more on testing

- one-way ANOVA
- multiple ~~comparison~~ regression
- multiple comparison
- contrast.

Exam 2 → this coming Tue

M 6

Ronald 3.1 - 3.5

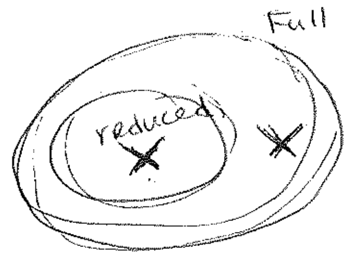
∗ Now we will talk about **testing**.

likelihood ratio test

∗∗ Testing linear parametric functions (first principles test)

∗∗ Testing models

∗∗ Confidence intervals and multiple comparisons



♣ Let's do **testing models**.

Start with a model that we assume valid

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{e}, \qquad \mathbf{e} \sim \mathrm{N}_n(\mathbf{0}, \sigma^2 I). \qquad \text{Full model} \tag{3}$$

Our wish is to reduce this model (i.e. simpler model, putting more constraints on the estimation space)

$$\mathbf{y} = \mathbf{X}_0\gamma + \mathbf{e}, \qquad \mathbf{e} \sim \mathrm{N}_n(\mathbf{0}, \sigma^2 I) \qquad C(\mathbf{X}_0) \subset C(\mathbf{X}). \tag{4}$$

Note: if the reduced model is correct, the big model is also correct. **Our question is whether the reduced model is correct.**

$*$ Ex3 (contd): Consider the one-way ANOVA model;

$$y_{ij} = \mu + \alpha_i + e_{ij}, i = 1, 2, 3 \text{ and } j = 1, 2.$$

- Define the reduced model to test for no treatment effects.

$$X = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \qquad X_0 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

- Define the reduced model to test $\alpha_1 - \alpha_3 = 0$. $\Leftrightarrow \alpha_1 = \alpha_3$

$$X_0 = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \end{bmatrix}$$

$*$ Ex4: Consider the full multiple regression the one-way ANOVA model;

$$y_i = \beta_0 + \beta_1 \cancel{x_{1i}} + \beta_2 x_{2i} + \cancel{\beta_3 x_{3i}} + e_i, i = 1, 2, 3 \quad \text{and} \quad j = 1, 2.$$

- Define the reduced model to test whether x_1 and x_3 are adding significantly to the explanatory capability of the regression model.

$$y_i = \gamma_0 + \gamma_2 x_{2i} + e_i$$

- Define the reduced model to test $\beta_2 - \beta_3 = 0$. ($\beta_2 = \beta_3$)

$$y_i = \gamma_0 + \gamma_1 x_{1i} + \gamma_2 (x_{2i} + x_{3i}) + e_i$$

$$\gamma = \begin{bmatrix} \gamma_0 \\ \gamma_1 \\ \gamma_2 \end{bmatrix} \qquad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}$$

$*$ Consider the following hypothesis.

$$H_0 : \mathrm{E}(\mathbf{y}) = \mathbf{X}_0\gamma \quad \text{for some } \gamma \qquad \Leftrightarrow \qquad H_0 : \mathrm{E}(\mathbf{y}) \in C(\mathbf{X}_0)$$

versus

$$H_1 : \underbrace{\mathrm{E}(\mathbf{y})}_{= X\beta} \in C(\mathbf{X}) \qquad \text{and} \qquad \mathrm{E}(\mathbf{y}) \notin C(\mathbf{X}_0)$$

$E(y) = X\beta \Rightarrow \hat{y} = P_X y$

Q: How to build a test statistic?

$C(X_0) \subset C(X)$

- Recall! Let $\mathbf{P}$ and $\mathbf{P}_0$ be the perpendicular projection operators onto $\underbrace{C(\mathbf{X})}_{Ha}$ and $\underbrace{C(\mathbf{X}_0)}_{Ho}$, respectively.

With $C(\mathbf{X}_0) \subset C(\mathbf{X})$, $\mathbf{P}-\mathbf{P}_0$ is the perpendicular projection operator onto the orthogonal complement of $C(\mathbf{X}_0)$ with respect to $C(\mathbf{X})$, that is, $C(\mathbf{P}-\mathbf{P}_0) = C(\mathbf{X}_0)^{\perp}_{C(\mathbf{X})}$.

- $\overbrace{\hat{\mathbf{y}} = \mathbf{P}\mathbf{y}}^{Ha}$ and $\overbrace{\hat{\mathbf{y}} = \mathbf{P}_0\mathbf{y}}^{H_0}$ are estimates of $\mathrm{E}(\mathbf{y})$ under models (1) and (2), respectively.

