LECTURE NOTE 7

NORMAL LINEAR MODEL

1. NORMAL DISTRIBUTION

We shall use $\mathcal{N}(0,1)$ to denote the standard normal distribution. It is a continuous distribution with a density function

$$f(w) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}w^2).$$

Probabilities are obtained by integrating the density: if the random variable W has a $\mathcal{N}(0,1)$ distribution, then

$$\text{Prob}(c \le W \le d) = \int_c^d f(w)\, dw.$$

The expectation of W is zero and the variance is one: $E(W) = 0$, $\text{Var}(W) = 1$.

A general normal distribution, with mean μ and variance σ^2, is obtained from a linear function of a standard normal: if $W \sim \mathcal{N}(0,1)$, then

$$\mu + \sigma W \ \sim \mathcal{N}(\mu, \sigma^2).$$

2. NORMAL LINEAR MODEL

We continue to assume random sampling:

$$(Y_i, Z_i) \overset{\text{i.i.d.}}{\sim} F \qquad (i = 1, \ldots, n).$$

In this note, Y_i is 1×1 ($M = 1$). We make the two assumptions in Claim 1 of Note 6:

$$E(Y_i \,|\, Z_i = z_i) = x_i'\beta \quad \text{and} \quad \text{Var}(Y_i \,|\, Z_i = z_i) = \sigma^2.$$

In addition, we assume that the conditional distribution of Y_i given $Z_i = z_i$ is normal:

$$Y_i \mid Z_i = z_i \sim \mathcal{N}(x_i'\beta, \sigma^2) \qquad (i = 1, \dots, n).$$

Define the prediction error

$$U_i \equiv Y_i - x_i'\beta,$$

and let

$$V_i = U_i/\sigma,$$

so that

$$V_i \mid Z_i = z_i \sim \mathcal{N}(0, 1).$$

Because of the random sampling, this implies that

$$V_i \mid Z = z \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1) \qquad (i = 1, \dots, n).$$

If W is $n \times 1$ with a joint normal distribution, with mean μ and covariance matrix Σ, we shall use the notation $W \sim \mathcal{N}(\mu, \Sigma)$. So we can write the normal linear model as

$$Y = x\beta + \sigma V, \quad V \mid Z = z \sim \mathcal{N}(0, I_n), \tag{1}$$

with

$$Y = \begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix}, \quad x = \begin{pmatrix} x_1' \\ \vdots \\ x_n' \end{pmatrix}, \quad V = \begin{pmatrix} V_1 \\ \vdots \\ V_n \end{pmatrix}$$

(and $x = g(z)$.) We are going to transform this model into the following canonical form:

$$Y^* = \begin{pmatrix} s \\ 0 \end{pmatrix} \beta + \sigma V^*, \qquad V^* \mid Z = z \sim \mathcal{N}(0, I_n). \tag{2}$$

This will make it simpler to obtain the distribution of the sum of squared residuals (from a least-squares fit), and to show that the least-squares estimator is independent of the sum of squared residuals.

3. CANONICAL FORM

A $n \times n$ matrix q is *orthogonal* if $q^{-1} = q'$. Note that if q is orthogonal then so is q', and $qq' = q'q = I_n$. Orthogonal matrices preserve the least-squares inner product:

Claim 1. If c and d are $n \times 1$ and q is an orthogonal $n \times n$ matrix, then

$$\langle qc, qd \rangle = \langle c, d \rangle.$$

Proof.

$$\langle qc, qd \rangle = (qc)'(qd) = c'q'qd = c'd = \langle c, d \rangle. \quad \diamond$$

So an orthogonal matrix preserves the least-squares norm:

$$||qc||^2 = \langle qc, qc \rangle = \langle c, c \rangle = ||c||^2.$$

Claim 2 (QR decomposition). If x is $n \times K$, then there is an orthogonal $n \times n$ matrix q and an upper triangular $n \times K$ matrix r such that

$$x = qr.$$

If $n > K$, then

$$r = \begin{pmatrix} s \\ 0 \end{pmatrix}, \quad s = \begin{pmatrix} s_{11} & s_{12} & \dots & s_{1K} \\ 0 & s_{22} & \dots & s_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & s_{KK} \end{pmatrix}.$$

See G. Golub and C. Van Loan, *Matrix Computations*, Third Edition, The Johns Hopkins University Press, 1996.

