LIN MOD M3 : 1

MAURICE
DIESENDRUCK

Given 1-way ANOVA model $y_{ij} = \mu + \alpha_i + e_{ij}$, where $i = 1, \dots, t$; $j = 1, \dots, N_i$; and $e_{ij} \sim N(0, \sigma^2)$. Show that $\alpha_1 = \alpha_2 = \dots = \alpha_t$ iff all contrasts $\lambda'\beta = 0$.

First, show implication of $\alpha_1 = \dots = \alpha_t$, i.e. what is actually being tested. Then equate that 'overall property' to a property of each individual contrast.

$$Y = \underset{X}{\begin{array}{c} \begin{array}{ccccc} \mu & \alpha_1 & \alpha_2 & \cdots & \alpha_t \end{array} \\ \left[\begin{array}{cccccc} 1 & 1 & & & & \\ \vdots & 1 & & & & \}N_1 \\ & 1 & & & & \\ & & 1 & & & \\ & & 1 & & & \}N_2 \\ & & 1 & & & \\ & & & \ddots & & \\ 1 & & & & & \end{array}\right] \end{array}} \underset{\beta}{\begin{bmatrix} \mu \\ \alpha_1 \\ \vdots \\ \alpha_t \end{bmatrix}} + \underline{e}$$

$\alpha_1 = \dots = \alpha_t$ implies a reduced model:

$$Y = J\mu + \underline{e}$$

Notation:

Consider $X = [\, J \mid X_1 \cdots X_t \,]$. Let $X_* = [X_1 \cdots X_t]$, M_* : ppm onto $C(X_*)$, M : ppm onto $C(X)$, and $M - M_*$: ppm onto $C(X - X_*)$.

The estimation space under H_0 is $C(J)$; while the test space is $C(X_*) = C(M_*)$.
(Prop B.32)

With an orthonormal basis $R = [R_1 \cdots R_t]$ of $C(X_*)$, can write $M_* = RR'$ (Th. B.35).

$$\boxed{\begin{array}{l} M_* = RR' = \sum_{i=1}^{t} R_i R_i' = \sum_{i=1}^{t} M_i \\ Y'M_*Y = \sum_{i=1}^{t} Y'M_iY \end{array}}$$

Note: Since R_i's are orthonormal, $M_iM_j = 0$, and functions $Y'M_iY$ and $Y'M_jY$ are independent. (Th. 1.3.7)

A hypothesis tested using $Y'M_*Y$ would test $0 = \beta'X'M_*X\beta$. Since M_* and M_i's are nonnegative definite, for $0 = \beta'X'M_*X\beta = \sum_{i=1}^{t} \beta'X'M_iX\beta$ to hold, $\beta'X'M_iX\beta > 0 \ \forall i$.

This implies $\beta'X'[R_i(R_i'R_i)^- R_i']X\beta = 0$, or $R_i'X\beta = 0 \ \forall i$.

Equivalently, if any $\beta'X'M_iX\beta > 0$, then $\beta'X'M_*X\beta = 0$, and H_0 no longer holds. ∎

$$y_i = \beta_0 + \beta_1 x_{i1} + \dots \beta_{p-1} x_{ip-1} + e_i$$

$$X = \left[J \;\vdots\; X_* \right], \quad X_* \text{ is } n \times (p-1)$$

$$\beta = \begin{bmatrix} \beta_0 \\ | \\ \beta_* \\ | \end{bmatrix}, \quad \beta_* \text{ is } (p-1) \times 1$$

Mean-centered

$$y_i = \gamma_0 + \gamma_1 (x_{i1} - \bar{x}_{\cdot 1}) + \dots + \gamma_{p-1} (x_{ip-1} - \bar{x}_{\cdot p-1}) + e_i$$

$$X = \left[J \;\vdots\; (I - M_J) z \right], \quad (I - M_J) z \text{ is } n \times (p-1)$$

$$\gamma = \begin{bmatrix} \gamma_0 \\ | \\ \gamma_* \\ | \end{bmatrix}, \quad \gamma_* \text{ is } (p-1) \times 1$$

$$\beta_0 + X_* \beta_* = E[Y] = \gamma_0 + \left[(I - M_J) z \right] \gamma_* \qquad \boxed{\star}$$

Average over observations:

$$\beta_0 + \beta_1 \bar{x}_1 + \dots + \beta_{p-1} \bar{x}_{p-1} = \gamma_0 + \underbrace{\gamma_1 (\bar{x}_1 - \bar{x}_1) + \dots + \gamma_{p-1} (\bar{x}_{\cdot p-1} - \bar{x}_{\cdot p-1})}_{0}$$

