University of Washington STATISTICS

# Applied Statistics and Experimental Design
# General Linear Model

Fritz Scholz

Fall Quarter 2008

# General Linear Hypothesis

We assume the data vector $\boldsymbol{Y} = (Y_1, \ldots, Y_N)'$ consists of independent $Y_i \sim \mathcal{N}(\mu_i, \sigma^2)$ random variables, for $i = 1, \ldots, N$.

We also have a data model hypothesis, namely that the mean vector $\boldsymbol{\mu}$ can be any point in a given $s$-dimensional linear subspace $\Pi_\Omega \subset R^N$, where $s < N$.

Many statistical problem can be formulated as follows:

we identify a linear subspace $\Pi_\omega$ of $\Pi_\Omega$ of dimension $s - r$, with $0 < r \leq s$, and we test the hypothesis

$$H_0 : \boldsymbol{\mu} \in \Pi_\omega \qquad \text{against the alternative} \qquad H_1 : \boldsymbol{\mu} \in \Pi_\Omega - \Pi_\omega \, .$$

To distinguish $\Pi_\omega$ from $\Pi_\Omega$ we may want to call $\Pi_\omega$ the test hypothesis.

# The Two-Sample Example

Let $Y_1,\ldots,Y_m \sim \mathcal{N}(\xi,\sigma^2)$ and $Y_{m+1},\ldots,Y_{m+n} \sim \mathcal{N}(\eta,\sigma^2)$ be independent samples.

Here $N = m+n$ and the data model below specifies or reflects two samples with common variance $\sigma^2$ and possibly different means $\xi$ and $\eta$.

$\Pi_\Omega$ is $2$-dimensional ($s = 2$) consisting of all vectors of the form

$$\boldsymbol{\mu} = (\mu_1,\ldots,\mu_N)' = \xi\underbrace{(\overbrace{1,\ldots,1}^{m\text{ 1's}},\overbrace{0,\ldots,0}^{n\text{ 0's}})'}_{\boldsymbol{a}_1} + \eta\underbrace{(\overbrace{0,\ldots,0}^{m\text{ 0's}},\overbrace{1,\ldots,1}^{n\text{ 1's}})'}_{\boldsymbol{a}_2} = \xi\boldsymbol{a}_1 + \eta\boldsymbol{a}_2$$

The orthogonal vectors $\boldsymbol{a}_1$ and $\boldsymbol{a}_2$ span $\Pi_\Omega$.

Here we want to test the hypothesis $H_0 : \xi = \eta$, i.e., $\boldsymbol{\mu} = (\mu_1,\ldots,\mu_N)' = \xi(\boldsymbol{a}_1 + \boldsymbol{a}_2)$

and $\Pi_\omega$ is the linear subspace spanned by $\mathbf{1} = \boldsymbol{a}_1 + \boldsymbol{a}_2 = (\overbrace{1,\ldots,1}^{N\text{ 1's}})'$,
i.e., $r = 1$ or $s - r = 1$.

# The $k$-Sample Example

Let $Y_{i,1},\ldots,Y_{i,n_i} \sim \mathcal{N}(\xi_i,\sigma^2),\ i=1,\ldots,k$ be independent samples.

Here $N = n_1+\ldots+n_k$ and we have used the traditional double indexing on the $Y$'s, but we could equally well have used a more awkward single index $i=1,\ldots,N$.

The data model below specifies or reflects $k$ samples with common variance $\sigma^2$ and possibly different means $\xi_1,\ldots,\xi_k$.

$\Pi_\Omega$ is $k$-dimensional ($s=k$) consisting of all vectors of the form

$$\begin{aligned}\boldsymbol{\mu} &= (\mu_{1,1}+\ldots+\mu_{k,n_k})' = \xi_1 \underbrace{(\overbrace{1,\ldots,1}^{n_1\text{ 1's}},\ \overbrace{0,\ldots,0}^{(N-n_1)\text{ 0's}})'}_{\boldsymbol{a}_1} + \ldots + \xi_k \underbrace{(\overbrace{0,\ldots,0}^{(N-n_k)\text{ 0's}},\overbrace{1,\ldots,1}^{n_k\text{ 1's}})'}_{\boldsymbol{a}_k} \\ &= \xi_1\boldsymbol{a}_1+\ldots+\xi_k\boldsymbol{a}_k \qquad \text{The orthogonal vectors } \boldsymbol{a}_1,\ldots,\boldsymbol{a}_k \text{ span } \Pi_\Omega.\end{aligned}$$

The vector $\boldsymbol{a}_i$ has $1$'s in positions $(i,1),\ldots,(i,n_i)$ and $0$'s in the remaining positions.

