6.3 The Sample Mean and the Sample Variance

Let $X_1, \ldots, X_n$ be independent $N(\mu, \sigma^2)$ random variables; we sometimes refer to them as a **sample** from a normal distribution. In this section, we will find the joint and marginal distributions of

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

These are called the **sample mean** and the **sample variance,** respectively. First note that because $\overline{X}$ is a linear combination of independent normal random variables, it is normally distributed with

$$E(\overline{X}) = \mu$$

$$\mathrm{Var}(\overline{X}) = \frac{\sigma^2}{n}$$

As a preliminary to showing that $\overline{X}$ and S^2 are independently distributed, we establish the following theorem.

THEOREM **A**

The random variable $\overline{X}$ and the vector of random variables $(X_1 - \overline{X}, X_2 - \overline{X}, \ldots, X_n - \overline{X})$ are independent.

Proof

At the level of this course, it is difficult to give a proof that provides sufficient insight into why this result is true; a rigorous proof essentially depends on geometric properties of the multivariate normal distribution, which this book does not cover. We present a proof based on moment-generating functions; in particular, we will show that the joint moment-generating function

$$M(s, t_1, \ldots, t_n) = E\{\exp[s\overline{X} + t_1(X_1 - \overline{X}) + \cdots + t_n(X_n - \overline{X})]\}$$

factors into the product of two moment-generating functions—one of $\overline{X}$ and the other of $(X_1 - \overline{X}), \ldots, (X_n - \overline{X})$. The factoring implies (Section 4.5) that the random variables are independent of each other and is accomplished through some algebraic trickery. First we observe that since

$$\sum_{i=1}^{n} t_i(X_i - \overline{X}) = \sum_{i=1}^{n} t_i X_i - n\overline{X}\,\overline{t}$$

then

$$s\overline{X} + \sum_{i=1}^{n} t_i(X_i - \overline{X}) = \sum_{i=1}^{n} \left[\frac{s}{n} + (t_i - \bar{t})\right] X_i$$

$$= \sum_{i=1}^{n} a_i X_i$$

where

$$a_i = \frac{s}{n} + (t_i - \bar{t})$$

Furthermore, we observe that

$$\sum_{i=1}^{n} a_i = s$$

$$\sum_{i=1}^{n} a_i^2 = \frac{s^2}{n} + \sum_{i=1}^{n} (t_i - \bar{t})^2$$

Now we have

$$M(s, t_1, \ldots, t_n) = M_{X_1 \cdots X_n}(a_1, \ldots, a_n)$$

and since the X_i are independent normal random variables, we have

$$M(s, t_1, \ldots, t_n) = \prod_{i=1}^{n} M_{X_i}(a_i)$$

$$= \prod_{i=1}^{n} \exp\left(\mu a_i + \frac{\sigma^2}{2} a_i^2\right)$$

$$= \exp\left(\mu \sum_{i=1}^{n} a_i + \frac{\sigma^2}{2} \sum_{i=1}^{n} a_i^2\right)$$

$$= \exp\left[\mu s + \frac{\sigma^2}{2}\left(\frac{s^2}{n}\right) + \frac{\sigma^2}{2} \sum_{i=1}^{n} (t_i - \bar{t})^2\right]$$

$$= \exp\left(\mu s + \frac{\sigma^2}{2n} s^2\right) \exp\left[\frac{\sigma^2}{2} \sum_{i=1}^{n} (t_i - \bar{t})^2\right]$$

The first factor is the mgf of $\overline{X}$. Since the mgf of the vector $(X_1 - \overline{X}, \ldots, X_n - \overline{X})$ can be obtained by setting $s = 0$ in M, the second factor is this mgf. ∎

COROLLARY A

$\overline{X}$ and S^2 are independently distributed.

Proof

This follows immediately since S^2 is a function of the vector $(X_1 - \overline{X}, \ldots, X_n - \overline{X})$, which is independent of $\overline{X}$. ■

The next theorem gives the marginal distribution of S^2.

THEOREM B

The distribution of $(n-1)S^2/\sigma^2$ is the chi-square distribution with $n-1$ degrees of freedom.

Proof

We first note that

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \mu)^2 = \sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma}\right)^2 \sim \chi_n^2$$

Also,

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \mu)^2 = \frac{1}{\sigma^2}\sum_{i=1}^{n}[(X_i - \overline{X}) + (\overline{X} - \mu)]^2$$

Expanding the square and using the fact that $\sum_{i=1}^{n}(X_i - \overline{X}) = 0$, we obtain

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \mu)^2 = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \overline{X})^2 + \left(\frac{\overline{X} - \mu}{\sigma/\sqrt{n}}\right)^2$$

This is a relation of the form $W = U + V$. Since U and V are independent by Corollary A, $M_W(t) = M_U(t)M_V(t)$. W and V both follow chi-square distributions, so

$$\begin{aligned} M_U(t) &= \frac{M_W(t)}{M_V(t)} \\ &= \frac{(1-2t)^{-n/2}}{(1-2t)^{-1/2}} \\ &= (1-2t)^{-(n-1)/2} \end{aligned}$$

The last expression is the mgf of a random variable with a χ_{n-1}^2 distribution. ■

One final result concludes this chapter's collection.