Basic Statistics and Linear Algebra

Mean as matrix multiplication: $\overline{\mathbf{X}} = \frac{1}{n}\mathbf{X}'\mathbf{1} = (\mathbf{X}_1 + \cdots + \mathbf{X}_n)/n$

Variance as matrix multiplication: $\mathbf{S} = \frac{1}{n-1}\mathbf{X}'\left(\mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}'\right)\mathbf{X} = \frac{1}{n-1}\Sigma_{j=1}^{n}(\mathbf{X}_j - \overline{\mathbf{X}})(\mathbf{X}_j - \overline{\mathbf{X}})'$

A matrix $\mathbf{A}$ is positive definite if $0 < \mathbf{x}'\mathbf{A}\mathbf{x}\ \forall \mathbf{x} \neq \mathbf{0}$

If two vectors have angle θ between them, then $\cos(\theta) = \frac{\mathbf{x}'\mathbf{y}}{\sqrt{\mathbf{x}'\mathbf{x}}\sqrt{\mathbf{y}'\mathbf{y}}}$

Transpose of a matrix: first row becomes first column (with entries in same order)

Matrix multiplication: take dot product of row with column

If $\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, then $|\mathbf{A}| = ad - bc$

If $\mathbf{A} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix}$ is 3x3, then

$$|\mathbf{A}| = a \cdot \begin{vmatrix} e & f \\ h & i \end{vmatrix} - b \cdot \begin{vmatrix} d & f \\ g & i \end{vmatrix} + c \cdot \begin{vmatrix} d & e \\ g & h \end{vmatrix}$$

The pattern of alternating + and - holds for larger matrices.

Finding eigenvalues: $|\mathbf{A} - (\lambda \cdot \mathbf{I})|$

Finding eigenvectors: $(\mathbf{A} - \lambda_i) \cdot \mathbf{e}_i = \mathbf{0}$

If $\mathbf{A} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, then $\mathbf{A}^{-1} = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$

Ch. 5 - Inferences About a Mean Vector

Hotelling's T^2: Tests for plausibility of multivariate mean vector

$T^2 = n(\overline{\mathbf{X}} - \mu_0)'(\mathbf{S})^{-1}(\overline{\mathbf{X}} - \mu_0) \sim \frac{(n-1)p}{n-p}F_{p,n-p}$

Reject $\mathrm{H}_0 : \mu = \mu_{\mathbf{0}}$ vs. $\mathrm{H}_1 : \mu \neq \mu_{\mathbf{0}}$ at the level α of significance if $T^2 > \frac{(n-1)p}{n-p}F_{p,n-p}(\alpha)$

$T^2$is invariant under changes in units of measurements for $\mathbf{X}$ of the form $\mathbf{Y} = \mathbf{C}\mathbf{X} + \mathbf{d}$, $\mathbf{C}$ nonsingular.

Confidence region for mean: $n(\overline{\mathbf{X}} - \mu)'(\mathbf{S})^{-1}(\overline{\mathbf{X}} - \mu) \leq \frac{(n-1)p}{n-p}F_{p,n-p}(\alpha)$

For 2 dimensional case, the center of the ellipse is at $\overline{\mathbf{x}}$.

Next find the eignevalues and eigenvectors for $\mathbf{S}$.

Half lengths of the major and minor axes of the ellipse are given by $\sqrt{\lambda_i}\sqrt{\frac{(n-1)p}{n-p}F_{p,n-p}(\alpha)}$ for each i

The axes lie along the eigenvectors $\mathbf{e}_i$ when they are plotted with $\overline{\mathbf{x}}$ as the origin

One-at-a-time confidence statements:

$\overline{x}_i - t_{n-1}(\alpha/2)\sqrt{\frac{s_{ii}}{n}} \leq \mu_i \leq \overline{x}_i + t_{n-1}(\alpha/2)\sqrt{\frac{s_{ii}}{n}}$

Simultaneous confidence statements:

$$\left(\mathbf{a}'\overline{\mathbf{X}} - \sqrt{\frac{(n-1)p}{n-p}F_{p,n-p}(\alpha)\mathbf{a}'\mathbf{S}\mathbf{a}}, \mathbf{a}'\overline{\mathbf{X}} + \sqrt{\frac{(n-1)p}{n-p}F_{p,n-p}(\alpha)\mathbf{a}'\mathbf{S}\mathbf{a}}\right)$$

will contain $\mathbf{a}'\mu$ with probability $1 - \alpha$.