# $H_0 : \xi_1 = \ldots = \xi_k$

Here we want to test the hypothesis $H_0 : \xi_1 = \ldots = \xi_k$, i.e.,

$$\boldsymbol{\mu} = (\mu_{1,1} + \ldots + \mu_{k,n_k})' = \xi_1 \boldsymbol{a}_1 + \ldots + \xi_k \boldsymbol{a}_k = \xi_1(\boldsymbol{a}_1 + \ldots + \boldsymbol{a}_k)$$

and $\Pi_\omega$ is the $(s - r)$ dimensional linear subspace spanned by

$$\mathbf{1} = \boldsymbol{a}_1 + \ldots + \boldsymbol{a}_k = (\overbrace{1, \ldots, 1}^{N \text{ 1's}})',$$

i.e., $s - r = k - r = 1$, or $r = k - 1$.

# Least Squares Estimates (LSE's)

The $\Pi_\Omega$-least squares estimate $\hat{\boldsymbol{\mu}} = \hat{\boldsymbol{\mu}}(\boldsymbol{Y})$ of $\boldsymbol{\mu}$ is the value $\boldsymbol{\mu}$ which minimizes

$$|\boldsymbol{Y} - \boldsymbol{\mu}|^2 = \sum_{i=1}^{N} (Y_i - \mu_i)^2 \qquad \text{over} \qquad \boldsymbol{\mu} \in \Pi_\Omega \,.$$

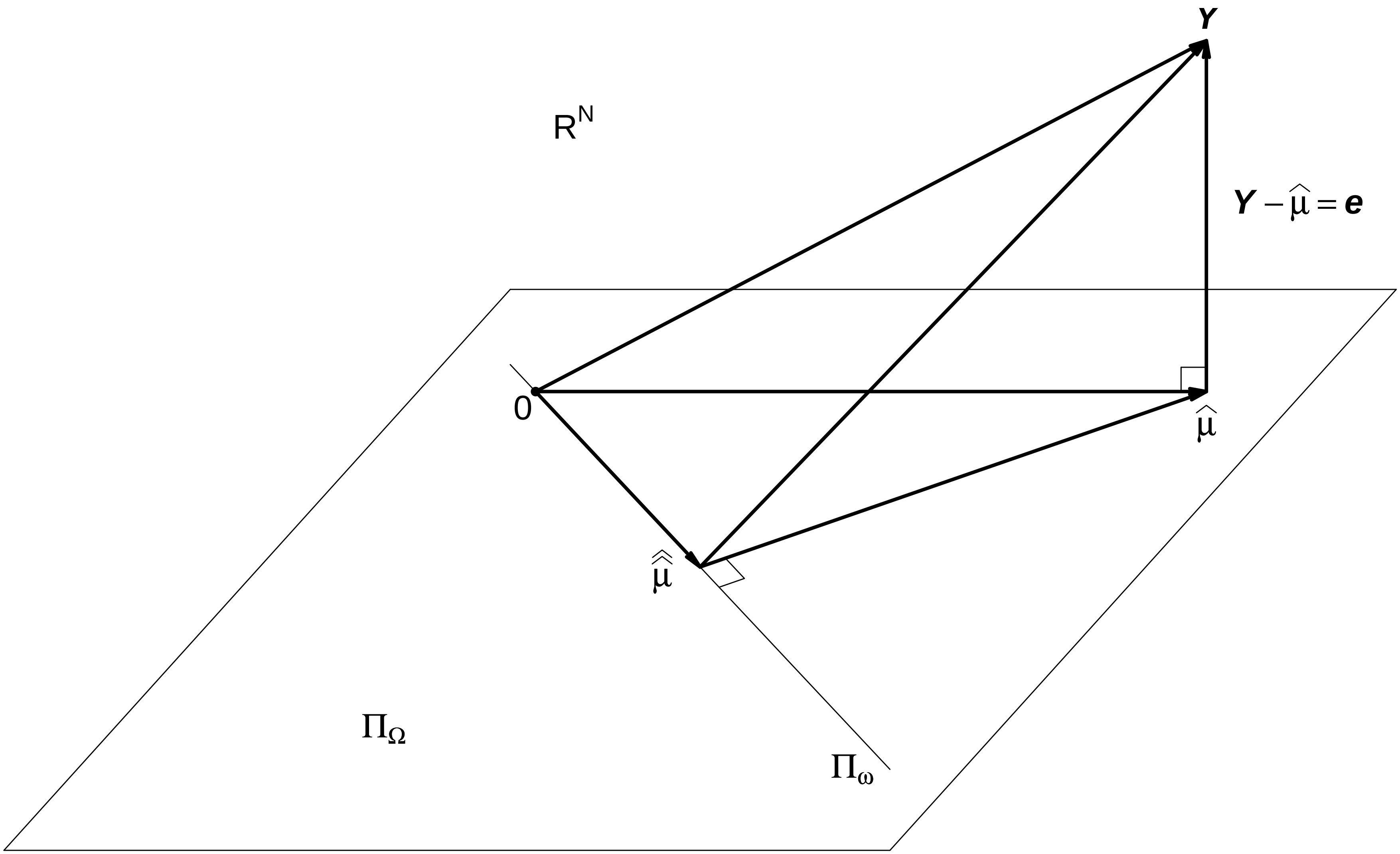
$\hat{\boldsymbol{\mu}} = \hat{\boldsymbol{\mu}}(\boldsymbol{Y})$ is the projection of $\boldsymbol{Y}$ onto $\Pi_\Omega$, or the point in $\Pi_\Omega$ closest to $\boldsymbol{Y}$.

