Lecture 8.

The multivariate normal distribution with mean \( \mathbf{a} = (a_1, \ldots, a_d) \) and covariance \( C = \{C_{i,j}\} \), a symmetric positive definite matrix, is given by the density
\[
\left( \frac{1}{\sqrt{2\pi}} \right)^d \frac{1}{|C|^{\frac{1}{2}}} \exp\left[ -\frac{1}{2} < \mathbf{x} - \mathbf{a}, C^{-1} \mathbf{x} - \mathbf{a} > \right] d\mathbf{x}
\]
where \( \mathbf{x} = (x_1, \ldots, x_d) \). By completing the square, it is easy to check that
\[
\int_{\mathbb{R}^d} \left( \frac{1}{\sqrt{2\pi}} \right)^d \frac{1}{|C|^{\frac{1}{2}}} \exp\left[ -\frac{1}{2} < \theta, \mathbf{x} > -\frac{1}{2} < \mathbf{x} - \mathbf{a}, C^{-1} \mathbf{x} - \mathbf{a} > \right] d\mathbf{x} = \exp\left[ -\frac{1}{2} < \theta, C\theta > \right]
\]
When \( C \) is degenerate the normal distribution does not have a density, but lives on suitable hyperplane, where it has a density with respect to the Lebesgue measure on that hyperplane. If \( \mathbf{y} = T\mathbf{x} + \mathbf{c} \) then
\[
E[\mathbf{y}] = \mathbf{b} = Ta + \mathbf{c}
\]
and
\[
E[(\mathbf{y} - \mathbf{b})(\mathbf{y} - \mathbf{b})'] = TCT'
\]
and the distribution is again normal. The sum of two independent normal random variable is again normal with both the mean and the variance adding up. Of special interest is the two dimensional situation. If the means are 0,
\[
f(x, y) = \frac{1}{2\pi \sigma_1 \sigma_2 \sqrt{1 - \rho^2}} \exp\left[ -\frac{1}{2(1 - \rho^2)} \left( \frac{x^2}{\sigma_1^2} - \frac{2\rho xy}{\sigma_1 \sigma_2} + \frac{y^2}{\sigma_2^2} \right) \right]
\]
\( \rho \) is the correlation coefficient given by \( \rho = \frac{E[xy]}{\sigma_1 \sigma_2} \). \(-1 \leq \rho \leq 1 \).
\[
y|x \simeq N\left[ \frac{\rho \sigma_2 x}{\sigma_1}, \sigma_2^2 (1 - \rho^2) \right]
\]

Regression. Fitting a straight line for the data. Given \( N \) points \( \{(x_i, y_i)\} \) minimize
\[
\sum_{i=1}^{N} (y_i - \alpha - \beta x_i)^2
\]
The minimizers are

\[
\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i \quad \bar{y} = \frac{1}{N} \sum_{i=1}^{N} x_i
\]

\[
s_{x}^{2} = \text{var} (x) = \frac{1}{N} \sum_{i=1}^{N} x_{i}^{2} - \bar{x}^{2} \quad s_{y}^{2} = \text{var} (y) = \frac{1}{N} \sum_{i=1}^{N} y_{i}^{2} - \bar{y}^{2}
\]

\[
\text{cov} (x, y) = \frac{1}{N} \sum_{i=1}^{N} (x_{i} - \bar{x})(y_{i} - \bar{y}) \]

\[
= \frac{1}{N} \sum_{i=1}^{n} x_{i} y_{i} - \bar{x} \bar{y}
\]

\[
= \bar{x} \bar{y} - \bar{x} \bar{y}
\]

\[
\beta = \frac{\text{cov} (xy)}{\text{var} (x)}
\]

\[
\alpha = \bar{y} - \beta \bar{x}
\]

In terms of the correlation coefficient

\[
r = \frac{\text{cov} (x, y)}{s_{x} s_{y}}
\]

\beta can be expressed as

\[
\beta = \frac{r s_{y}}{r_{x}}
\]

Testing that the correlation is 0 in a bivariate (normal) data.

\[
r = \frac{\text{cov} (x, y)}{\sqrt{\text{var} (x) \text{var} (y)}}
\]

If \(x\) and \(y\) are independent what is the distribution of \(r\)?