Bonferroni method of multiple comparisons:

$\overline{x}_i - t_{n-1}\left(\frac{\alpha}{2p}\right)\sqrt{\frac{s_{ii}}{n}} \leq \mu_i \leq \overline{x}_i + t_{n-1}\left(\frac{\alpha}{2p}\right)\sqrt{\frac{s_{ii}}{n}}$

Large Sample Mean Inference: If $n - p$ large, then $\mathbf{a}'\overline{\mathbf{X}} \pm \sqrt{\chi_p^2(\alpha)}\sqrt{\frac{\mathbf{a}'\mathbf{S}\mathbf{a}}{n}}$ will contain $\mathbf{a}'\mu$ for every $\mathbf{a}$ with probability $1 - \alpha$.

EM algorithm for missing data: Let $\mathbf{T_1} = \Sigma_{j=1}^{n}\mathbf{X_j} = n\overline{\mathbf{X}}$ and $\mathbf{T_2} = \Sigma_{j=1}^{n}\mathbf{X}_j\mathbf{X}_j' = (n-1)S + n\overline{\mathbf{X}}\overline{\mathbf{X}}'$

Estimation step: Compute the revised maximum likelihood estimates $\tilde{\mu} = \frac{\widetilde{\mathbf{T}}_1}{n}$, $\tilde{\mathbf{\Sigma}} = \frac{1}{n}\widetilde{\mathbf{T}}_2 - \tilde{\mu}\tilde{\mu}'$

Prediction step: For each vector $\mathbf{x_j}$ with missing values, let $\mathbf{x_j}^{(1)}$ denote the missing components and $\mathbf{x_j}^{(2)}$ denote those components which are available. Thus, $\mathbf{x_j}' = \left[\mathbf{x_j}^{(1)'} \quad \mathbf{x_j}^{(2)'}\right]$

Now, $\tilde{\mathbf{x}}_j^{(1)} = \tilde{\mu}^{(1)} + \tilde{\mathbf{\Sigma}}_{12}\tilde{\mathbf{\Sigma}}_{22}^{-1}(\mathbf{x}_j^{(2)} - \tilde{\mu}^{(2)})$

$\widetilde{\mathbf{x}_j^{(1)}\mathbf{x}_j^{(1)'}} = \tilde{\mathbf{\Sigma}}_{11} - \tilde{\mathbf{\Sigma}}_{12}\tilde{\mathbf{\Sigma}}_{22}^{-1}\tilde{\mathbf{\Sigma}}_{21} + \tilde{\mathbf{x}}_j^{(1)}\tilde{\mathbf{x}}_j^{(1)'}$

$\widetilde{\mathbf{x}_j^{(1)}\mathbf{x}_j^{(2)'}} = \tilde{\mathbf{x}}_j^{(1)}\mathbf{x}_j^{(2)'}$

Now repeat estimation step using the predicted data.