Let $\underset{(p-1) \times 1}{\bar{X}_*} = \left[\bar{x}_{\cdot 1}, \dots, \bar{x}_{\cdot p-1} \right]'$

$$\beta_0 + \bar{X}_*' \beta_* = \gamma_0$$

$$\beta_0 = \gamma_0 - \bar{X}_*' \beta_*$$

Substitute β_0 into $\boxed{\star}$: $\cancel{\gamma_0} - \bar{X}_*' \beta_* + X_* \beta_* = \cancel{\gamma_0} + \left[(I - M_J) z \right] \gamma_*$

$$\left[\underset{n \times (p-1)}{X_*} - \underset{n \times (p-1)}{\mathbb{1}_n \bar{X}_*'} \right] \beta_* = \left[(I - M_J) z \right] \gamma_*$$

For each i, $\left[(x_{i1} - \bar{x}_{\cdot 1}), \dots, (x_{ip-1} - \bar{x}_{\cdot p-1}) \right] \beta_* \equiv \left[(I - M_J) z \right] \beta_*$

$$\boxed{\text{Therefore } \beta_* = \gamma_*, \text{ and } \beta_0 = \gamma_0 - \bar{x}_*' \gamma_*}$$

Book denotes $\bar{X}_*'$ as $\left(\frac{1}{n}\right) J_1^n Z$, so $\beta_0 = \gamma_0 - \left(\frac{1}{n}\right) J_1^n Z \gamma_*$

From above, in the mean-centered version, $C((I - M_J) z) = C(J)^{\perp}_{C(X)} = C(M - M_J)$

Normal equation $X'X\gamma = X'Y$: $X'X = \underset{p \times n}{\begin{bmatrix} - J - \\ - (I - M_J) z - \end{bmatrix}} \underset{n \times p}{\begin{bmatrix} | & | \\ J & (I - M_J) z \\ | & | \end{bmatrix}} = \underset{p \times p}{\begin{bmatrix} n & — 0 — \\ | & \\ 0 & [(I - M_J) z]' [(I - M_J) z] \\ | & \end{bmatrix}}$,

$[X'X]^{-1} = \underset{p \times p}{\begin{bmatrix} \frac{1}{n} & — 0 — \\ | & \\ 0 & \dfrac{1}{[(I - M_J) z]' [(I - M_J) z]} \\ | & \end{bmatrix}}$ is invertible because X is full rank , $X'Y = \underset{p \times 1}{\begin{bmatrix} \Sigma y_i \\ | \\ [(I - M_J) z]' Y \end{bmatrix}}$

The L.S. estimate of γ is $\hat{\gamma} = (X'X)^{-1}X'Y = \begin{bmatrix} \frac{\sum_i y_i}{n} \\ \frac{[(I-M_J)Z]'Y}{[(I-M_J)Z]'[(I-M_J)Z]} \end{bmatrix} = \begin{bmatrix} \bar{y} \\ \frac{[(I-M_J)Z]'Y}{[(I-M_J)Z]'[(I-M_J)Z]} \end{bmatrix} = \begin{bmatrix} \hat{\gamma}_0 \\ \hat{\gamma}_* \end{bmatrix}$

From previous result, know that $\gamma_* = \beta_*$, so $\hat{\gamma}_* = \hat{\beta}_*$. This is useful in that we can relate Y under the reduced model, with $X = \left[J \,\vdots\, (I-M_J)Z\right]$, to the Least Squares estimate $\hat{\beta}_*$.

For L.S., generally $MY = X\hat{\beta}$. Here $M_{(I-M_J)Z}Y = X\hat{\beta}_*$. $M_{(I-M_J)Z}$ is defined below.

SSReg $= Y'(M-M_J)Y$. Since $C(M-M_J) = C((I-M_J)Z)$, can write ppm onto $(I-M_J)Z$ as

$$(I-M_J)Z\left[Z'(I-M_J)'(I-M_J)Z\right]^{-}Z'(I-M_J)' = M_{(I-M_J)Z}$$

$= Y'(I-M_J)Z\left[Z'(I-M_J)'(I-M_J)Z\right]^{-}Z'(I-M_J)'Y$	Replacing $M-M_J$ with $M_{(I-M_J)Z}$
$= \hat{\beta}_*'Z'(I-M_J)Z\left[Z'(I-M_J)'(I-M_J)Z\right]^{-}Z'(I-M_J)'Z\hat{\beta}_*$	Use normal equations to get $\hat{\beta}_*$, then $E[Y] = Z\hat{\beta}_*$
$= \hat{\beta}_*'Z'(I-M_J)Z\left[Z'(I-M_J)Z\right]^{-}Z'(I-M_J)'Z\hat{\beta}_*$	Collapse ppm's, since they are idempotent.
$= \hat{\beta}_*'Z'(I-M_J)Z\hat{\beta}_*$ $\blacksquare$	Allow generalized inverse to cancel out.

