_1^2 + \sigma_2^2}\right)$$

$$V[X_1 - X_2] = 1^2\, V[X_1] + (-1)^2\, V[X_2]$$

More Examples:

1. A certain beverage company is suspected of under filling its cans of soft drink. The company advertises that its cans contain, on the average, 12 ounces of soda with standard deviation 0.4 ounce.

Compute the probability that a random sample of 50 cans produces a sample mean fill of 11.9 ounces or less.

$n = 50 \quad \mu_{\bar{x}} = 11.9 \quad \sigma_{\bar{x}} = .4/\sqrt{50}$

$$P(\bar{x} \leq 11.9) = P\left(z \leq \frac{11.9 - 12}{.4/\sqrt{50}}\right)$$

$$\approx .0385$$

2. Suppose that the high daily temperatures in a small town in the eastern United States have a mean of 58.6◦F and a standard deviation of 9.8◦F.

a) Suppose that a random sample of 16 high temperatures was chosen and the sample mean was recorded. Give the values of the mean and the standard deviation of the sample mean.

$\mu_{\bar{x}} = 58.6°F \quad \sigma_{\bar{x}} = 9.8/\sqrt{16} = 2.45°F$

b) If a random sample of size 4 of average high daily temperatures is selected, find the probability that the mean of this sample of average high daily temperatures is less than 57 ◦F.

$P(\bar{x} < 57) = \text{normalcdf}\left(-9999, 57, 58.6, \frac{9.8}{\sqrt{4}}\right)$

$\approx .3720$

c) If a random sample of size 25 of average high daily temperatures is selected, find the probability that the mean of this sample of average high daily temperatures is between 57◦F and 61◦F.

$n = 25$

$P(57 < \bar{x} < 61) = \text{normalcdf}(57, 61, 58.6, 9.8/\sqrt{25})$

$= \text{pnorm}(61, 58.6, 9.8/\sqrt{25}) - \text{pnorm}\left(57, 58.6, \frac{9.8}{\sqrt{25}}\right)$

$\approx .6825$

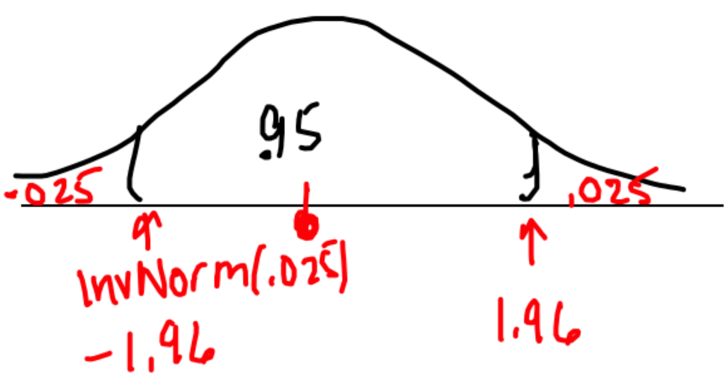
$\mu = 1500 \quad \sigma = 100$

3. A certain brand of light bulb has a mean lifetime of 1500 hours with a standard deviation of 100 hours. If the bulbs are sold in boxes of 25, what are the <u>parameters</u> of the distribution of sample means? $n = 25$

$$\mu_{\bar{x}} = 1500$$

$$\sigma_{\bar{x}} = 100/\sqrt{25} = 20$$

Find the interval that would contain the middle 95% of $\bar{x}$'s



$$\underline{1500} \pm 1.96(\underline{20})$$

$$(1460.8, 1539.2)$$