$E(y) = X_0 \cdot \gamma$

$E(y) = X\beta$

reduced model

- **If model (2) is true,** $X_0\gamma = X\beta$
  ($\Leftrightarrow$ H_0 is true)

$*$ $\mathbf{Py}$ and $\mathbf{P}_0\mathbf{y}$ are estimates of the same quantity

$*$ $\mathrm{E}(\mathbf{P}-\mathbf{P}_0)\mathbf{y} = \mathbf{0}$.

$$\Rightarrow \text{small } \underbrace{\mathbf{Py}}_{\hat{y} \text{ under } H_A} - \overbrace{\mathbf{P}_0\mathbf{y}}^{= \hat{y} \text{ under } H_0} = (\mathbf{P}-\mathbf{P}_0)\mathbf{y}$$

reduced model

- **If model (2) is NOT true,**

($\Leftrightarrow$ Ho is not true)

$*$ $\mathrm{E}(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta} \in C(\mathbf{X}) = C(\mathbf{P})$

$*$ $\mathbf{P}_0\mathrm{E}(\mathbf{y}) = \mathbf{P}_0\mathbf{X}\boldsymbol{\beta} \neq \mathrm{E}(\mathbf{y})$

$*$ $\mathbf{Py}$ and $\mathbf{P}_0\mathbf{y}$ estimate different things.

$\Rightarrow$ large $\mathbf{Py} - \mathbf{P}_0\mathbf{y} = (\mathbf{P} - \mathbf{P}_0)\mathbf{y}$

$(P_k - P_0)$

$n \times 1$

- The difference is $\mathbf{P}\mathbf{y} - \mathbf{P}_0\mathbf{y} = (\mathbf{P} - \mathbf{P}_0)\mathbf{y}$.

* The length of the difference is $\|(P-P_0)y - 0\|^2 = \mathbf{y}^T(\mathbf{P} - \mathbf{P}_0)\mathbf{y}$

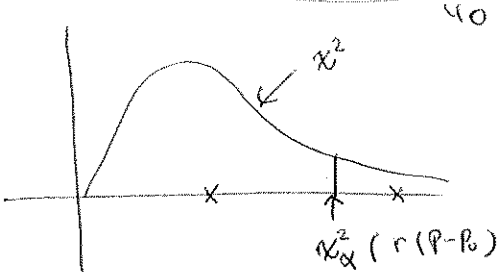
$$= \|(P-P_0)y\|^2 = \left((P-P_0)y\right)^T\left((P-P_0)y\right) = y^T(P-P_0)y$$

* The distribution of $\mathbf{y}^T\underbrace{(\mathbf{P} - \mathbf{P}_0)}_{= r(X) - r(X_0)}\mathbf{y}$?

$$\frac{y^T(P-P_0)y}{\sigma^2} \sim \chi^2\left(r(P-P_0), \underbrace{\frac{(X\beta)^T(P-P_0)(X\beta)}{2\cdot\sigma^2}}_{=\phi}\right)$$

H_0

large $y^T(P-P_0)y$ is an evidence ~~for~~ against H_0

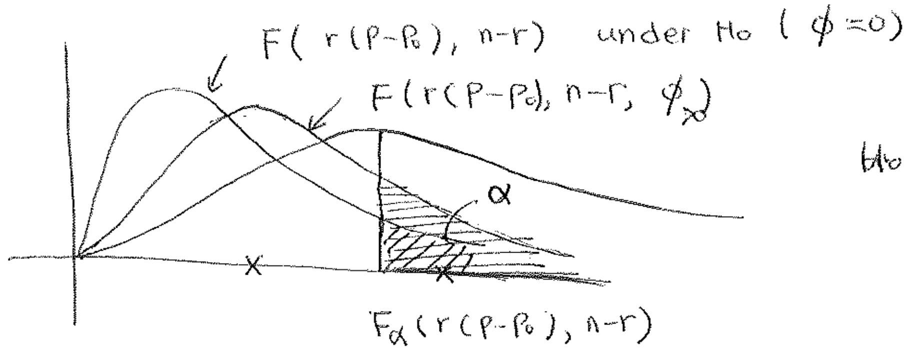


$$\hat{\sigma}^2 = MSE = \frac{y^T(I-P)y}{(n-r)}$$

$$\frac{y^T(I-P)y}{\sigma^2} \sim \chi^2(n-r, 0)$$

$$\frac{\dfrac{y^T(P-P_0)y}{\cancel{\sigma^2}} \Big/ r(P-P_0)}{\dfrac{y^T(I-P)y}{\cancel{\sigma^2}} \Big/ (n-r)} \sim F(r(P-P_0), n-r, \phi)$$

$$= \frac{\dfrac{y^T(P-P_0)y}{r(P-P_0)}}{\dfrac{y^T(I-P)y}{(n-r)}}$$



H_0 is not true, $\Rightarrow \phi \neq 0$

$\phi > 0$

power $=$ P(reject H_0 | H_0 is not true)

$= P\left(F > F_\alpha(r(P-P_0), n-r) \mid F \sim F(r(P-P_0), n-r, \phi)\right)$

larger ϕ ($\|Py - P_0y\|^2$ becomes larger)

- Let's estimate σ^2.

- Recall: model (1) is assumed to be valid.
  Full model

$*$ MSE is the unbiased estimator of σ^2.
$*$ The distribution of $\mathbf{y}^T(\mathbf{P} - \mathbf{P}_0)\mathbf{y}$?