We shall use the following property of joint normality:

Claim 3. If d_1 is $m \times n$, d_2 is $m \times 1$, with d_1 and d_2 not random, and if W is $n \times 1$ with $W \sim \mathcal{N}(\mu, \Sigma)$, then

$$d_1 W + d_2 \sim \mathcal{N}(d_1\mu + d_2, d_1 \Sigma d_1').$$

Claim 4. If $V \sim \mathcal{N}(0, I_n)$ and q is an $n \times n$ orthogonal (nonrandom) matrix, then

$$qV \sim \mathcal{N}(0, I_n).$$

Proof.

$$qV \sim \mathcal{N}(0, qI_nq') = \mathcal{N}(0, I_n).$$

Now we can transform the normal linear model in (1) into the canonical form in (2). We shall assume that $n > K$ and that there is no exact linear dependence connecting the columns of x: if c is $K \times 1$ and $xc = 0$, then $c = 0$. This implies that $x'x$ is nonsingular and so s in the QR decomposition of x is nonsingular. From (1),

$$Y = q \begin{pmatrix} s \\ 0 \end{pmatrix} \beta + \sigma V.$$

Multiply both sides of this equation by q':

$$q'Y = \begin{pmatrix} s \\ 0 \end{pmatrix} \beta + \sigma q'V.$$

Define

$$Y^* = q'Y, \quad V^* = q'V,$$

and use Claim 4 to obtain our canonical form:

$$Y^* = \begin{pmatrix} s \\ 0 \end{pmatrix} \beta + \sigma V^*, \qquad V^* \,|\, Z = z \sim \mathcal{N}(0, I_n).$$

Define

$$Y^*_{(1)} = \begin{pmatrix} Y_1 \\ \vdots \\ Y_K \end{pmatrix}, \quad V^*_{(1)} = \begin{pmatrix} V_1 \\ \vdots \\ V_K \end{pmatrix},$$

and

$$Y^*_{(2)} = \begin{pmatrix} Y_{K+1} \\ \vdots \\ Y_n \end{pmatrix}, \quad V^*_{(2)} = \begin{pmatrix} V_{K+1} \\ \vdots \\ V_n \end{pmatrix}.$$

Then we can write the canonical form as

$$Y^*_{(1)} = s\beta + \sigma V^*_{(1)}, \tag{3}$$

$$Y^*_{(2)} = \sigma V^*_{(2)}, \tag{4}$$

where s is a $K \times K$ nonsingular matrix and, conditional on $Z = z$, the components of $V^*_{(1)}$ and $V^*_{(2)}$ are all i.i.d. $\mathcal{N}(0, 1)$.

The least squares estimator solves

$$b = \arg\min_d ||Y - xd||^2.$$

From Claim 1, and using the QR decomposition of x,

$$\begin{aligned} ||Y - xd||^2 = ||q'(Y - xd)||^2 &= ||Y^* - \begin{pmatrix} s \\ 0 \end{pmatrix} d||^2 \\ &= ||Y^*_{(1)} - sd||^2 + ||Y^*_{(2)}||^2. \end{aligned}$$

So we have

$$b = s^{-1}Y^*_{(1)} = \beta + \sigma s^{-1} V^*_{(1)}, \tag{5}$$

and the sum of squared residuals is

$$\begin{aligned} \text{SSR} \equiv \min_d ||Y - xd||^2 &= ||Y - xb||^2 \\ &= ||Y_{(2)}||^2 = \sigma^2 ||V_{(2)}||^2 \\ &= \sigma^2 \sum_{i=K+1}^{n} (V_i^*)^2. \end{aligned} \tag{6}$$

Since the least-squares estimator depends on V_i^* only for $i = 1, \ldots, K$, and since SSR depends on V_i^* only for $i = K + 1, \ldots, n$, we have b independent of SSR conditional on $Z = z$.