MD

LINMOD M3 : 3

For model with intercept, expect $\underline{y}$ and $\hat{\underline{y}}$ to share the same mean, since

$$y_i = \hat{y}_i + e_i \Rightarrow \sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \hat{y}_i + \sum_{i=1}^{n} e_i, \quad \text{where } e_i \sim N(0,\sigma^2), \text{ so } \sum_{i=1}^{n} e_i = 0$$

$$\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} \hat{y}_i \Rightarrow E[\underline{y}] = \frac{\sum_{i=1}^{n} y_i}{n} = \bar{y}\mathbb{1}_n = \frac{\sum_{i=1}^{n} \hat{y}_i}{n} = E[\hat{\underline{y}}]$$

Note: $\bar{\underline{y}} = \bar{y} \cdot \mathbb{1}_n$

Correlation of $\underline{y}$ and $\hat{\underline{y}}$ $\equiv$ cosine of angle between their mean-centered vectors $\underline{y} - \bar{\underline{y}}$ and $\hat{\underline{y}} - \bar{\underline{y}}$,

because $\text{corr}(\underline{y}, \hat{\underline{y}}) = \dfrac{\sum_{i=1}^{n} (y_i - \bar{y})(\hat{y}_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2}\sqrt{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}}$ and letting $u = \underline{y} - \bar{\underline{y}}$, $v = \hat{\underline{y}} - \bar{\underline{y}}$

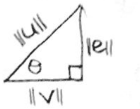
$$= \frac{u_1 v_1 + \ldots + u_n v_n}{\sqrt{u_1^2 + \ldots + u_n^2}\sqrt{v_1^2 + \ldots + v_n^2}} = \frac{u'v}{\sqrt{u'u}\sqrt{v'v}} = \frac{u \cdot v}{\|u\|\,\|v\|} = \cos\theta$$

where θ is the angle between u and v.

Geometric argument, using Pythagorean, u, and v, finds $\cos\theta$.

First, note:

1. Model has intercept, so $\mathbb{1}_n \in C(x)$ and $\bar{y}\mathbb{1}_n \equiv \bar{\underline{y}} \in C(x)$
   Since $\hat{\underline{y}}$ is a projection onto $C(x)$, $\hat{\underline{y}} \in C(x) \Rightarrow v = \hat{\underline{y}} - \bar{\underline{y}} \in C(x)$
2. Given orthogonal decomposition of $\underline{y} = \hat{\underline{y}} + \underline{e}$, where $\hat{\underline{y}} \in C(x)$ and $e \in C(x)^{\perp} \Rightarrow \underline{e} = \underline{y} - \hat{\underline{y}} \in C(x)^{\perp}$.
3. Recognize that $u = \underline{y} - \bar{\underline{y}}$, "triangulates" v and e, as the hypoteneuse:



since $e \perp v$.

Pythagorean Theorem states $\|u\|^2 = \|v\|^2 + \|e\|^2 \Rightarrow \|\underline{y} - \bar{\underline{y}}\|^2 = \|\hat{\underline{y}} - \bar{\underline{y}}\|^2 + \|\underline{y} - \hat{\underline{y}}\|^2$

$$\Rightarrow \underbrace{\sum_{i=1}^{n} (y_i - \bar{y})^2}_{SS_{total}} = \underbrace{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}_{SS_{regr}} + \underbrace{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}_{SS_{error}}$$

, which is equivalent to

Finally, $\cos\theta = \dfrac{\|v\|}{\|u\|} \Rightarrow (\cos\theta)^2 = \left[\text{corr}(\underline{y}, \hat{\underline{y}})\right]^2 = \dfrac{\|v\|^2}{\|u\|^2} = \dfrac{SS_{reg}}{SS_{total}} = R^2$ ■

$\boxed{\text{LINMOD M3 : 4}}$

MD

Let $\varepsilon_j = y_j - \hat{E}[y_j|x]$ for $j=1,2$. $\boxed{\text{Show } E[\varepsilon_j] = 0.}$

$$E[\varepsilon_j] = E\left[y_j - \hat{E}[y_j|x] \right] = E\left[E\left[y_j - \hat{E}[y_j|x] \mid x \right] \right] = E[0] = 0 \quad \blacksquare$$

Uses law of iterated expectations, as in Thm 6.3.1.