♠ Assume that the model

$$H_a: \mathbf{y} = \mathbf{X}\beta + \mathbf{e}, \quad \mathbf{e} \sim \mathrm{N}_n(\mathbf{0}, \sigma^2 I)$$

holds for some values of β and σ^2 and consider the reduced model

$$H_0: \mathbf{y} = \mathbf{X}_0\gamma + \mathbf{e}, \quad \mathbf{e} \sim \mathrm{N}_n(\mathbf{0}, \sigma^2 I), \quad C(\mathbf{X}_0) \in C(\mathbf{X}).$$

Let $\mathbf{P}$ and $\mathbf{P}_0$ be the perpendicular projection matrices into $C(\mathbf{X})$ and $C(\mathbf{X}_0)$, respectively. Then

- $$\frac{\mathbf{y}^T(\mathbf{P}-\mathbf{P}_0)\mathbf{y}/r(\mathbf{P}-\mathbf{P}_0)}{\mathbf{y}^T(I-\mathbf{P})\mathbf{y}/r(I-\mathbf{P})} \sim F(r(\mathbf{P}-\mathbf{P}_0), r(I-\mathbf{P}), \beta^T \underbrace{\mathbf{X}^T(\mathbf{P}-\mathbf{P}_0)\mathbf{X}\beta/(2\sigma^2)}_{=0 \text{ under } H_0})$$

- The reduced model is true iff

$$F = \frac{\mathbf{y}^T(\mathbf{P}-\mathbf{P}_0)\mathbf{y}/r(\mathbf{P}-\mathbf{P}_0)}{\mathbf{y}^T(I-\mathbf{P})\mathbf{y}/r(I-\mathbf{P})} \sim F(r(\mathbf{P}-\mathbf{P}_0), r(I-\mathbf{P}), 0)$$

$*$ Note that a nonzero noncentrality parameter shifts the (central) F distribution to the right.

$*$ Consider the following hypothesis.

$$H_0 : \mathrm{E}(\mathbf{y}) = \mathbf{X}_0\gamma \quad \text{for some } \gamma \qquad \Leftrightarrow \qquad H_0 : \mathrm{E}(\mathbf{y}) \in C(\mathbf{X}_0)$$

versus

$$H_1 : \mathrm{E}(\mathbf{y}) \in C(\mathbf{X}) \qquad \text{and} \qquad \mathrm{E}(\mathbf{y}) \notin C(\mathbf{X}_0)$$

Reject H_0 if

$$F = \frac{\mathbf{y}^T(\mathbf{P} - \mathbf{P}_0)\mathbf{y}/r(\mathbf{P} - \mathbf{P}_0)}{\mathbf{y}^T(I - \mathbf{P})\mathbf{y}/r(I - \mathbf{P})} > F_\alpha(r(\mathbf{P} - \mathbf{P}_0), r(I - \mathbf{P})).$$

♣ The first principles F test for the general linear hypothesis is the same as the full versus reduced F test as we saw.

♣ The F test that we developed is equivalent to the likelihood ratio test for the same hypothesis.

Our model : $y \sim N_n(X\beta, \sigma^2 I_n)$

→ • $H_0: \Lambda^T\beta = m$ w/ $\Lambda^T\beta$ estimable ($\Leftrightarrow$ columns of Λ are in $\mathcal{C}(X^T)$)

$\Updownarrow$

• $H_0: y = X_0\beta_0 + e$

parameter space : $\Omega_0 = \{(\beta, \sigma^2) : \Lambda^T\beta = m, \ \sigma^2 > 0\}$: the model restricted by H_0

$$\Omega = \{(\beta, \sigma^2) : \beta \in \mathbb{R}^p, \ \sigma^2 > 0\}$$

$\Omega = \begin{bmatrix}\beta_0\\ \beta_1\\ \beta_2\end{bmatrix} \in \mathbb{R}^3$ $\quad \begin{bmatrix}\beta_0\\ \beta_1\\ \beta_2\end{bmatrix} \in \mathbb{R}^3$ w/ $\beta_1 = 0$ $= \Omega_0$

Likelihood Ratio Test

$$L(\beta, \sigma^2) = \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2}(y - X\beta)^T(y - X\beta)\right)$$

$$\Rightarrow \quad \ell(\beta, \sigma^2) = -\frac{n}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(y - X\beta)^T(y - X\beta)$$

$$\phi(y) = \frac{\max_{\Omega_0} L(\beta, \sigma^2)}{\max_{\Omega} L(\beta, \sigma^2)} < c \quad \Rightarrow \text{ reject } H_0$$