From (5), the distribution of the least-squares estimator is joint normal:

$$b \,|\, Z = z \sim \mathcal{N}(\beta, \sigma^2 s^{-1} s^{-1\prime}) = \mathcal{N}(\beta, \sigma^2 (x'x)^{-1}).$$

Definition 1. If V_i i.i.d $\mathcal{N}(0,1)$, then

$$\sum_{i=1}^{m} V_i^2 \sim \text{Chi}^2(m).$$

We shall also use the notation $\chi^2(m)$. The parameter m is called the degrees of freedom. Note that the mean of a random variable with a chi-square distribution is

$$E[\text{Chi}^2(m)] = E(\sum_{i=1}^{m} V_i^2) = m.$$

From (6), the sum of squared residuals is distributed as σ^2 times a random variable with a chi-square distribution, with $n-K$ degrees of freedom:

$$\text{SSR} \,|\, Z = z \sim \sigma^2 \text{Chi}^2(n-K).$$

We can obtain an unbiased estimator for σ^2 by dividing SSR by the degrees of freedom:

$$\hat{\sigma}^2 \equiv \frac{\text{SSR}}{n-K}, \tag{7}$$

$$E(\hat{\sigma}^2 \,|\, Z = z) = \sigma^2.$$

4. CONFIDENCE INTERVAL

We shall obtain a confidence interval for a linear combination of the coefficients:

$$l'\beta = \sum_{j=1}^{K} l_j \beta_j.$$

Define the *standard error*

$$\text{SE} = [\hat{\sigma}^2 l'(x'x)^{-1} l]^{1/2}.$$

Our confidence interval will be based on the t-distribution.

Definition 2. If the random variables W and S are independent, with $W \sim \mathcal{N}(0,1)$, $S \sim \text{Chi}^2(m)$, then

$$\frac{W}{(S/m)^{1/2}} \sim t(m).$$

Claim 5.

$$\frac{l'(b-\beta)}{\mathrm{SE}} \,|\, Z = z \sim t(n-K).$$

Proof. Conditional on $Z = z$,

$$\frac{l'(b-\beta)}{[\sigma^2 l'(x'x)^{-1}l]^{1/2}} \Big/ \left[\frac{\mathrm{SSR}}{\sigma^2(n-K)}\right]^{1/2} \sim \mathcal{N}(0,1) \Big/ \left[\frac{\mathrm{Chi}^2(n-K)}{n-K}\right]^{1/2}$$
$$\sim t(n-K),$$

where we have used the independence of b and SSR. We can cancel σ^2 in the numerator and denominator on the left-hand side, and simplify to obtain

$$\frac{l'(b-\beta)}{\mathrm{SE}} \sim t(n-K). \quad \diamond$$

Note that because the conditional distribution of $l'(b-\beta)/\mathrm{SE}$ given $Z = z$ does not depend on z, we have $l'(b-\beta)/\mathrm{SE}$ independent of Z. The ratio $l'(b-\beta)/\mathrm{SE}$ is called a *pivot* for $l'\beta$. It depends on the unknown parameters only through $l'\beta$, and it has a known distribution. This leads to a confidence interval for $l'\beta$.

The $t-$distribution is available in tables and in computer programs. Suppose that $n - K = 30$. We have

$$\mathrm{Prob}(t(30) > 2.04) = .025,$$

and since the t-distribution is symmetric about zero,

$$\mathrm{Prob}(|t(30)| \leq 2.04) = .95.$$

Then Claim 5 gives

$$\mathrm{Prob}(-2.04 \leq \frac{l'\beta - l'b}{\mathrm{SE}} \leq 2.04 \,|\, Z = z) = .95,$$

and so

$$\mathrm{Prob}(l'b - 2.04 \cdot \mathrm{SE} \leq l'\beta \leq l'b + 2.04 \cdot \mathrm{SE} \,|\, Z = z) = .95. \tag{8}$$

Because the conditional probability in (8) does not depend on z, the unconditional probability is also equal to .95. We can write this as

$$\text{Prob}(\beta \in [l'b \pm 2.04 \cdot \text{SE}]) = .95.$$

As $n - K$ increases from 30 to infinity, the 97.5 percentile of the t-distribution decreases from 2.04 to a limiting value of 1.96 (which is the 97.5 percentile of a standard normal distribution).