There is a theme that comes up often in linear modeling. Let \(x_{1}, x_{2}, \ldots, x_{n}\) be \(n\) in dependent Gaussian random variables with mean 0 and variance 1. Let

\[
\{0\} = V_{0} \subset V_{1} \subset V_{2} \cdots \subset V_{k} \subset V_{k+1} = R^{n}
\]
be an increasing family of subspaces of $R^n$. Then $R_n = \bigoplus_{j=0}^{k} W_j$ where for $j = 0, 1, \ldots, k$

$$W_j = V_j \cap V_{j-1}$$

Then the quadratic form

$$\|x\|^2 = Q(x) = \sum_{i=1}^{n} x_i^2$$

can be written as

$$Q(x) = \sum_{i=0}^{k} Q_i(x) = \sum_{i=0}^{k} \|P_i x\|^2$$

where $\{P_i\}$ are orthogonal projections on to the subspaces $\{W_i\}$. Then $\{Q_i(x)\}$ are mutually independent and for every $i$, $Q_i(x)$ is distributed as a $\chi^2$ with $d_{i+1} - d_i$ degrees of freedom, where $d_i = \text{dim } V_i$.

We note that $r$ can be represented as

$$\frac{Z}{\sqrt{Z^2 + \chi^2_{n-2}}}$$

$$\sum_i (x_i - \bar{x})(y_i - \bar{y}) \sqrt{\text{var } (y)} = Z = \sum b_i x_i$$

with $\sum_i b_i = 0$ and $\sum_i b_i^2 = 1$. Treating $\{y_i\}$ as just constants, not all of them equal so that $s_y > 0$, we have three mutually orthogonal subspaces in $R^n$

$$W_1 = \{c(1, 1, \ldots, 1)\}$$
$$W_2 = \{c(y_1 - \bar{y}, y_2 - \bar{y}, \ldots, y_n - \bar{y})\}$$
$$W_3 = (W_1 \cup W_2)^\perp$$

of dimensions 1, 1, $n - 2$ and projections $P_1, P_2, P_3$ respectively.

$$\sum_{i=1}^{n} x_i^2 = [\sqrt{n\bar{x}}]^2 + ns^2$$

$$= [\sum_{i=1}^{n} a_i x_i]^2 + [\sum_{i=1}^{n} b_i x_i]^2 + Q(x)$$

$$= Z_1^2 + Z_2^2 + Q(x)$$
with $a_i \equiv \frac{1}{\sqrt{n}}$, $b_i = \frac{y_i - \bar{y}}{s_y}$. $Z_1$ and $Z_2$ are normal with mean 0 and some variance $\sigma^2$ and $Q(x) = ns_x^2 - Z_2^2$ is a $\sigma^2\chi^2_{n-2}$ and they are independent. Finally

$$r = \frac{Z_2}{\sqrt{ns_x^2}}$$

$$\frac{r}{\sqrt{1 - r^2}}\sqrt{n - 2} = \frac{Z_2}{\sqrt{Q(x)}}\sqrt{n - 2} \approx t_{n-2}$$
General linear regression models. Suppose we have a variable $Y$ that we want to predict using $\{X_j\}, \ j = 1, 2, \ldots, k$. We make a model of the form

$$Y = a_0 + a_1X_1 + \cdots + a_kX_k + Z$$

We can dispense with $a_0$ by adding $X_0$ which is always 1. We have data $\{y_i; x_{i,1}, x_{i,2}, \ldots, x_{i,k}\}$ for $i = 1, 2, \ldots, N$

$$y_i = \sum_{j=1}^{k} a_j x_{i,j} + z_j$$

$z_j$ are assumed to be independent normal random variables with mean 0 and some variance $\sigma^2$. So the general linear model in matrix notation is

$$y = Xa + z$$

$\{y_i\}$ and $\{x_{i,j}\}$ are given. We need to estimate $\{a_j\}$ and $\sigma^2$. We do least square approximation.

$$\sum_{i=1}^{N} [y_i - \sum_{j=1}^{k} a_j x_{i,j}]^2$$

is to be minimized. Or minimize

$$||y - Xa||^2$$

over choices of $a$. Leads to

$$\langle y - Xa, Xc \rangle = 0$$

for all $c$.

$$X^*Xa = X^*y$$

$$\hat{a} = [X^*X]^{-1}X^*y$$

The residual sum of squares is

$$||y - X\hat{a}||^2 = \langle y, y \rangle - 2\langle y, X\hat{a} \rangle + \langle X\hat{a}, X\hat{a} \rangle$$

$$\langle y, X\hat{a} \rangle = \langle X^*y, [X^*X]^{-1}X^*y \rangle$$
\langle \hat{X}a, X\hat{a} \rangle = \langle X[X^*X]^{-1}X^*y, X[X^*X]^{-1}X^*y \rangle = \langle X^*y, [X^*X]^{-1}X^*y \rangle