Goal: want ~~to~~ find ~~such~~ c such that

$$P(\phi(y) < c \mid H_0) = \alpha$$

Step 1: Find σ^2 that maximizes $\ell(\beta, \sigma^2)$ for any given β

$$\frac{\partial \ell(\beta, \sigma^2)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}(y - X\beta)^T(y - X\beta) = 0$$

$$\Rightarrow \quad \sigma^2 = \frac{(y - X\beta_0)^T(y - X\beta)}{n} = \frac{Q(\beta)}{n}$$

$$\hat{\sigma}_0^2 = \frac{Q(\hat{\beta}_0)}{n}$$

$$\hat{\sigma}^2 = \frac{Q(\hat{\beta})}{n}$$

Step 2 Find c

$$\phi(y) = \frac{\left(\frac{\cancel{n}}{2\cancel{\pi}Q(\hat{\beta}_0)}\right)^{\frac{n}{2}} \exp\left(-\underbrace{\frac{1}{2\hat{\sigma}_0^2}(y - X\hat{\beta}_0)^T(y - X\hat{\beta}_0)}_{=\cancel{-\frac{n}{2}}}\right)}{\left(\frac{\cancel{n}}{2\cancel{\pi}Q(\hat{\beta})}\right)^{\frac{n}{2}} \cancel{\exp\left(-\frac{n}{2}\right)}}$$

$\hat{\beta}_0$: MLE under Ω_0

$\hat{\beta}$: MLE under Ω

$$= \left(\frac{Q(\hat{\beta}_0)}{Q(\hat{\beta})}\right)^{-\frac{n}{2}}$$

$$\Rightarrow \quad \text{Reject } H_0 \text{ if } \phi(y) = \left(\frac{Q(\hat{\beta}_0)}{Q(\hat{\beta})}\right)^{-\frac{n}{2}} < c$$

$$\frac{Q(\hat{\beta}_0)}{Q(\hat{\beta})} > c^{-\frac{2}{n}}$$

$$\Leftrightarrow \quad \frac{Q(\hat{\beta}_0) - Q(\hat{\beta})}{Q(\hat{\beta})} > c^{-\frac{2}{n}} - 1$$

$$\Leftrightarrow \quad \frac{(Q(\hat{\beta}_0) - Q(\hat{\beta}))/s}{Q(\hat{\beta})/(n-r)} > \underbrace{(c^{-\frac{2}{n}} - 1) \cdot \frac{n-r}{s}}_{F_\alpha(s,\, n-r)}$$

Now look at $Q(\hat{\beta})$ and $Q(\hat{\beta}_0)$

- $\hat{\beta}$ minimizes $Q(\beta) = (y - X\beta)^T(y - X\beta)$ for $\beta \in \mathbb{R}^p$

$$\Rightarrow \quad Q(\hat{\beta}) = \| y - Py \|^2$$

- Similarly: $\hat{\beta}_0$ minimizes $Q(\beta) = (y - X\beta)^T(y - X\beta)$ for

$\beta \in \mathbb{R}^p$ w/ constraints imposed by $\underset{\text{(testable)}}{\underline{\Lambda^T\beta = m}}$

$$\Rightarrow \quad Q(\hat{\beta}_0) = \| y - P_0 y \|^2$$

$$= y^T(I - P_0)y \qquad = y^T(I - P)y$$

$\text{rank}(\Lambda)$ ← $p \times s$ $\qquad X : n \times p$

$\Lambda^T\beta$

$$F = \frac{\left(\| y - P_0 y \|^2 - \| y - Py \|^2 \right) / s}{\| y - Py \|^2 / (n - r)} \sim F(s, n-r)$$

$\Rightarrow$ The F ~~what~~ we developed is equiv. to the LRT.

- Th 6.1 If $\Lambda^T\boldsymbol{\beta}$ is a set of linearly independent estimable functions, and $\hat{\boldsymbol{\beta}}_0$ is part of a solution to the restricted normal equations with constraint $\Lambda^T\boldsymbol{\beta} = \mathbf{m}$, then

$$\begin{aligned} Q(\hat{\boldsymbol{\beta}}_0) - Q(\hat{\boldsymbol{\beta}}) &= (\hat{\boldsymbol{\beta}}_0 - \hat{\boldsymbol{\beta}})^T\mathbf{X}^T\mathbf{X}(\hat{\boldsymbol{\beta}}_0 - \hat{\boldsymbol{\beta}}) \\ &= (\Lambda^T\hat{\boldsymbol{\beta}} - \mathbf{m})^T(\Lambda^T(\mathbf{X}^T\mathbf{X})^-\Lambda)^{-1}(\Lambda^T\hat{\boldsymbol{\beta}} - \mathbf{m}). \end{aligned}$$

$\Rightarrow$ Testing on $\Lambda^T\beta = m$ is equiv. to Testing by model comparison.

diet: A, B, C, D

$$\hat\beta = (X^TX)^{-1}X^Ty$$

$y_{ij} = \theta_i + e_{ij}$ $\hat\theta_i = \bar y_{i\cdot}$

$\Rightarrow \hat y_{ij} = \hat\theta_i = \bar y_{i\cdot}$

$i=1,2,3,4$

$j=1,\dots,n_i$ $(4, 6, 6, 8)$

$N = 4+6+6+8 = 24$

Coagulation Example– revisit

$$X = \begin{bmatrix} 1_4 & 0_4 & 0_4 & 0_4 \\ 0_6 & 1_6 & 0_6 & 0_6 \\ 0_6 & 0_6 & 1_6 & 0_6 \\ 0_8 & 0_8 & 0_8 & 1_8 \end{bmatrix} \quad \beta = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \theta_3 \\ \theta_4 \end{bmatrix}$$

