CHAPTER 4

APPLICATIONS OF STATISTICAL INFERENCE

4.1 SUMMARY OF NECESSARY THEORETICAL RESULTS

In order to carry out many more statistical inferences other than the simple confidence intervals that we have considered, we need some additional theoretical results which we give in this section. Some of these will be proved in Chapter 6; so we state a few of them as theorems and we give an example of each.

Remark If the instructor so chooses, Section 6.2 on the moment-generating function (m.g.f.) of linear functions can be studied at this time.

Theorem 4.1-1 Let $X_1, X_2, \ldots, X_n$ be n independent chi-square random variables with $r_1, r_2, \ldots, r_n$ degrees of freedom, respectively. Then the random variable $Y = X_1 + X_2 + \cdots + X_n$ has a chi-square distribution with $r_1 + r_2 + \cdots + r_n$ degrees of freedom.

EXAMPLE 4.1-1 Let X_1, X_2, X_3 be three independent chi-square random variables with $r_1 = 2$, $r_2 = 5$, $r_3 = 4$ degrees of freedom. Then the random variable $Y = X_1 + X_2 + X_3$ is $\chi^2(11)$ and $P(Y < 19.68) = 0.95$ from Table IV.

Theorem 4.1-2 If $X_1, X_2, \ldots, X_n$ are independent normal random variables with means $\mu_1, \mu_2, \ldots, \mu_n$ and variances $\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2$, respectively, then the random variable $Y = \sum_{i=1}^n a_i X_i$ has a normal distribution with mean $\mu_Y = \sum_{i=1}^n a_i \mu_i$ and variance $\sigma_Y^2 = \sum_{i=1}^n a_i^2 \sigma_i^2$.

EXAMPLE 4.1-2 Let X_1, X_2, X_3 be three independent normal random variables with means $\mu_1 = 4$, $\mu_2 = -5$, $\mu_3 = 2$ and variances $\sigma_1^2 = 4$, $\sigma_2^2 = 16$, $\sigma_3^2 = 9$, respectively. Then the random variable $Y = X_1 - X_2$ is normal with $\mu_Y = (1)(4) + (-1)(-5) = 9$ and $\sigma_Y^2 = (1)^2(4) + (-1)^2(16) = 20$. Moreover, the random variable $Z = X_1 + X_2 + X_3$ is normal with mean $\mu_Z = 4 + (-5) + 2 = 1$ and variance $\sigma_Z^2 = 4 + 16 + 9 = 29x$.

Theorem 4.1-3 Let $\overline{X}$ and S^2 be the mean and the variance of a random sample of size n from a distribution which is $N(\mu, \sigma^2)$. Then $\overline{X}$ and S^2 are independent random variables with distributions such that

$$\overline{X} \quad \text{is} \quad N(\mu, \sigma^2/n)$$

and

$$\frac{(n-1)S^2}{\sigma^2} \quad \text{is} \quad \chi^2(n-1).$$

EXAMPLE 4.1-3 Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution which is $N(\mu, \sigma^2)$. We know from Section 3.4 that $(X_i - \mu)^2/\sigma^2$ is $\chi^2(1)$, $i = 1, 2, \ldots, n$. From Theorem 4.1-1, it must be true that $\sum_{i=1}^n (X_i - \mu)^2/\sigma^2$ is $\chi^2(n)$ as we are adding n independent chi-square variables together. From Theorem 4.1-3, note that if we replace μ by its estimator $\overline{X}$ we obtain $(n-1)S^2/\sigma^2 = \sum_{i=1}^n (X_i - \overline{X})/\sigma^2$, which is $\chi^2(n-1)$. That is, we have lost one degree of freedom by replacing a parameter in a chi-square variable by its estimator. Later, we see that this is generalized: If p parameters in a $\chi^2(r)$ random variable, $p < r$, are replaced by "reasonable" estimators, the resulting chi-square variable has $r - p$ degrees of freedom.

Before we close this section on results needed to make a number of statistical inferences, we consider two important distributions. We do not try to develop them here, but only give the definitions of these two random variables and refer the reader to Tables VI and VII in which a few probabilities concerning these T and F random variables are given. Also see Examples 5.2-12 and 5.2-13 for the p.d.f.s and some graphs for the T and F distributions.

Let

$$T = \frac{Z}{\sqrt{U/r}},$$

where Z has a $N(0,1)$ distribution, U has a $\chi^2(r)$ distribution, and Z and U are independent. We say that T has a Student's t distribution with r degrees of freedom. For illustration, if $r = 10$, $P[T \le 2.764] = 0.99$ from Table VI in the Appendix. In general, from Table VI, we can find $t_\alpha(r)$ so that $P[T > t_\alpha(r)] = \alpha$.

EXAMPLE 4.1-4 (continuation of Example 4.1-3) With Theorem 4.1-3, $\overline{X}$ has a $N(\mu, \sigma^2/n)$ distribution so that

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \quad \text{is} \quad N(0,1).$$

Of course $(n-1)S^2/\sigma^2$ is $\chi^2(n-1)$ and $\overline{X}$ and S^2 are independent. Hence

$$T = \frac{\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\dfrac{(n-1)S^2}{\sigma^2}\Big/(n-1)}} = \frac{\overline{X}-\mu}{S/\sqrt{n}}$$

has a Student's t distribution with $n-1$ degree of freedom.