\|y - X\hat{a}\|^2 = \langle y, y \rangle - \langle X^*y, [X^*X]^{-1}X^*y \rangle

\frac{1}{n - k} \|y - X\hat{a}\|^2 = \hat{\sigma}^2

F Test. In general in order to test if a particular linear model is valid, we need to compare the error we get from fitting the model to the intrinsic error or noise level \( \sigma^2 \). If \( \sigma^2 \) is large it is impossible to say anything because the data is corrupted by very high noise. If the model is correctly specified with \( k \) linearly independent parameters \( a_1, a_2, \ldots, a_k \), then we can fit the model and the residual sum of squares \( \|x - X\hat{a}\|^2 \) provides an estimate \( \hat{\sigma}^2 \) of \( \sigma^2 \). If we want to test whether \( a = (a_1, a_2, \ldots, a_k) \) is from a subspace \( S \subset \mathbb{R}^k \) of dimension \( d \) can be tested by examining

\[ R_1(x) = \inf_{a \in S} \|x - Xa\|^2 \geq \inf_{a \in \mathbb{R}^k} \|x - Xa\|^2 = R(x) \]

If

\[ R_1(x) - R(x) = Q_1(x) \]

Then \( Q_1 \) and \( R_1 \) are independent \( \chi^2 \) and

\[ \frac{Q_1(x)}{k-d} \frac{k-d}{R(x)} \frac{R(x)}{n-k} \sim F_{k-d,n-k} \]

Example 1. Testing all means are 0. We have a collection of Normal variables \( \{x_{i,j}\} \), \( 1 \leq j \leq n_i \) and for \( i = 1, 2, \ldots, k \) they have \( E[x_{i,j}] = \mu_i \) and \( \text{var}(x_{i,j}) = \sigma^2 \) for all \( i \) and \( j \). Then

\[
X_{n \times k} = \begin{pmatrix}
1 & 0 & 0 & \cdots & 0 \\
1 & 0 & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 1 & 0 & \cdots & 0 \\
0 & 1 & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & 1 \\
0 & 0 & 0 & \cdots & 1 \\
\vdots & \vdots & \vdots & \ddots & \vdots 
\end{pmatrix}
\]
\[ n = n_1 + \cdots + n_k, \quad x_i = \mu_i + z_i. \quad \mu_1 = \cdots = \mu_{n_1} = a_1, \mu_{n_1+1} = \cdots = \mu_{n_1+n_2} = a_2 \text{ etc.} \]

The rows come in blocks of sizes \( n_1, n_2, \ldots, n_k \). If we minimize over \( H \)

\[
\inf_{a_1, \ldots, a_k} \sum_{i=1}^{k} \sum_{j=1}^{n_i} (x_{i,j} - a_i)^2 = \sum_{i} n_i s_i^2
\]

where for each \( i \), \( s_i^2 \) is the variance of \( n_i \) variables \( \{x_{i,j}\} \). If all the \( a_i \) are 0 the infimum is just \( \sum_{i,j} x_{i,j}^2 = Q(x) \).

\[ Q(x) = \sum_{i} n_i s_i^2 + R(x) \]

and

\[
\frac{1}{k} \sum_{i} n_i s_i^2 = F_{1,n-k}
\]

**Example 2.** Testing all means are equal. \( \mu_j = \mu + a_j, \sum a_j = 0 \).

\[
X_{n \times k} = \begin{pmatrix}
1 & 1 & 0 & 0 & \cdots & 0 \\
1 & 1 & 0 & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
1 & 0 & 1 & 0 & \cdots & 0 \\
1 & 0 & 1 & 0 & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
1 & 0 & 0 & 0 & \cdots & 1 \\
1 & 0 & 0 & 0 & \cdots & 1 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots 
\end{pmatrix}
\]

\[ \sum_{i} n_i \bar{x}_i^2 = n \bar{x}^2 + \sum_{i} n_i (\bar{x}_i - \bar{x})^2 \]

where \( \bar{x} = \frac{1}{n} \sum_{i=1}^{k} n_i \bar{x}_i \)

\[
\frac{n \bar{x}^2}{1/n - k} \simeq F_{1,n-k}
\]

or

\[
\frac{\bar{x}}{\sqrt{\frac{1}{n} R(x)}} \sqrt{n - k} \simeq t_{n-k}
\]