The $\Pi_\omega$-least squares estimate $\hat{\hat{\boldsymbol{\mu}}} = \hat{\hat{\boldsymbol{\mu}}}(\boldsymbol{Y})$ of $\boldsymbol{\mu}$ is the value $\boldsymbol{\mu}$ which minimizes

$$|\boldsymbol{Y} - \boldsymbol{\mu}|^2 = \sum_{i=1}^{N} (Y_i - \mu_i)^2 \qquad \text{over} \qquad \boldsymbol{\mu} \in \Pi_\omega \,.$$

$\hat{\hat{\boldsymbol{\mu}}} = \hat{\hat{\boldsymbol{\mu}}}(\boldsymbol{Y})$ is the projection of $\boldsymbol{Y}$ onto $\Pi_\omega$, or the point in $\Pi_\omega$ closest to $\boldsymbol{Y}$.

# General Linear Model Schematic Diagram



# Comments on the Previous Diagram

$\hat{\boldsymbol{\mu}} \in \Pi_\Omega$ is the orthogonal projection of the data vector $\boldsymbol{Y}$ onto $\Pi_\Omega$,

i.e., it is the best explanation of $\boldsymbol{Y}$ in terms of $\Pi_\Omega$ (Least Squares Distance).

The orthogonal complement $\boldsymbol{e} = \boldsymbol{Y} - \hat{\boldsymbol{\mu}}$ is the residual error vector,

i.e., that part of $\boldsymbol{Y}$ not explained by $\hat{\boldsymbol{\mu}} \in \Pi_\Omega$. $\boldsymbol{Y} = \hat{\boldsymbol{\mu}} + (\boldsymbol{Y} - \hat{\boldsymbol{\mu}}) = \hat{\boldsymbol{\mu}} \overset{\perp}{+} \boldsymbol{e}$.

$\hat{\hat{\boldsymbol{\mu}}}$ is the orthogonal projection of the data vector $\boldsymbol{Y}$ onto $\Pi_\omega$,

i.e., it is the best explanation of $\boldsymbol{Y}$ in terms of $\Pi_\omega$ (Least Squares Distance).

$\hat{\hat{\boldsymbol{\mu}}}$ is also the orthogonal projection of $\hat{\boldsymbol{\mu}}$ onto $\Pi_\omega$,

i.e., it is the best explanation of $\hat{\boldsymbol{\mu}}$ in terms of $\Pi_\omega$. $\hat{\boldsymbol{\mu}} = \hat{\hat{\boldsymbol{\mu}}} \overset{\perp}{+} (\hat{\boldsymbol{\mu}} - \hat{\hat{\boldsymbol{\mu}}})$.

The orthogonal complement $\hat{\boldsymbol{\mu}} - \hat{\hat{\boldsymbol{\mu}}}$ is that part of $\hat{\boldsymbol{\mu}}$ that cannot be explained by $\Pi_\omega$.
It expresses the estimated discrepancy of the unknown $\boldsymbol{\mu}$ from $\Pi_\omega$.

# Orthogonal Decomposition

We can view $\boldsymbol{Y}$ as the orthogonal sum of the following three component vectors

$$\boldsymbol{Y} = \hat{\hat{\boldsymbol{\mu}}} \overset{\perp}{+} (\hat{\boldsymbol{\mu}} - \hat{\hat{\boldsymbol{\mu}}}) \overset{\perp}{+} (\boldsymbol{Y} - \hat{\boldsymbol{\mu}}) = \hat{\hat{\boldsymbol{\mu}}} \overset{\perp}{+} (\hat{\boldsymbol{\mu}} - \hat{\hat{\boldsymbol{\mu}}}) \overset{\perp}{+} \boldsymbol{e}$$

Orthogonality $\implies$ these three components are independent of each other.