Note: $\hat{E}[y_j|x] = \mu_{y_j} + (x-\mu_x)\beta_* = \mu_{y_j} + (x-\mu_x)V_{xx}^{-1}V_{xy_j}$, and therefore

$\varepsilon_j = (y_j - \mu_{y_j}) - (x-\mu_x)V_{xx}^{-1}V_{xy_j}$.

For $\rho_{y\cdot x} = \text{corr}(\varepsilon_1, \varepsilon_2) = \dfrac{\text{cov}(\varepsilon_1, \varepsilon_2)}{\cdots} = 0 \Rightarrow \text{cov}(\varepsilon_1, \varepsilon_2) = 0$

$\text{cov}(\varepsilon_2, \varepsilon_1) = \left[\text{cov}(\varepsilon_1, \varepsilon_2)\right]' = 0' = 0$.

$\boxed{\text{Now show } \text{cov}(\varepsilon_2, \varepsilon_1) = \text{cov}(y_2, \varepsilon_1).}$

$$\text{cov}(\varepsilon_2, \varepsilon_1) = \text{cov}\left(y_2 - \hat{E}[y_2|x], \varepsilon_1 \right) = \text{cov}(y_2, \varepsilon_1) - \cancelto{0}{\text{cov}\left(\hat{E}[y_2|x], \varepsilon_1 \right)} \quad \blacksquare$$

where the second term $= 0$, because predictions are orthogonal to errors, so their correlation (and thus covariance) is 0.

Together, $\rho_{yx} = 0 \Rightarrow \text{cov}(\varepsilon_2, \varepsilon_1) = \text{cov}(y_2, \varepsilon_1) = 0 \;\}$ No correlation between ε_1 and ε_2, an no correlation between y_2 and ε_1, or equivalently between y_1 and ε_2.

$\boxed{\text{Show } \text{cov}(\varepsilon_j, x-\mu_x) = 0.}$

$$\begin{aligned}
\text{cov}(\varepsilon_j, x-\mu_x) &= \text{cov}\left((y_j - \mu_{y_j}) - (x-\mu_x)V_{xx}^{-1}V_{xy_j}, \; x-\mu_x \right) \\
&= \text{cov}\left(y_j - \mu_{y_j}, \; x-\mu_x \right) - \text{cov}\left((x-\mu_x)V_{xx}^{-1}V_{xy_j}, \; x-\mu_x \right) \\
&= V_{y_j x} - V_{xx} \cdot V_{xx}^{-1}V_{xy_j} = 0 \quad \blacksquare
\end{aligned}$$

There is no correlation between the errors and the mean-centered data.

A best linear predictor $\hat{E}[y_2|x, y_1]$ predicts based on some linear function of x and y_1: $f(x, y_1)$.

This $f(x, y_1)$ can be equivalently expressed without y_1, using ε_1 instead, as $g(x-\mu_x, \varepsilon_1)$.

Since y_2 is not correlated to ε_1 $(\text{cov}(y_2, \varepsilon_1) = 0)$, an equivalent predictor of y_2 can be based on $g(x-\mu_x)$ or $g'(x)$, so that $\hat{E}[y_2|x] = \hat{E}[y_2|x, y_1]$.

LINMOD M3 : 5

5A. Group (i,j) = (1,1) has lowest average score : 0.912 .

Group (i,j) = (1,2) has highest average score : 2.854 .

Group (i,j) = (1,1) has the lowest absolute score : 0.155 .

Group (i,j) = (2,1) has the highest absolute score : 4.285 .

5B. (i) The null hypothesis being tested is $H_0 : R^2 = 0$, or that beyond the mean, the regression covariates <u>are not</u> informative for estimating the response.

The alternative hypothesis is $H_a : R^2 > 0$, or that the regression covariates, beyond the mean <u>are</u> informative, and "explain" some of the variance in the response.

(ii) The F-statistic at 5 and 26 DF is 2.746, with p-value 0.04032, indicating that under H_0, the probability under repeated trials of explaining that much or more of the variability in the response is ~ 4% .

I would be skeptical of concluding that major and background are informative, but would say that the low p-value suggests that more exploration for relationship <u>is warranted</u>.

5C. The lm p-value does not contradict the anova p-value, because anova's output depends on factor order, and demonstrates <u>sequential</u> changes to sum of squares with each category of covariates ; in contrast, the p-value in lm is the maginal test of that covariate being 0, compared to the baseline case.

ANOVA results indicate that, <u>taken together</u>, BG factors <u>do not</u> significantly contribute to response variability.