$r(X) = 4 = r$

```
> g1 <- lm(coag ~ diet - 1, coagulation)
> summary(g1)

Call:
lm(formula = coag ~ diet - 1, data = coagulation)

Residuals:
   Min     1Q Median     3Q    Max
 -5.00  -1.25   0.00   1.25   5.00

Coefficients:
      Estimate Std. Error t value Pr(>|t|)
dietA  61.0000     1.1832   51.55   <2e-16 ***
dietB  66.0000     0.9661   68.32   <2e-16 ***
dietC  68.0000     0.9661   70.39   <2e-16 ***
dietD  61.0000     0.8367   72.91   <2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 2.366 on 20 degrees of freedom
Multiple R-squared:  0.9989, Adjusted R-squared:  0.9986
F-statistic:  4399 on 4 and 20 DF,  p-value: < 2.2e-16
```

$\leftarrow H_0: \theta_1 = 0$

$\leftarrow H_0: \theta_2 = 0$

$\vdots$

r

$N - r = 24 - 4 = 20$

Coagulation Example– revisit

$\theta_1 = \theta_2 = \theta_3 = \theta_4 = \mu$

$y_{ij} = \mu + e_{ij}$

$\Rightarrow \hat{\mu} = \bar{y}_{\cdot\cdot}$

$\hat{y}_{ij} = \hat{\mu} = \bar{y}_{\cdot\cdot}$

$\text{rank}(X_0) = 1$

$(X_0^T X_0)^{-1} = \frac{1}{N}$

$$X_0 = \begin{bmatrix} 1_4 \\ 1_6 \\ 1_6 \\ 1_8 \end{bmatrix}$$

```
> gnull <- lm(coag ~ 1, coagulation)
> summary(gnull)

Call:
lm(formula = coag ~ 1, data = coagulation)

Residuals:
   Min     1Q Median     3Q    Max
 -8.00  -2.25  -0.50   3.00   7.00

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  64.0000     0.7848   81.55   <2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 3.845 on 23 degrees of freedom
```

Coagulation Example– revisit

```
> anova(gnull, g1)
Analysis of Variance Table

Model 1: coag ~ 1
Model 2: coag ~ diet - 1
  Res.Df RSS Df Sum of Sq      F    Pr(>F)
1     23 340
2     20 112  3       228 13.571 4.658e-05 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
>
```

$N - r(X_0)$

$24 - 1$

$24 - 4$

$N - r(X)$

$340 - 112$

$\sim F(3, 20)$

~~reject H_0 F~~

reject H_0 if

$F > F_\alpha(3, 20)$

$P(F > 13.571)$

$$\underbrace{r(X) - r(X_0)}_{e(X_0)} = 4 - 1 = 3$$

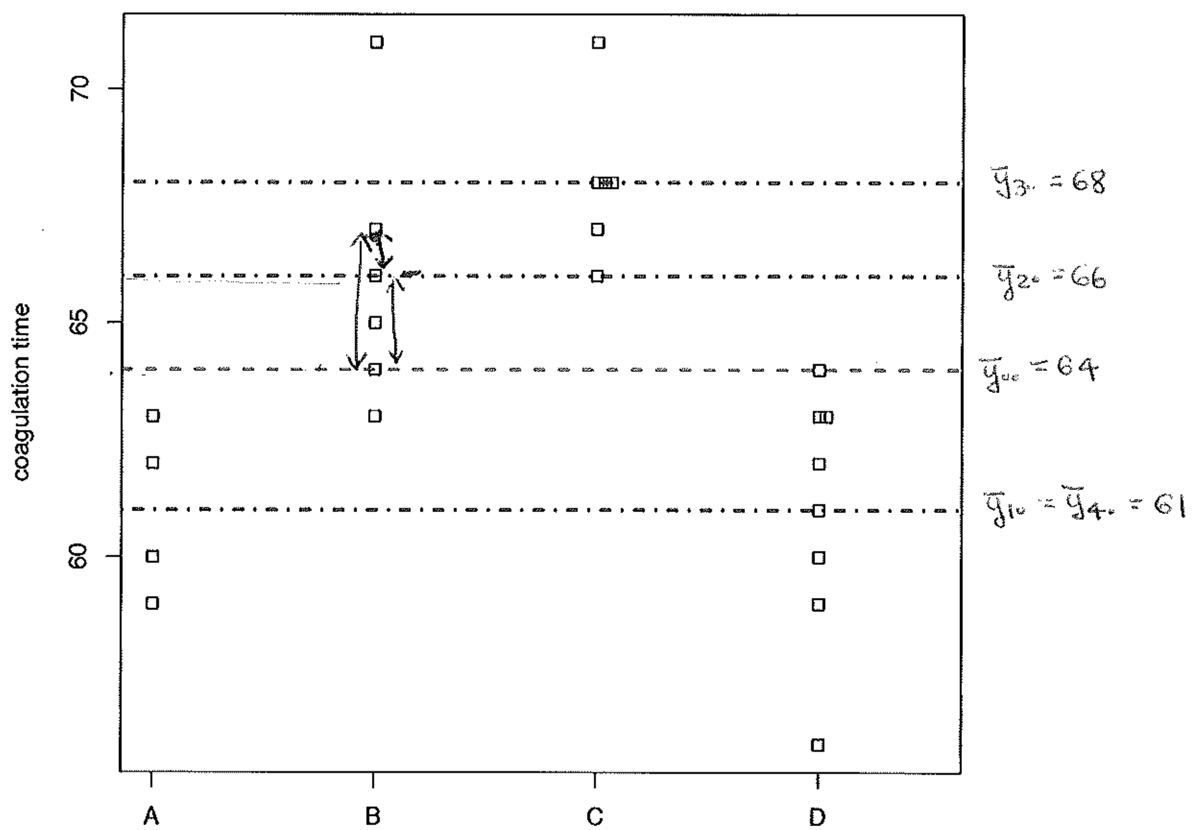
$$Y^T(I - P_0)Y = \|Y - P_0 Y\|^2 = \sum_i \sum_j (y_{ij} - \bar{y}_{..})^2 = 340$$