Pythagoras $\times$ 4

$$\begin{aligned}
|\boldsymbol{Y}|^2 &= |\hat{\boldsymbol{\mu}}|^2 + |\boldsymbol{Y} - \hat{\boldsymbol{\mu}}|^2 = |\hat{\boldsymbol{\mu}}|^2 + |\boldsymbol{e}|^2 \\
|\boldsymbol{Y}|^2 &= |\hat{\hat{\boldsymbol{\mu}}}|^2 + |\boldsymbol{Y} - \hat{\hat{\boldsymbol{\mu}}}|^2 \\
|\hat{\boldsymbol{\mu}}|^2 &= |\hat{\hat{\boldsymbol{\mu}}}|^2 + |\hat{\boldsymbol{\mu}} - \hat{\hat{\boldsymbol{\mu}}}|^2 \\
|\boldsymbol{Y} - \hat{\hat{\boldsymbol{\mu}}}|^2 &= |\hat{\boldsymbol{\mu}} - \hat{\hat{\boldsymbol{\mu}}}|^2 + |\boldsymbol{Y} - \hat{\boldsymbol{\mu}}|^2 = |\hat{\boldsymbol{\mu}} - \hat{\hat{\boldsymbol{\mu}}}|^2 + |\boldsymbol{e}|^2 \\
|\boldsymbol{Y}|^2 &= |\hat{\hat{\boldsymbol{\mu}}}|^2 + \overbrace{|\hat{\boldsymbol{\mu}} - \hat{\hat{\boldsymbol{\mu}}}|^2 + |\boldsymbol{e}|^2}^{|\boldsymbol{Y} - \hat{\hat{\boldsymbol{\mu}}}|^2} \qquad \text{double Pythagoras}
\end{aligned}$$

# LSE's $=$ MLE's (Maximum Likelihood Estimates

Under the normal distribution model for the $Y_i$ the LSE's are also the maximum likelihood estimates (MLE's) of $\boldsymbol{\mu}$ w.r.t. to the respective model constraints $\boldsymbol{\mu} \in \Pi_\Omega$ and $\boldsymbol{\mu} \in \Pi_\omega$.

This follows immediately from the likelihood function for the observed $\boldsymbol{Y} = \boldsymbol{y} = (y_1, \ldots, y_N)'$

$$L(\boldsymbol{\mu}, \sigma) = f_{\boldsymbol{\mu},\sigma}(y_1, \ldots, y_N) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^N \exp\left(-\frac{\sum_{i=1}^N (y_i - \mu_i)^2}{2\sigma^2}\right)$$

which is maximized over $\Pi_\Omega$ ($\Pi_\omega$) by minimizing $\sum_{i=1}^N (y_i - \mu_i)^2$ w.r.t. $\boldsymbol{\mu} \in \Pi_\Omega(\Pi_\omega)$ and by taking $\hat{\sigma}^2 = \sum_{i=1}^N (y_i - \hat{\mu}_i)^2/N$ and $\hat{\hat{\sigma}}^2 = \sum_{i=1}^N (y_i - \hat{\hat{\mu}}_i)^2/N$, respectively.

# General Linear Model Theorem (Proof Appendix A)

Theorem: Under the general linear model assumption we have:

$$|\hat{\hat{\boldsymbol{\mu}}} - \hat{\boldsymbol{\mu}}|^2 = \sum_{i=1}^{N} (\hat{\hat{\mu}}_i - \hat{\mu}_i)^2 = |\boldsymbol{Y} - \hat{\hat{\boldsymbol{\mu}}}|^2 - |\boldsymbol{Y} - \hat{\boldsymbol{\mu}}|^2 = \sum_{i=1}^{N} (Y_i - \hat{\hat{\mu}}_i)^2 - \sum_{i=1}^{N} (Y_i - \hat{\mu}_i)^2 \sim \sigma^2 \chi^2_{r,\lambda}$$

is independent of $$SSE = |\boldsymbol{Y} - \hat{\boldsymbol{\mu}}|^2 = \sum_{i=1}^{N} (Y_i - \hat{\mu}_i)^2 \sim \sigma^2 \chi^2_{N-s}$$

and thus $$F = \frac{|\hat{\hat{\boldsymbol{\mu}}} - \hat{\boldsymbol{\mu}}|^2 / r}{|\boldsymbol{Y} - \hat{\boldsymbol{\mu}}|^2 / (N-s)} \sim F_{r,N-s,\lambda} \, ,$$