Ch. 6 - Comparisons of Several Multivariate Means

Comparing mean vectors from two populations:

$\mathbf{S}_{pooled} = \frac{n_1-1}{n_1+n_2-2}\mathbf{S_1} + \frac{n_2-1}{n_1+n_2-2}S_2 \qquad T^2 = [\overline{\mathbf{X}}_\mathbf{1} - \overline{\mathbf{X}}_\mathbf{2} - (\mu_1 - \mu_2)]'\left[\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\mathbf{S}_{pooled}\right]^{-1}[\overline{\mathbf{X}}_\mathbf{1} - \overline{\mathbf{X}}_\mathbf{2} - (\mu_1 - \mu_2)] \sim \frac{(n_1+n_2-2)p}{(n_1+n_2-p-1)}F_{p,n_1+n_2-p-1}$

So reject $\mathrm{H}_0 : \mu_1 - \mu_2 = \delta_0$ if

$T^2 > c^2 = \frac{(n_1+n_2-2)p}{(n_1+n_2-p-1)}F_{p,n_1+n_2-p-1}$

Simultaneous confidence intervals: Using same c^2 as above, with probability $1 - \alpha$,

$$\mathbf{a}'(\overline{\mathbf{X}}_1 - \overline{\mathbf{X}}_2) \pm c\sqrt{\mathbf{a}'\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\mathbf{S}_{pooled}\mathbf{a}}$$

will cover $\mathbf{a}'(\mu_1 - \mu_2)$ for all $\mathbf{a}$

Two-sample situation when $\Sigma_1 \neq \Sigma_2$:

Let the sample sizes be such that $n_1 - p$ and $n_2 - p$ are large. Then an approximate $100(1-\alpha)\%$ confidence ellipsoid for $\mu_1 - \mu_2$ is given by all $\mu_1 - \mu_2$ satisfying

$$[\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2 - (\mu_1 - \mu_2)]'\left[\frac{1}{n_1}\mathbf{S}_1 + \frac{1}{n_2}\mathbf{S}_2\right]^{-1}[\overline{\mathbf{x}}_1 - \overline{\mathbf{x}}_2 - (\mu_1 - \mu_2)] \leq \chi_p^2(\alpha)$$

One-way MANOVA: $\mathbf{X}_{\ell j} = \mu + \tau_\ell + \mathbf{e}_{\ell j}$ with $j = 1, 2, \cdots, n_\ell$ and $\ell = 1, 2, \cdots, g$

μ is overall mean, τ_ℓ represents the ℓth treatment effect. $\mathbf{e}_{\ell j}$ are independent $N_p(\mathbf{0}, \mathbf{\Sigma})$.

A vector of observations may be decomposed as suggested by model. Thus, $\mathbf{x}_{\ell j} = \overline{\mathbf{x}} + (\overline{\mathbf{x}}_\ell - \overline{\mathbf{x}}) + (\mathbf{x}_{\ell j} - \overline{\mathbf{x}}_\ell)$ which is

observation = overall sample mean $(\hat{\mu})$ + estimated treatment effect $(\hat{\tau}_\ell)$ + residual $(\hat{\mathbf{e}}_{\ell j})$

The MANOVA table to for comparing population mean vectors is given by

Analogous to the univariate result, the hypothesis of no treatment effects,

$$H_0\text{: } \boldsymbol{\tau}_1 = \boldsymbol{\tau}_2 = \cdots = \boldsymbol{\tau}_g = \mathbf{0}$$

is tested by considering the relative sizes of the treatment and residual sums of squares and cross products. Equivalently, we may consider the relative sizes of the residual and total (corrected) sum of squares and cross products. Formally, we summarize the calculations leading to the test statistic in a MANOVA table.

MANOVA Table for Comparing Population Mean Vectors

Source of variation	Matrix of sum of squares and cross products (SSP)	Degrees of freedom (d.f.)
Treatment	$\mathbf{B} = \sum_{\ell=1}^{g} n_\ell(\overline{\mathbf{x}}_\ell - \overline{\mathbf{x}})(\overline{\mathbf{x}}_\ell - \overline{\mathbf{x}})'$	$g - 1$
Residual (Error)	$\mathbf{W} = \sum_{\ell=1}^{g}\sum_{j=1}^{n_\ell}(\mathbf{x}_{\ell j} - \overline{\mathbf{x}}_\ell)(\mathbf{x}_{\ell j} - \overline{\mathbf{x}}_\ell)'$	$\sum_{\ell=1}^{g} n_\ell - g$
Total (corrected for the mean)	$\mathbf{B} + \mathbf{W} = \sum_{\ell=1}^{g}\sum_{j=1}^{n_\ell}(\mathbf{x}_{\ell j} - \overline{\mathbf{x}})(\mathbf{x}_{\ell j} - \overline{\mathbf{x}})'$	$\sum_{\ell=1}^{g} n_\ell - 1$