Let

$$F = \frac{U_1/r_1}{U_2/r_2},$$

where U_i is $\chi^2(r_i)$, $i = 1, 2$, and U_1 and U_2 are independent. Then we say that F has a Fisher's F distribution (some call it Snedecor's F) with r_1 and r_2 degrees of freedom, denoted by $F(r_1, r_2)$. For illustration, if $r_1 = 5$ and $r_2 = 10$, then $P(F \le 4.24) = 0.975$ and $P(F > 5.64) = 0.01$ from Table VII in the Appendix. In general from that table, we can find F_α so that $P(F > F_\alpha) = \alpha$ for selected values of α.

EXAMPLE 4.1-5 Let $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_m$ be independent random samples from $N(\mu_X, \sigma_X^2)$ and $N(\mu_Y, \sigma_Y^2)$, respectively. We know, from Theorem 4.1-3, that $(n-1)S_X^2/\sigma_X^2$ is $\chi^2(n-1)$ and $(m-1)S_Y^2/\sigma_Y^2$ is $\chi^2(m-1)$ and the two are independent since the Xs and Ys are independent. Thus

$$F = \frac{\dfrac{(n-1)S_X^2}{\sigma_X^2(n-1)}}{\dfrac{(m-1)S_Y^2}{\sigma_Y^2(m-1)}} = \frac{S_X^2\sigma_Y^2}{S_Y^2\sigma_X^2}$$

has an F distribution with $n-1$ and $m-1$ degrees of freedom.

EXERCISES 4.1

4.1-1 Let X_1 and X_2 be independent chi-square variables with $r_1 = 4$ and $r_2 = 2$ degrees of freedom, respectively.

(a) Determine $P(2.204 < X_1 + X_2 < 16.81)$.

(b) Find $P(12.59 < X_1 + X_2)$.

4.1-2 Let X_1 and X_2 be independent random variables. Let $Y = X_1 + X_2$ be $\chi^2(14)$ and let X_1 be $\chi^2(3)$.

(a) Guess the distribution of X_2. Note that we prove this in Section 6.2.

(b) Determine $P(3.053 < X_2 < 24.72)$.

4.1-3 Let the independent random variables X_1 and X_2 be $N(\mu_1 = 3, \sigma_1^2 = 9)$ and $N(\mu_2 = 6, \sigma_2^2 = 16)$, respectively. Determine $P(-10 < Y < 5)$, where $Y = X_1 - X_2$.

4.1-4 Three random steps in series are needed to complete a certain procedure. The means and the standard deviations of the respective steps are $\mu_1 = 6$ hours, $\mu_2 = 4$ hours, $\mu_3 = 5$ hours and $\sigma_1 = 2$ hours, $\sigma_2 = 2$ hours, $\sigma_3 = 3$ hours. Assuming independence and normal distributions, compute the probability that the procedure will be completed in less than 20 hours.

4.2 CONFIDENCE INTERVALS USING χ^2, F, AND T

In this section, we assume that all samples arise from normal distributions. However, in actual practice, we must always question the assumption of normality and make certain that those underlying distributions are at least approximately normal before using this theory.

We use the fact (Theorem 4.1-3) that $W = (n-1)S^2/\sigma^2$ is $\chi^2(n-1)$ to find a confidence interval for σ^2. From Table IV in the Appendix with $n-1$ degrees of freedom select a and b such that

$$P\left(a \le \frac{(n-1)S^2}{\sigma^2} \le b\right) = 1 - \alpha,$$

where $P(W \le a) = P(W \ge b) = \alpha/2$ so that $a = \chi^2_{1-\alpha/2}(n-1)$ and $b = \chi^2_{\alpha/2}(n-1)$. (Note that $\chi^2_p(r)$ is a number that cuts off p probability to the *right* of it for a $\chi^2(r)$ random variable.) Then, solving the inequalities, we have

$$1 - \alpha = P\left(\frac{a}{(n-1)S^2} \le \frac{1}{\sigma^2} \le \frac{b}{(n-1)S^2}\right)$$

$$= P\left(\frac{(n-1)S^2}{b} \le \sigma^2 \le \frac{(n-1)S^2}{a}\right).$$

Thus the probability that the random interval $[(n-1)S^2/b, (n-1)S^2/a]$ contains the unknown σ^2 is $1-\alpha$. Once the values of $X_1, X_2, \ldots, X_n$ are observed to be $x_1, x_2, \ldots, x_n$ and s^2 computed, then the interval $[(n-1)s^2/b, (n-1)s^2/a]$ is a $100(1-\alpha)\%$ confidence interval for σ^2. It follows that a $100(1-\alpha)\%$ confidence interval for σ, the standard deviation, is given by

$$\left[\sqrt{\frac{(n-1)s^2}{b}}, \sqrt{\frac{(n-1)s^2}{a}}\right] = \left[\sqrt{\frac{n-1}{b}}\,s, \sqrt{\frac{n-1}{a}}\,s\right]$$

$$= \left[\sqrt{\frac{n-1}{\chi^2_{\alpha/2}(n-1)}}\,s, \sqrt{\frac{n-1}{\chi^2_{1-\alpha/2}(n-1)}}\,s\right].$$

EXAMPLE 4.2-1

Assume that the time in days required for maturation of seeds of a species of Guardiola, a flowering plant found in Mexico, is $N(\mu, \sigma^2)$. A random sample of $n = 13$ seeds, both parents having narrow leaves, yielded $\bar{x} = 18.97$ days and

$$12s^2 = \sum_{i=1}^{13}(x_i - \bar{x})^2 = 128.41.$$

A 90% confidence interval for σ^2 is

$$\left[\frac{128.41}{21.03}, \frac{128.41}{5.226}\right] = [6.11, 24.57]$$