The noncentrality parameter is

$$\lambda = |\hat{\hat{\boldsymbol{\mu}}}(\boldsymbol{\mu}) - \hat{\boldsymbol{\mu}}(\boldsymbol{\mu})|^2 / \sigma^2 = |\hat{\hat{\boldsymbol{\mu}}}(\boldsymbol{\mu}) - \boldsymbol{\mu}|^2 / \sigma^2 \, ,$$

where $\hat{\hat{\boldsymbol{\mu}}}(\boldsymbol{\mu})$ is to be viewed as $\Pi_\omega$-LSE when $\boldsymbol{Y} = \boldsymbol{\mu}$ and $\hat{\boldsymbol{\mu}}(\boldsymbol{\mu})$ is to be viewed as $\Pi_\Omega$-LSE when $\boldsymbol{Y} = \boldsymbol{\mu}$. Of course, $\hat{\boldsymbol{\mu}}(\boldsymbol{\mu}) = \boldsymbol{\mu}$ since $\boldsymbol{\mu} \in \Pi_\Omega$.

# General Linear Model Theorem (2-Sample Case)

Here the $\Pi_\Omega$-LSE of $\boldsymbol{\mu}$ minimizes

$$\sum_{i=1}^{N}(Y_i-\mu_i)^2=\sum_{i=1}^{m}(Y_i-\xi)^2+\sum_{i=m+1}^{m+n}(Y_i-\eta)^2$$

$$\Rightarrow \quad \hat{\boldsymbol{\mu}}=(\hat\xi,\ldots,\hat\xi,\hat\eta,\ldots,\hat\eta)'=\hat\xi\boldsymbol{a}_1+\hat\eta\boldsymbol{a}_2 \text{ with } \hat\xi=\bar Y_1=\sum_{i=1}^{m}Y_i/m \text{ and } \hat\eta=\sum_{i=m+1}^{m+n}Y_i/n.$$

and the $\Pi_\omega$-LSE of $\boldsymbol{\mu}$ minimizes

$$\sum_{i=1}^{N}(Y_i-\mu_i)^2=\sum_{i=1}^{N}(Y_i-\mu)^2 \quad \Rightarrow \quad \hat{\hat{\boldsymbol{\mu}}}=(\hat\mu,\ldots,\hat\mu)'=\hat{\hat{\mu}}\,\mathbf{1} \quad \text{with} \quad \hat{\hat{\mu}}=\bar Y=\sum_{i=1}^{N}Y_i/N\,.$$

$$\bar Y=(m\bar Y_1+n\bar Y_2)/N \quad \Rightarrow \quad |\hat{\hat{\boldsymbol{\mu}}}-\hat{\boldsymbol{\mu}}|^2=m(\bar Y_1-\bar Y)^2+n(\bar Y_2-\bar Y)^2=\frac{mn}{N}(\bar Y_1-\bar Y_2)^2$$

$$\text{Thus} \quad F=\frac{|\hat{\hat{\boldsymbol{\mu}}}-\hat{\boldsymbol{\mu}}|^2/r}{|\boldsymbol{Y}-\hat{\boldsymbol{\mu}}|^2/(N-s)}=\frac{([\bar Y_1-\bar Y_2]/\sqrt{1/m+1/n})^2/1}{\sum_{i=1}^{m+n}(Y_i-\bar Y)^2/(m+n-2)}\sim F_{1,N-2,\lambda}$$

the square of the 2-sample $t$-statistic. $F$-test $\iff$ two-sided $t$-test.

# Noncentrality Parameter in the 2-Sample Case

To find the noncentrality parameter $\lambda$ we replace $\boldsymbol{Y}$ by $\boldsymbol{\mu} = (\xi, \ldots, \xi, \eta, \ldots, \eta)'$ in

$$|\hat{\hat{\boldsymbol{\mu}}}(\boldsymbol{Y}) - \hat{\boldsymbol{\mu}}(\boldsymbol{Y})|^2 = \frac{mn}{N}(\bar{Y}_1 - \bar{Y}_2)^2$$

$$\Longrightarrow \quad |\hat{\hat{\boldsymbol{\mu}}}(\boldsymbol{\mu}) - \hat{\boldsymbol{\mu}}(\boldsymbol{\mu})|^2 = \frac{mn}{N}(\xi - \eta)^2 = \left( \frac{(\xi - \eta)}{\sqrt{1/m + 1/n}} \right)^2$$