One test for $\mathrm{H}_0 : \tau_1 = \tau_2 = \cdots = \tau_g = \mathbf{0}$ (ie. no treatment effect). Called *Wilks' lambda*.

The exact distribution for Wilks' lambda is given for special cases in the following table:

and related to the likelihood ratio criterion.[2] The exact distribution of Λ^* can be derived for the special cases listed in Table 6.3. For other cases and large sample sizes, a modification of Λ^* due to Bartlett (see [4]) can be used to test H_0.

Table 6.3 Distribution of Wilks' Lambda, $\Lambda^* = |\mathbf{W}|/|\mathbf{B} + \mathbf{W}|$

No. of variables	No. of groups	Sampling distribution for multivariate normal data
$p = 1$	$g \geq 2$	$\left(\frac{\Sigma n_\ell - g}{g - 1}\right)\left(\frac{1 - \Lambda^*}{\Lambda^*}\right) \sim F_{g-1, \Sigma n_\ell - g}$
$p = 2$	$g \geq 2$	$\left(\frac{\Sigma n_\ell - g - 1}{g - 1}\right)\left(\frac{1 - \sqrt{\Lambda^*}}{\sqrt{\Lambda^*}}\right) \sim F_{2(g-1), 2(\Sigma n_\ell - g - 1)}$
$p \geq 1$	$g = 2$	$\left(\frac{\Sigma n_\ell - p - 1}{p}\right)\left(\frac{1 - \Lambda^*}{\Lambda^*}\right) \sim F_{p, \Sigma n_\ell - p - 1}$
$p \geq 1$	$g = 3$	$\left(\frac{\Sigma n_\ell - p - 2}{p}\right)\left(\frac{1 - \sqrt{\Lambda^*}}{\sqrt{\Lambda^*}}\right) \sim F_{2p, 2(\Sigma n_\ell - p - 2)}$

Simultaneous confidence intervals for treatment effects: Let $n = \Sigma_{k=1}^{g} n - k$. For the MANOVA model above, with confidence at least $(1-\alpha)$, $\tau_{ki} - \tau_{\ell i}$ belongs to

$$\overline{x}_{ki} - \overline{x}_{ki} \pm t_{n-g}\left(\frac{\alpha}{pg(g-1)}\right)\sqrt{\frac{w_{ii}}{n-g}\left(\frac{1}{n_k}+\frac{1}{n_\ell}\right)}$$

where w_{ii} is the ith diagonal element of $\mathbf{W}$ from the MANOVA table

Two-way MANOVA: We specify the two-way fixed-effects model for a *vector* response consisting of p components

$\mathbf{X}_{\ell kr} = \mu + \tau_\ell + \beta_k + \gamma_{\ell k} + \mathbf{e}_{\ell kr}$ with $\ell = 1, 2, \cdots, g$, $k = 1, 2, \cdots, b$, $r = 1, 2, \cdots, n$, where,

$\Sigma_{\ell=1}^{g}\tau_\ell = \Sigma_{k=1}^{b}\beta_k = \Sigma_{\ell=1}^{g}\gamma_{\ell k} = \Sigma_{k=1}^{b}\gamma_{\ell k} = \mathbf{0}$