$$\underbrace{Y^T(I - P)Y}_{e(X)} = \|Y - PY\|^2 = \sum_i \sum_j (y_{ij} - \bar{y}_{i.})^2 = 112$$

$$F = \frac{\left(Y^T(I - P_0)Y - Y^T(I - P)Y\right) / r(P - P_0)}{Y^T(I - P)Y / (N - r)}$$

$$= \frac{(340 - 112) / 3}{112 / (24 - 4)} = 13.571$$

Coagulation Example– revisit



$$R^2 = \frac{\sum(\hat{y}_{ij} - \bar{y}_{\cdot\cdot})^2}{\sum(y_{ij} - \bar{y}_{\cdot\cdot})^2} = \frac{\sum(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2}{\sum(y_{ij} - \bar{y}_{\cdot\cdot})^2} = \cancel{\frac{\sum(y_{ij} - \bar{y}_{i\cdot})^2}{\sum(y_{ij} - \bar{y}_{\cdot\cdot})^2}}$$

$$= 1 - \frac{\sum(\bar{y}_{i\cdot} - y_{ij})^2}{\sum(y_{ij} - \bar{y}_{\cdot\cdot})^2} = 1 - \frac{112}{340} = 0.6706$$

$$y_{ij} = \mu + \alpha_i + e_{ij}$$

$$X' = \begin{bmatrix} 1_4 & 0_4 & \cdots & 0_4 \\ 1_6 & 1_6 & \vdots & 0_6 \\ 1_6 & 0_6 & \vdots & 0_6 \\ 1_8 & 0_8 & \cdots & 1_8 \end{bmatrix}$$

$$e(X') = e(X)$$

Coagulation Example– revisit

```
> options(contrasts=c("contr.sum", "contr.poly"))
> g2 <- lm(coag ~ diet, coagulation)
> summary(g2)

Call:
lm(formula = coag ~ diet, data = coagulation)

Residuals:
   Min     1Q Median     3Q    Max
 -5.00  -1.25   0.00   1.25   5.00

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  64.0000     0.4979 128.537  < 2e-16 ***
diet1        -3.0000     0.9736  -3.081 0.005889 **
diet2         2.0000     0.8453   2.366 0.028195 *
diet3         4.0000     0.8453   4.732 0.000128 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 2.366 on 20 degrees of freedom
Multiple R-squared:  0.6706, Adjusted R-squared:  0.6212
F-statistic: 13.57 on 3 and 20 DF,  p-value: 4.658e-05
```

$p = 5$

$r(X') = 4$

$\sum \alpha_i = 0$

$\hat{\mu} = \bar{y}_{\cdot\cdot}$

$\hat{\alpha}_i = \bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot}$

$\hat{y}_{ij} = \bar{y}_{i\cdot}$

$= \hat{\mu} + \hat{\alpha}_i$

Coagulation Example– revisit

```
> anova(g2)
Analysis of Variance Table

Response: coag
          Df Sum Sq Mean Sq F value    Pr(>F)
diet       3    228    76.0  13.571 4.658e-05 ***
Residuals 20    112     5.6
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

SSR/3

SSM

SSR

SSE/20

SSE

228 + 112 = 340

Economic Dataset Example

* Taken from Faraway's book (page 28).

* Consider an old economic dataset on 50 different countries.

* These data are averages from 1960 to 1970 to remove business cycle or other short-term fluctuation.

* Variables

X_3 `dpi`: per capita disposable income in US dollars

X_4 `ddpi`: the percentage rate of change in per capita disposable income

y `sr`: aggregate personal saving divided by disposable income.

X_1 `pop15` and X_2 `pop75`: the percentage of population under 15 and over 75, respectively

Economic Dataset Example (contd)