Thus the noncentrality parameter is

$$\lambda = \frac{\left( [\xi - \eta]/\sqrt{1/m + 1/n} \right)^2}{\sigma^2} = \delta^2$$

where $\delta$ is the noncentrality parameter in the two-sample $t$-test, namely

$$\delta = \frac{\xi - \eta}{\sigma\sqrt{1/m + 1/n}} \,.$$

# General Linear Model Theorem (k-Sample Case)

Here the $\Pi_\Omega$-LSE of $\boldsymbol{\mu}$ minimizes

$$\sum_{i=1}^{N}(Y_i-\mu_i)^2=\sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{i,j}-\xi_i)^2$$

$$\Rightarrow \quad \hat{\boldsymbol{\mu}}=(\hat{\xi}_1,\ldots,\hat{\xi}_1,\ldots,\hat{\xi}_k\ldots,\hat{\xi}_k)'=\hat{\xi}_1\,\boldsymbol{a}_1+\ldots+\hat{\xi}_k\boldsymbol{a}_k \text{ with } \hat{\xi}_i=\bar{Y}_{i\bullet}=\sum_{j=1}^{n_i}Y_{i,j}/n_i.$$

and the $\Pi_\omega$-LSE of $\boldsymbol{\mu}$ minimizes

$$\sum_{i=1}^{N}(Y_i-\mu_i)^2=\sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{i,j}-\mu)^2 \;\Rightarrow\; \hat{\hat{\boldsymbol{\mu}}}=(\hat{\mu},\ldots,\hat{\mu})'=\hat{\hat{\mu}}\,\mathbf{1}$$

with $\hat{\hat{\mu}}=\bar{Y}_{\bullet\bullet}=\sum_{i=1}^{k}\sum_{j=1}^{n_i}Y_{i,j}/N=\bar{Y}_{\bullet\bullet}=\sum_{i=1}^{k}\bar{Y}_{i\bullet}n_i/N \quad\Rightarrow\quad |\hat{\boldsymbol{\mu}}-\hat{\hat{\boldsymbol{\mu}}}|^2=\sum_{i=1}^{k}\sum_{j=1}^{n_i}(\bar{Y}_{i\bullet}-\bar{Y}_{\bullet\bullet})^2$

Thus $\quad F=\dfrac{|\hat{\hat{\boldsymbol{\mu}}}-\hat{\boldsymbol{\mu}}|^2/r}{|\boldsymbol{Y}-\hat{\boldsymbol{\mu}}|^2/(N-s)}=\dfrac{\sum_{i=1}^{k}\sum_{j=1}^{n_i}(\bar{Y}_{i\bullet}-\bar{Y}_{\bullet\bullet})^2/(k-1)}{\sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{i,j}-\bar{Y}_{\bullet\bullet})^2/(N-k)}\sim F_{k-1,N-k,\lambda}$

# Noncentrality Parameter in the k-Sample Case

To get the noncentrality parameter $\lambda$ we replace $\boldsymbol{Y}$ by $\boldsymbol{\mu} = (\xi_1, \ldots, \xi_1, \ldots, \xi_k, \ldots, \xi_k)'$ in

$$|\hat{\hat{\boldsymbol{\mu}}}(\boldsymbol{Y}) - \hat{\boldsymbol{\mu}}(\boldsymbol{Y})|^2 = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (\bar{Y}_{i\bullet} - \bar{Y}_{\bullet\bullet})^2$$

$$\Longrightarrow \quad |\hat{\hat{\boldsymbol{\mu}}}(\boldsymbol{\mu}) - \hat{\boldsymbol{\mu}}(\boldsymbol{\mu})|^2 = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (\xi_i - \bar{\xi})^2 = \sum_{i=1}^{k} n_i (\xi_i - \bar{\xi})^2$$

with $\quad \bar{\xi} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} \mu_{i,j}/N = \sum_{i=1}^{k} \xi_i n_i / N$

Thus the noncentrality parameter is

$$\lambda = \frac{\sum_{i=1}^{k} n_i (\xi_i - \bar{\xi})^2}{\sigma^2} \, .$$

# Appendix A: General Linear Model Theorem Proof

The proof is essentially that given in *Testing Statistical Hypotheses, 3rd Edition*, chapter 7, by E.L. Lehmann and J.P. Romano (2005)

There it is presented in the context of certain optimality properties and followed by many explicit examples.

The proof is first given in a case of special linear subspaces $\tilde{\Pi}_\Omega$ and $\tilde{\Pi}_\omega \subset \tilde{\Pi}_\Omega$, where the statement of the theorem is immediate from the definitions of $\chi^2_f$, $\chi^2_{g,\lambda}$ and $F_{f,g,\lambda}$.

Then it is argued that the general case can always be orthonormally transformed to the special case and that this does not change the meaning of LSE, i.e., LSE's in one framework rotate to LSE's in the other framework.
Distances and independence don't change under orthonormal tranforms.

# General Linear Model Theorem: Special Case

$U_i \sim \mathcal{N}(\nu_i, \sigma^2), \;\; i = 1, \ldots, N$ with model hypothesis $\nu_{s+1} = \ldots = \nu_N = 0$.

This describes an $s$-dimensional linear subspace $\tilde{\Pi}_\Omega$ of $R^N$.