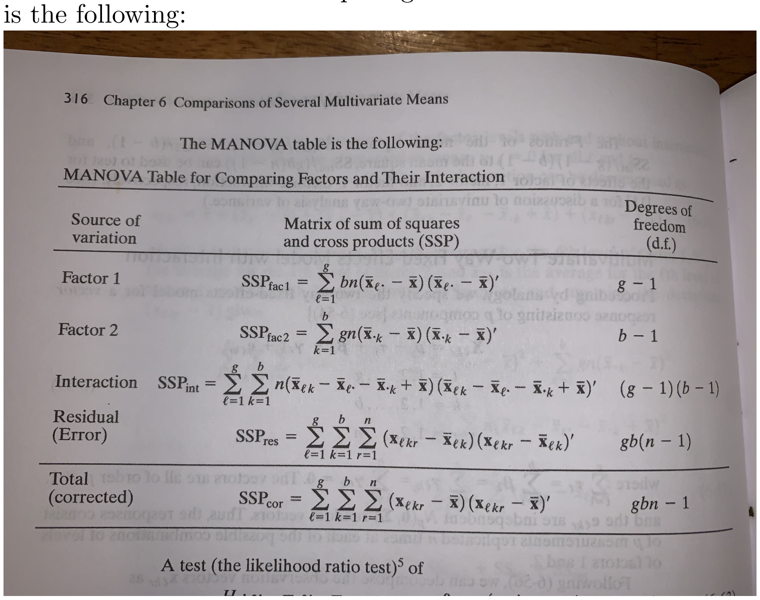
The vectors are all of order $p \times 1$ and the $\mathbf{e}_{\ell kr}$ are independent $N_p(\mathbf{0}, \mathbf{\Sigma})$

We can decompose the observation vectors $\mathbf{x}_{\ell kr}$ as

$\mathbf{x}_{\ell kr} = \overline{\mathbf{x}} + (\overline{\mathbf{x}}_{\ell\bullet} - \overline{\mathbf{x}}) + (\overline{\mathbf{x}}_{\bullet k} - \overline{\mathbf{x}}) + (\overline{\mathbf{x}}_{\ell k} - \overline{\mathbf{x}}_{\ell\bullet} - \overline{\mathbf{x}}_{\bullet k} + \overline{\mathbf{x}}) + (\mathbf{x}_{\ell kr} - \overline{\mathbf{x}}_{\ell k})$

where $\overline{\mathbf{x}}$ is the overall average of the observation vectors, $\overline{\mathbf{x}}_{\ell\bullet}$ is the average of the observation vectors at the ℓth level of factor 1, $\overline{\mathbf{x}}_{\bullet k}$ is the average of the observation vectors at the kth level of factors 2, $\overline{\mathbf{x}}_{\ell k}$ is the average of the observation vectors at the ℓth level of factor 1 *and* the kth level of factor 2.

The MANOVA table for comparing factors and their interactions is the following:



316 Chapter 6 Comparisons of Several Multivariate Means

The MANOVA table is the following:

MANOVA Table for Comparing Factors and Their Interaction

Source of variation	Matrix of sum of squares and cross products (SSP)	Degrees of freedom (d.f.)
Factor 1	$\text{SSP}_{\text{fac1}} = \sum_{\ell=1}^{g} bn(\overline{\mathbf{x}}_{\ell\cdot} - \overline{\mathbf{x}})(\overline{\mathbf{x}}_{\ell\cdot} - \overline{\mathbf{x}})'$	$g-1$
Factor 2	$\text{SSP}_{\text{fac2}} = \sum_{k=1}^{b} gn(\overline{\mathbf{x}}_{\cdot k} - \overline{\mathbf{x}})(\overline{\mathbf{x}}_{\cdot k} - \overline{\mathbf{x}})'$	$b-1$
Interaction	$\text{SSP}_{\text{int}} = \sum_{\ell=1}^{g}\sum_{k=1}^{b} n(\overline{\mathbf{x}}_{\ell k} - \overline{\mathbf{x}}_{\ell\cdot} - \overline{\mathbf{x}}_{\cdot k} + \overline{\mathbf{x}})(\overline{\mathbf{x}}_{\ell k} - \overline{\mathbf{x}}_{\ell\cdot} - \overline{\mathbf{x}}_{\cdot k} + \overline{\mathbf{x}})'$	$(g-1)(b-1)$
Residual (Error)	$\text{SSP}_{\text{res}} = \sum_{\ell=1}^{g}\sum_{k=1}^{b}\sum_{r=1}^{n} (\mathbf{x}_{\ell kr} - \overline{\mathbf{x}}_{\ell k})(\mathbf{x}_{\ell kr} - \overline{\mathbf{x}}_{\ell k})'$	$gb(n-1)$
Total (corrected)	$\text{SSP}_{\text{cor}} = \sum_{\ell=1}^{g}\sum_{k=1}^{b}\sum_{r=1}^{n} (\mathbf{x}_{\ell kr} - \overline{\mathbf{x}})(\mathbf{x}_{\ell kr} - \overline{\mathbf{x}})'$	$gbn-1$