```
> rm(list=ls(all=TRUE))
> library(faraway)
> data(savings)
> g <- lm(sr ~ pop15 + pop75 + dpi + ddpi, savings)
> summary(g)
Call:
lm(formula = sr ~ pop15 + pop75 + dpi + ddpi, data = savings)

Residuals:
    Min      1Q  Median      3Q     Max
-8.2422 -2.6857 -0.2488  2.4280  9.7509

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 28.5660865  7.3545161   3.884 0.000334 ***
pop15       -0.4611931  0.1446422  -3.189 0.002603 **
pop75       -1.6914977  1.0835989  -1.561 0.125530
dpi         -0.0003369  0.0009311  -0.362 0.719173
ddpi         0.4096949  0.1961971   2.088 0.042471 *
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
Residual standard error: 3.803 on 45 degrees of freedom
Multiple R-squared:  0.3385, Adjusted R-squared:  0.2797
F-statistic: 5.756 on 4 and 45 DF,  p-value: 0.0007904
```

$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_4 \end{bmatrix}$ $\quad \dfrac{(\hat{\beta}_1 - 0)}{\sqrt{\hat{\sigma}^2 \lambda^T (X^T X)^{-1} \lambda}} \sim t$

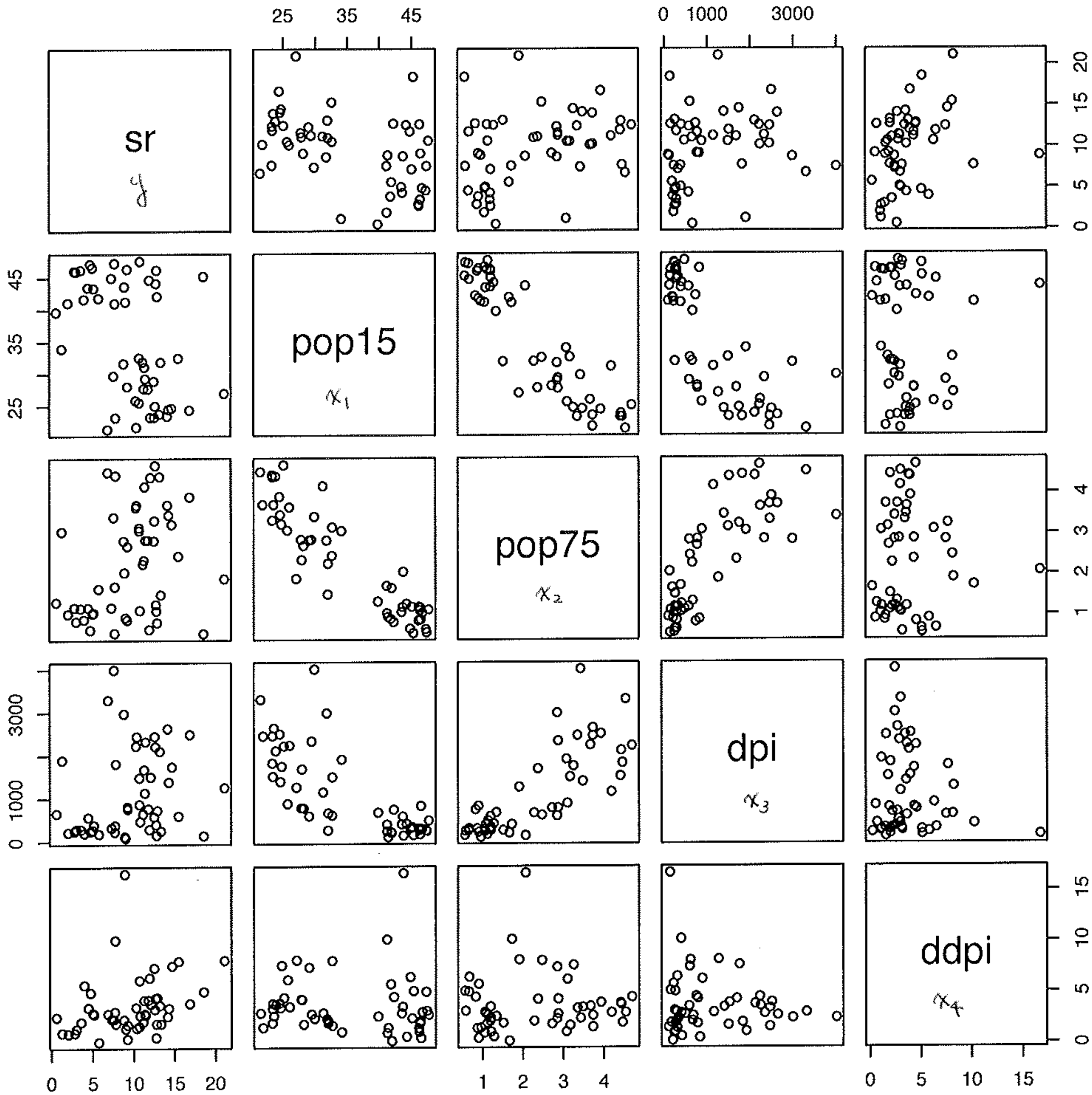
$\lambda = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$

$H_0: \beta_1 = 0$

$H_0: \beta_1 = \beta_1 = \beta_3 = \beta_4 = 0$ vs H_1: at least one $\beta_j \neq 0$

model w/ intercept.

$\Rightarrow$ reject H_0



sr
y
pop15
x_1
pop75
x_2
dpi
x_3
ddpi
x_4

Economic Dataset Example (contd)