Test $H_0 : \nu_1 = \ldots = \nu_r = 0$ (in addition to the model hypothesis).

This describes an $(s-r)$-dimensional subspace $\tilde{\Pi}_\omega \subset \tilde{\Pi}_\Omega$.

$$\Longrightarrow \quad \sum_{i=s+1}^{N} U_i^2/\sigma^2 \sim \chi^2_{N-s} \qquad \text{and} \qquad \sum_{i=1}^{r} U_i^2/\sigma^2 \sim \chi^2_{r,\lambda}$$

are independent and with noncentrality parameter $\lambda = \sum_{i=1}^{r} \nu_i^2/\sigma^2$

The natural test rejects $H_0$ when the corresponding $F$-statistic is too large, where

$$F = \frac{\sum_{i=1}^{r} U_i^2/r}{\sum_{i=s+1}^{N} U_i^2/(N-s)} \sim F_{r,N-s,\lambda} \, .$$

# LSE View in Special Case

The $\tilde{\Pi}_\Omega$-LSE $\hat{\boldsymbol{\nu}}$ minimizes

$$|\boldsymbol{U}-\boldsymbol{\nu}|^2 = \sum_{i=1}^{N}(U_i-\nu_i)^2 = \sum_{i=1}^{s}(U_i-\nu_i)^2 + \sum_{i=s+1}^{N} U_i^2 \quad \text{over} \quad \boldsymbol{\nu}\in\tilde{\Pi}_\Omega$$

i.e., $\hat{\nu}_i = U_i$ for $i=1,\ldots,s$ and $\hat{\nu}_i = 0$ for $i=s+1,\ldots,N$.

Similarly, the $\tilde{\Pi}_\omega$-LSE $\hat{\hat{\boldsymbol{\nu}}}$ minimizes

$$|\boldsymbol{U}-\boldsymbol{\nu}|^2 = \sum_{i=1}^{N}(U_i-\nu_i)^2 = \sum_{i=1}^{r}U_i^2 + \sum_{i=r+1}^{s}(U_i-\nu_i)^2 + \sum_{i=s+1}^{N} U_i^2 \quad \text{over} \quad \boldsymbol{\nu}\in\tilde{\Pi}_\omega$$

i.e., $\hat{\hat{\nu}}_i = U_i$ for $i=r+1,\ldots,s$ and $\hat{\nu}_i = 0$ for $i=1,\ldots,r,s+1,\ldots,N$.

Clearly

$$|\boldsymbol{U}-\hat{\boldsymbol{\nu}}|^2 = \sum_{i=s+1}^{N} U_i^2 \quad \text{and} \quad |\hat{\hat{\boldsymbol{\nu}}}-\hat{\boldsymbol{\nu}}|^2 = |\boldsymbol{U}-\hat{\hat{\boldsymbol{\nu}}}|^2 - |\boldsymbol{U}-\hat{\boldsymbol{\nu}}|^2 = \sum_{i=1}^{r} U_i^2$$

$$\Rightarrow \quad F = \frac{|\hat{\hat{\boldsymbol{\nu}}}-\hat{\boldsymbol{\nu}}|^2/r}{|\boldsymbol{U}-\hat{\boldsymbol{\nu}}|^2/(N-s)} = \frac{\sum_{i=1}^{r}U_i^2/r}{\sum_{i=s+1}^{N}U_i^2/(N-s)}$$

# LSE View of the Noncentrality Parameter

Using $\boldsymbol{U} = \boldsymbol{\nu} \in \tilde{\Pi}_\Omega$ in the $\tilde{\Pi}_\omega$-LSE derivation of $\boldsymbol{\nu}_\omega$ we have

$$
\begin{aligned}
|\boldsymbol{U} - \boldsymbol{\nu}_\omega|^2 &= |\boldsymbol{\nu} - \boldsymbol{\nu}_\omega|^2 = \sum_{i=1}^{N} (\nu_i - \nu_{\omega,i})^2 \\
&= \sum_{i=1}^{r} \nu_i^2 + \sum_{i=r+1}^{s} (\nu_i - \nu_{\omega,i})^2 + \sum_{i=s+1}^{N} \nu_i^2 \quad \text{since } \boldsymbol{\nu}_\omega \in \tilde{\Pi}_\omega \subset \tilde{\Pi}_\Omega \\
&= \sum_{i=1}^{r} \nu_i^2 + \sum_{i=r+1}^{s} (\nu_i - \nu_{\omega,i})^2 \quad \text{since } \boldsymbol{\nu} \in \tilde{\Pi}_\Omega \\
&= \sum_{i=1}^{r} \nu_i^2 \quad \text{after minimizing over } \tilde{\Pi}_\omega\text{, i.e., } \nu_{\omega,i} = \nu_i, i = r+1, \ldots, s.
\end{aligned}
$$

$$
\Longrightarrow \qquad \lambda = |\boldsymbol{U} - \boldsymbol{\nu}_\omega|^2/\sigma^2 = \sum_{i=1}^{r} \nu_i^2/\sigma^2 \,.
$$

# Orthonormal Transform to the Special Case

The general case can always be orthonormally transformed to the special case.