A test (the likelihood ratio test)[5] of

A test of $\text{H}_0 : \gamma_{11} = \gamma_{12} = \cdots = \gamma_{gb} = \mathbf{0}$ (ie. no interaction effects) vs. H_1 : at least one $\gamma_{\ell k} \neq \mathbf{0}$ is conducted by rejecting H_0 for small values of the ratio $\Lambda^* = \frac{|\text{SSP}_{\text{res}}|}{|\text{SSP}_{\text{int}} + \text{SSP}_{\text{res}}|}$

Reject H_0 as above at the α level if

$$-\left[gb(n-1) - \frac{p+1-(g-1)(b-1)}{2}\right]\ln\Lambda^* > \chi^2_{(g-1)(b-1)p}(\alpha)$$

Now test for factor 1 main effects. Let $\text{H}_0 = \tau_1 = \tau_2 = \cdots = \tau_g = \mathbf{0}$ and H_1 : at least one $\tau_\ell \neq \mathbf{0}$. Let,

$\Lambda^* = \frac{|\text{SSP}_{\text{res}}|}{|\text{SSP}_{\text{fac1}} + \text{SSP}_{\text{res}}|}$

Reject H_0 at the level α if

$$-\left[gb(n-1) - \frac{p+1-(g-1)}{2}\right]\ln\Lambda^* > \chi^2_{(g-1)p}(\alpha)$$

Lastly, factor 2 effects are tested by considering $\text{H}_0 : \beta_1 = \beta_2 = \cdots = \beta_b = \mathbf{0}$ and H_1 : at least one $\beta_k \neq \mathbf{0}$. Let,

$\Lambda^* = \frac{|\text{SSP}_{\text{res}}|}{|\text{SSP}_{\text{fac2}} + \text{SSP}_{\text{res}}|}$

Reject H_0 at the level α if

$$-\left[gb(n-1) - \frac{p+1-(b-1)}{2}\right]\ln\Lambda^* > \chi^2_{(b-1)p}(\alpha)$$

The $100(1-\alpha)\%$ simultaneous confidence intervals for $\tau_{\ell i} - \tau_{mi}$ are $\tau_{\ell i} - \tau_{mi}$ belongs to $(\overline{x}_{\ell\bullet i} - \overline{x}_{m\bullet i} \pm t_v\left(\frac{\alpha}{pg(g-1)}\right)\sqrt{\frac{E_{ii}}{v}\frac{2}{bn}}$

where $v = gb(n-1)$, E_{ii} is the ith diagonal element of $\mathbf{E} = \text{SSP}_{\text{res}}$, and $\overline{x}_{\ell\bullet i} - \overline{x}_{m\bullet i}$ is the ith component of $\overline{\mathbf{x}}_{\ell\bullet} - \overline{\mathbf{x}}_{m\bullet}$.