```
> g2 <- lm(sr ~ pop75 + dpi + ddpi, savings)
> summary(g2)

Call:
lm(formula = sr ~ pop75 + dpi + ddpi, data = savings)

Residuals:
    Min      1Q  Median      3Q     Max
-8.0577 -3.2144  0.1687  2.4260 10.0763

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 5.4874944  1.4276619   3.844  0.00037 ***
pop75       0.9528574  0.7637455   1.248  0.21849
dpi         0.0001972  0.0010030   0.197  0.84499
ddpi        0.4737951  0.2137272   2.217  0.03162 *
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 4.164 on 46 degrees of freedom
Multiple R-squared:  0.189, Adjusted R-squared:  0.1361
F-statistic: 3.573 on 3 and 46 DF,  p-value: 0.02093
```

pop15

$\hat{\beta} = (X^T X)^{-1} X^T y$

Economic Dataset Example (contd)

w/o pop15   w/ pop15

```
> anova(g2, g)
Analysis of Variance Table

Model 1: sr ~ pop75 + dpi + ddpi
Model 2: sr ~ pop15 + pop75 + dpi + ddpi
  Res.Df    RSS Df Sum of Sq      F   Pr(>F)
1     46 797.72
2     45 650.71  1    147.01 10.167 0.002603 **
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
>
```

$(-3.189)^2$

Economic Dataset Example (contd)

```
> g3 <- lm(sr ~ pop15 + dpi, savings)
> summary(g3)
```

pop75 → ddpi

```
Call:
lm(formula = sr ~ pop15 + dpi, data = savings)

Residuals:
    Min      1Q  Median      3Q     Max
-8.1167 -2.6564 -0.0053  1.4831 10.9760

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) 22.7126386  4.1520364   5.470 1.69e-06 ***
pop15       -0.3303269  0.0949204  -3.480  0.00109 **
dpi         -0.0013107  0.0008767  -1.495  0.14159
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 3.979 on 47 degrees of freedom
Multiple R-squared:  0.2435, Adjusted R-squared:  0.2113
F-statistic: 7.564 on 2 and 47 DF,  p-value: 0.001419
```

Economic Dataset Example (contd)

```
> anova(g3, g)
Analysis of Variance Table

Model 1: sr ~ pop15 + dpi
Model 2: sr ~ pop15 + pop75 + dpi + ddpi
  Res.Df    RSS Df Sum of Sq      F  Pr(>F)
1     47 744.12
2     45 650.71  2    93.411 3.2299 0.04889 *
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

$\Lambda^T = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$ $\Lambda^T\beta = 0$

Economic Dataset Example (contd)

```
> gr <- lm(sr ~ I(pop15 + pop75) + dpi + ddpi, savings)
> summary(gr)
```

$\beta_1 = \beta_2$ $\quad y_i = \beta_0 + \beta_1(x_{1i} + x_{2i}) + \beta_3 x_{3i} + \beta_4 x_{4i} + e_i$

$\beta_1 x_{1i} + \beta_2 x_{2i}$

```
Call:
lm(formula = sr ~ I(pop15 + pop75) + dpi + ddpi, data = savings)

Residuals:
   Min     1Q Median     3Q    Max
-7.787 -2.767 -0.125  1.744 10.342

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)      21.6093051  4.8833633   4.425 5.87e-05 ***
I(pop15 + pop75) -0.3336331  0.1038679  -3.212  0.00241 **
dpi              -0.0008451  0.0008444  -1.001  0.32212
ddpi              0.3909649  0.1968714   1.986  0.05302 .
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 3.827 on 46 degrees of freedom
Multiple R-squared:  0.3152, Adjusted R-squared:  0.2705
F-statistic: 7.056 on 3 and 46 DF,  p-value: 0.0005328
```

Economic Dataset Example (contd)

```
> anova(gr, g)
Analysis of Variance Table

Model 1: sr ~ I(pop15 + pop75) + dpi + ddpi
Model 2: sr ~ pop15 + pop75 + dpi + ddpi
  Res.Df    RSS Df Sum of Sq      F Pr(>F)
1     46 673.63
2     45 650.71  1    22.915 1.5847 0.2146
```