Let $C$ be an orthonormal matrix, with the first $s$ rows $\boldsymbol{c}'_i, i = 1, \ldots, s$, spanning $\Pi_\Omega$ and $\boldsymbol{c}'_i, i = r+1, \ldots, s$, spanning $\Pi_\omega$. Such orthonormal basis vectors can always be constructed via the Gram-Schmidt process.

Transform $\boldsymbol{U} = C\boldsymbol{Y}$ with mean vector $\boldsymbol{\nu} = C\boldsymbol{\mu}$

and $U_1, \ldots, U_N$ are again independent with common variance $\sigma^2$. Now note that

$$\boldsymbol{\mu} \in \Pi_\Omega \Leftrightarrow \boldsymbol{\mu} \perp \boldsymbol{c}'_i \text{ or } \nu_i = \boldsymbol{c}'_i\boldsymbol{\mu} = 0 \text{ for } i = s+1, \ldots, N \Leftrightarrow \boldsymbol{\nu} \in \tilde{\Pi}_\Omega$$

and $$\boldsymbol{\mu} \in \Pi_\omega \Leftrightarrow \boldsymbol{\mu} \perp \boldsymbol{c}'_i \text{ or } \nu_i = \boldsymbol{c}'_i\boldsymbol{\mu} = 0 \text{ for } i = 1, \ldots, r, s+1, \ldots, N \Leftrightarrow \boldsymbol{\nu} \in \tilde{\Pi}_\omega\,.$$

# Orthonormal Transform and Independence

Suppose $V_1, \ldots, V_N$ are i.i.d. $\sim \mathcal{N}(0,1)$ then $\boldsymbol{Z} = C\boldsymbol{V}$ has components $Z_1, \ldots, Z_n$ i.i.d. $\sim \mathcal{N}(0,1)$

$$f(\boldsymbol{v}) = \left(1/\sqrt{2\pi}\right)^N \exp\left(-\sum_{i=1}^{N} v_i^2/2\right) = \left(1/\sqrt{2\pi}\right)^N \exp\left(-|\boldsymbol{v}|^2/2\right)$$

has constant density if and only if $|\boldsymbol{v}|$ is constant, i.e., the constant level density contour surfaces consist of points that are equidistant from the origin.

Since orthonormal transformations $\boldsymbol{z} = C\boldsymbol{v}$ preserve distances ($|\boldsymbol{v}| = |\boldsymbol{z}|$) "it follows" that the transformed vector $\boldsymbol{Z}$ has the same density, i.e., $f(\boldsymbol{z})$.

The general independence result may be seen as follows

$$\boldsymbol{Y} = \boldsymbol{\mu} + \sigma \boldsymbol{V} \quad \text{has independent components } Y_1, \ldots, Y_N.$$

$$\Longrightarrow \quad \boldsymbol{U} = C\boldsymbol{Y} = C\boldsymbol{\mu} + \sigma C\boldsymbol{V} = \boldsymbol{\nu} + \sigma \boldsymbol{Z} \quad \text{has independent components } U_1, \ldots, U_N.$$

# Orthonormal Transform and LSE's

The principle of LSE's is based on minimizing distances.

Orthonormal transforms don't change distances. $|\boldsymbol{b}|^2 = \boldsymbol{b}'\boldsymbol{b} = \boldsymbol{a}'C'C\boldsymbol{a} = \boldsymbol{a}'\boldsymbol{a} = |\boldsymbol{a}|^2$.

Thus the LSE's (w.r.t. $\Pi_\Omega$ or $\Pi_\omega$) based on the untransformed $\boldsymbol{Y}$ simply transform to the LSE's w.r.t. to the transformed $\tilde{\Pi}_\Omega$ or $\tilde{\Pi}_\omega$ and based on the transformed $\boldsymbol{U}$.

Numerator and denominator of $F$ involve distances, which are unchanged as we pass from $\boldsymbol{Y}$ to $\boldsymbol{U} = C\boldsymbol{Y}$.

The same applies to the characterization of the noncentrality parameter which is characterized as LSE for a specific value $\boldsymbol{Y} = \boldsymbol{\mu}$ or $\boldsymbol{U} = \boldsymbol{\nu}$, respectively.