Similarly, the $100(1-\alpha)\%$ simultaneous confidence intervals for $\beta_{ki} - \beta_{qi}$ are $\beta_{ki} - \beta_{qi}$ belongs to $(\overline{x}_{\bullet ki} - \overline{x}_{\bullet qi} \pm t_v\left(\frac{\alpha}{pb(b-1)}\right)\sqrt{\frac{E_{ii}}{v}\frac{2}{gn}}$

where v and E_{ii} are as just defined and $\overline{x}_{\bullet ki} - \overline{x}_{\bullet qi}$ is the ith component of $\overline{\mathbf{x}}_{\bullet k} - \overline{\mathbf{x}}_{\bullet q}$

Comment: We have considered the multivariate two-way model with replications. That is, the model allows for n replications of the responses at each combination of factor levels. This enables us to examine the "interaction" of the factors. If only one observation vector is available at each combination of factor levels, the two-way model does not allow for the possibility of a general interaction term $\gamma_{\ell k}$. The corresponding MANOVA table includes only factor 1, factor 2, and residual sources of variation as components of the total variation.

Ch. 7 - Multivariate Linear Regression Models

Classical linear regression model: $\mathbf{Y} = \mathbf{Z}\beta + \epsilon$ with $\mathbf{Y}$ having dimension $(n \times 1)$, $\mathbf{Z}$ $(n \times (r+1))$, and ϵ $(n \times 1)$.

In addition, $E(\epsilon) = \mathbf{0}$ [size$(n \times 1)$] and $\text{Cov}(\epsilon) = \sigma^2\mathbf{I}$ [size$(n \times n)$]

Method of least squares: selects $\mathbf{b}$ so as to minimize the sum of the squares of the differences: $S(\mathbf{b}) = \Sigma_{j=1}^{n}(y_j - b_0 - b_1 z_{j1} - \cdots - b_r z_{jr})^2 = (\mathbf{y} - \mathbf{Zb})'(\mathbf{y} - \mathbf{Zb})$

Least squares estimate of $\beta : \hat{\beta} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}$

$\hat{\mathbf{y}} = \mathbf{Z}\hat{\beta} = \mathbf{Hy}$ denotes the fitted values of $\mathbf{y}$, where $\mathbf{H} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$

The **residuals** are $\hat{\epsilon} = \mathbf{y} - \hat{\mathbf{y}} = [\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}']\mathbf{y} = (\mathbf{I} - \mathbf{H})\mathbf{y}$

and satisfy $\mathbf{Z}'\hat{\epsilon} = \mathbf{0}$ and $\hat{\mathbf{y}}'\hat{\epsilon} = 0$.

Also, the **residual sum of squares** $= \hat{\epsilon}'\hat{\epsilon} = \mathbf{y}'[\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}']\mathbf{y} = \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{Z}\hat{\beta}$

Sum of squares decomposition: $\Sigma_{j=1}^{n}(y_j - \overline{y})^2 = \Sigma_{j=1}^{n}(\hat{y}_j - \overline{y})^2 + \Sigma_{j=1}^{n}\hat{\epsilon}_j^2$

Under the general linear regression model, the least squares estimator $\hat{\beta}$ has $E(\hat{\beta}) = \beta$ and $\text{Cov}(\hat{\beta}) = \sigma^2(\mathbf{Z}'\mathbf{Z})^{-1}$.

$E(\hat{\epsilon}) = \mathbf{0}$ and $\text{Cov}(\hat{\epsilon}) = \sigma^2[\mathbf{I} - \mathbf{H}]$

Confidence region for β: $100(1-\alpha)\%$ confidence region for β given by

$(\beta - \hat{\beta})'\mathbf{Z}'\mathbf{Z}(\beta - \hat{\beta}) \leq (r+1)s^2 F_{r+1, n-r-1}(\alpha)$

where $s^2 = \hat{\epsilon}'\hat{\epsilon}/(n-r-1)$

Simultaneous $100(1-\alpha)\%$ confidence intervals for the β_i are given by,

$\hat{\beta}_i \pm \sqrt{\widehat{\text{Var}}(\hat{\beta}_i)}\sqrt{(r+1)F_{r+1,n-r-1}(\alpha)}, i = 0, 1, \ldots, r$

where $\widehat{\text{Var}}(\hat{\beta}_i)$ is the diagonal element of $s^2(\mathbf{Z}'\mathbf{Z})^{-1}$ corresponding to $\hat{\beta}_i$