5. CONFIDENCE ELLIPSE

We shall obtain a confidence region for two or more linear combinations of the coefficients. Let L be $h \times K$ so that $L\beta$ is $h \times 1$ and

$$Lb \,|\, Z = z \;\sim\; \mathcal{N}(L\beta, \sigma^2 L(x'x)^{-1}L').$$

Define

$$\hat{\text{Var}}(Lb) \equiv \hat{\sigma}^2 L(x'x)^{-1}L'.$$

Claim 6. If $W \sim \mathcal{N}(\mu, \Sigma)$ is $h \times 1$ and Σ is positive definite, then

$$(W - \mu)'\Sigma^{-1}(W - \mu) \;\sim\; \text{Chi}^2(h).$$

Proof. Since the $h \times h$ matrix Σ is positive definite and symmetric, there is a $h \times h$ matrix $\Sigma^{1/2}$ that is positive definite and symmetric such that

$$\Sigma = \Sigma^{1/2}\Sigma^{1/2}.$$

Then

$$\Sigma^{-1} = \Sigma^{-1/2}\Sigma^{-1/2}$$

with $\Sigma^{-1/2} = (\Sigma^{1/2})^{-1}$.

$$Q \equiv \Sigma^{-1/2}(W - \mu) \;\sim\; \mathcal{N}(0, \Sigma^{-1/2}\Sigma\Sigma^{-1/2}) = \mathcal{N}(0, I_h)$$

implies that

$$(W-\mu)'\Sigma^{-1}(W-\mu) = Q'Q = \sum_{j=1}^{h} Q_j^2 \sim \text{Chi}^2(h). \quad \diamond$$

Our confidence region for $L\beta$ will be based on the F-distribution.

Definition 3. If the random variables S_1 and S_2 are independent, with $S_1 \sim \text{Chi}^2(h)$ and $S_2 \sim \text{Chi}^2(m)$, then

$$\frac{S_1/h}{S_2/m} \sim F(h,m).$$

The parameters h and m of the F-distribution are called the numerator and denominator degrees of freedom.

Claim 7. Conditional on $Z=z$,

$$(Lb-L\beta)'[\hat{\text{Var}}(Lb)]^{-1}(Lb-L\beta)/h \sim F(h, n-K).$$

Proof. Conditional on $Z=z$,

$$\frac{(Lb-L\beta)'[\sigma^2 L(x'x)^{-1}L']^{-1}(Lb-L\beta)/h}{\text{SSR}/[\sigma^2(n-K)]} \sim \frac{\text{Chi}^2(h)/h}{\text{Chi}^2(n-K)/(n-K)} \sim F(h,n-K),$$

where we have used the independence of b and SSR. We can cancel σ^2 in the numerator and denominator on the left-hand side, and then simplify to obtain the result. $\diamond$

Note that because the conditional distribution of

$$(Lb-L\beta)'[\hat{\text{Var}}(Lb)]^{-1}(Lb-L\beta)/h \tag{9}$$

given $Z=z$ does not depend on z, the expression in (9) is independent of Z. The expression in (9) is a pivot for $L\beta$. It depends on the unknown parameters (β,σ^2) only through $L\beta$, and it has a known distribution. This leads to a confidence region for $L\beta$.

For example, suppose that $h=2$ with

$$L = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \end{pmatrix} \quad \text{and} \quad L\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}.$$

Suppose that $n-K=30$. The F-distribution is available in tables and computer programs. We have

$$\text{Prob}(F(2,30)>3.32)=.05.$$

Then Claim 7 gives

$$\text{Prob}([\begin{pmatrix}\beta_1\\ \beta_2\end{pmatrix}-\begin{pmatrix}b_1\\ b_2\end{pmatrix}]'[\hat{\text{Var}}\begin{pmatrix}b_1\\ b_2\end{pmatrix}]^{-1}[\begin{pmatrix}\beta_1\\ \beta_2\end{pmatrix}-\begin{pmatrix}b_1\\ b_2\end{pmatrix}]/2\leq 3.32)=.95. \tag{10}$$

(The probability conditional on $Z=z$ is also .95.) The confidence region consists of the values for (β_1,β_2) that satisfy the inequality in (10). This gives the interior of an ellipse which is centered at the least-squares values (b_1,b_2).