If the errors ϵ in the linear regression model are normally distributed, then a $100(1-\alpha)\%$ **confidence interval** for $E(Y_0|\mathbf{z}_0) = \mathbf{z}_0'\beta$ is given by,

$\mathbf{z_0}'\beta \pm t_{n-r-1}\left(\frac{\alpha}{2}\right)\sqrt{(\mathbf{z_0}'(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{z_0})s^2}$

A prediction interval for Y_0 is given by,

$\mathbf{z_0}'\beta \pm t_{n-r-1}\left(\frac{\alpha}{2}\right)\sqrt{(1 + \mathbf{z_0}'(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{z_0})s^2}$

Multivariate Multiple Regression

$\mathbf{Y} = \mathbf{Z}\beta + \epsilon$ where $\mathbf{Y}$ has size $(n \times m)$, $\mathbf{Z}(n \times (r+1))$, $\beta((r+1) \times m)$, and $\epsilon(n \times m)$

Simply stated, the ith response $\mathbf{Y}_{(i)}$ follows the linear regression model $\mathbf{Y}_{(i)} = \mathbf{Z}\beta_{(i)} + \epsilon_{(i)}$, $i = 1, 2, \ldots, m$.

Each $\hat{\beta}_i$ is defined the same as the single-response solution as well. Collecting these univariate estimates, we obtain,

$\hat{\beta} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y}$

For any choise of parameters $\mathbf{B}$, the matrix of errors is $\mathbf{Y} - \mathbf{ZB}$. The error sum of squares and cross products matrix is $(\mathbf{Y} - \mathbf{ZB})'(\mathbf{Y} - \mathbf{ZB})$

Predicted values: $\hat{\mathbf{Y}} = \mathbf{Z}\hat{\beta} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y}$

Residuals: $\hat{\epsilon} = \mathbf{Y} - \hat{\mathbf{Y}} = [\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}']\mathbf{Y}$

Total sum of squares and cross products decomposition: $\mathbf{Y}'\mathbf{Y} = \hat{\mathbf{Y}}'\hat{\mathbf{Y}} + \hat{\epsilon}'\hat{\epsilon}$

Residual sum of squares and cross products can also be written as $\hat{\epsilon}'\hat{\epsilon} = \mathbf{Y}'\mathbf{Y} - \hat{\mathbf{Y}}'\hat{\mathbf{Y}} = \mathbf{Y}'\mathbf{Y} - \hat{\beta}'\mathbf{Z}'\mathbf{Z}\hat{\beta}'$

$\hat{\mathbf{\Sigma}} = \frac{1}{n}\hat{\epsilon}'\hat{\epsilon}$

Confidence ellipsoid for $\beta'\mathbf{z}_0$ is given by

$$(\beta'\mathbf{z_0} - \hat{\beta}'\mathbf{z_0})'\left(\frac{n}{n-r-1}\hat{\mathbf{\Sigma}}\right)^{-1}\left(\frac{\hat{\beta}'\mathbf{z_0} - \beta'\mathbf{z_0}}{\sqrt{\mathbf{z_0}'(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{z_0}}}\right)$$

Prediction ellipsoid for $\mathbf{Y_0}$

$$\mathbf{z_0}'\hat{\beta}_{(\mathbf{i})} \pm \sqrt{\left(\frac{m(n-r-1)}{n-r-m}\right)F_{m,n-r-m}(\alpha)}\sqrt{(1 + \mathbf{z_0}'(\mathbf{Z}'\mathbf{Z})^{-1})\mathbf{z_0})\left(\frac{n}{n-r-1}\hat{\sigma}_{ii}\right.}$$