# 다변량 분산분석 & 회귀분석
## MANOVA & Multivariate Reg.

H. Park

HUFS

# 단변량 t-test

$$x_i \sim N(\mu, \sigma^2), \qquad i = 1, 2, \cdots n$$

$$\rightarrow \quad \bar{x} \sim N(\mu, \sigma^2/n)$$

$\rightarrow$

$H_0 : \mu = \mu_0 \quad \rightarrow \quad z_0 =$

if $\sigma^2$ is unknown $\rightarrow$ $$t_0 = \sqrt{\frac{n(\bar{x} - \mu_0)^2}{s^2}} \sim t_{n-1}$$

$\rightarrow$

# 다변량 test

일반선형가정 $H_0 : A\boldsymbol{\mu} = \boldsymbol{b}$ $\longleftrightarrow$ $H_0 : A\boldsymbol{\mu} - \boldsymbol{b} = \mathbf{0}$

$$\boldsymbol{x}_i \sim N_p(\boldsymbol{\mu}, \Sigma) \longrightarrow A\boldsymbol{x}_i - \boldsymbol{b} \sim N_q(A\boldsymbol{\mu} - \boldsymbol{b}, A\Sigma A^T)$$

If $H_0$ is true, $\dfrac{n-q}{q(n-1)} T_0^2 \sim F_{q,n-q}$

$$T_0^2 = n(A\bar{\boldsymbol{x}} - \boldsymbol{b})^T (ASA^T)^{-1} (A\bar{\boldsymbol{x}} - \boldsymbol{b})$$

(e.g.) $\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \sim N_3\left( \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 \end{pmatrix} \right)$

$$H_0 : \begin{pmatrix} \mu_2 = 12\mu_1 + 2 \\ \mu_2 = 5\mu_3 \end{pmatrix} \longleftrightarrow \begin{pmatrix} 12 & -1 & 0 \\ 0 & 1 & -5 \end{pmatrix} \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{pmatrix} - \begin{pmatrix} -2 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

# example (sweat data) p.69

```
data b; set a;

v1=12*y1-y2+2;

v2=y2-5*y3;

proc glm;

model v1 v2 = /nouni;

manova h=INTERCEPT ;  run;
```

$$\frac{n-q}{q(n-1)}T_0^2 = 4.36$$

reject $H_0$

MANOVA Test Criteria and Exact F Statistics for the Hypothesis of No Overall Intercept Effect
H = Type III SSCP Matrix for Intercept
E = Error SSCP Matrix

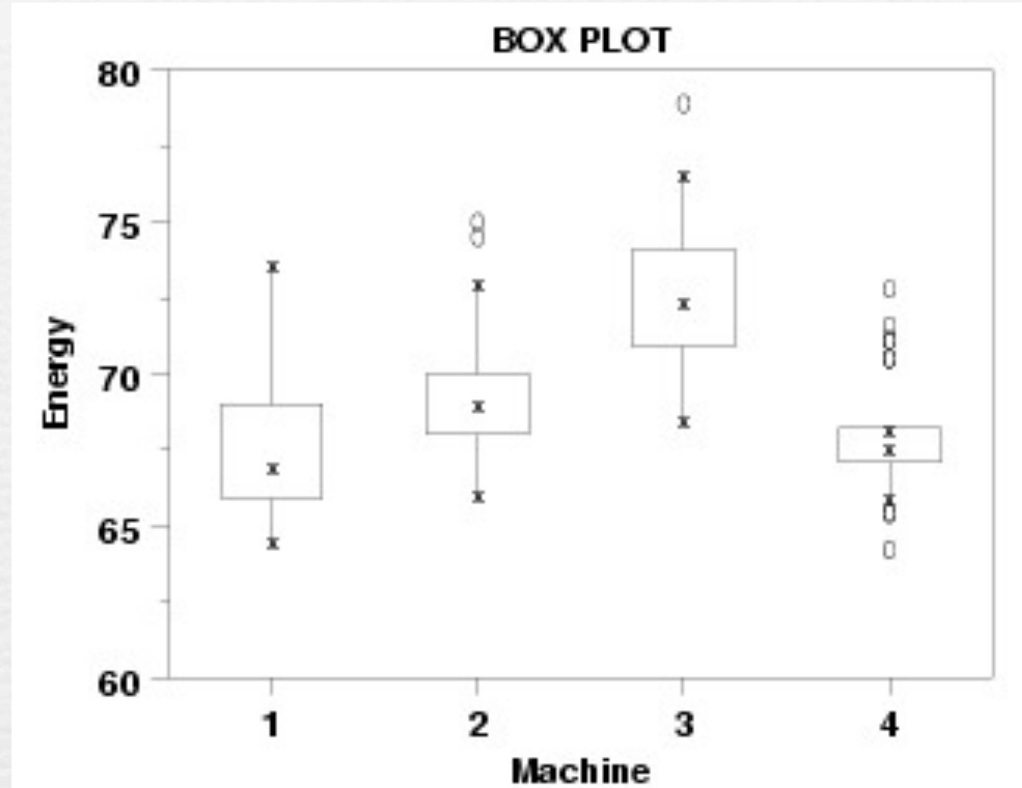
S=1 M=0 N=8

| Statistic | Value | F Value | Num DF | Den DF | Pr > F |
|---|---|---|---|---|---|
| Wilks' Lambda | 0.67374574 | 4.36 | 2 | 18 | 0.0286 |
| Pillai's Trace | 0.32625426 | 4.36 | 2 | 18 | 0.0286 |
| Hotelling-Lawley Trace | 0.48423944 | 4.36 | 2 | 18 | 0.0286 |
| Roy's Greatest Root | 0.48423944 | 4.36 | 2 | 18 | 0.0286 |

# 단변량 일원배치법(ANOVA)

$$
\begin{aligned}
y_{ij} &= \text{ } ith \text{ group and } jth \text{ observation} \\
&= \qquad
\end{aligned}
$$

$$\epsilon_{ij} \sim \ i.i.d. \ N(0, \sigma^2)$$




$$\sum_i^k \sum_j^{n_i} (y_{ij} - \bar{y}_{..})^2 = \sum_i^k \sum_j^{n_i} \qquad + \sum_i^k \sum_j^{n_i} \qquad$$

$$
\begin{aligned}
SST &= SSB + SSE \\
SS(total) &= SS(\qquad) + SS(\qquad)
\end{aligned}
$$

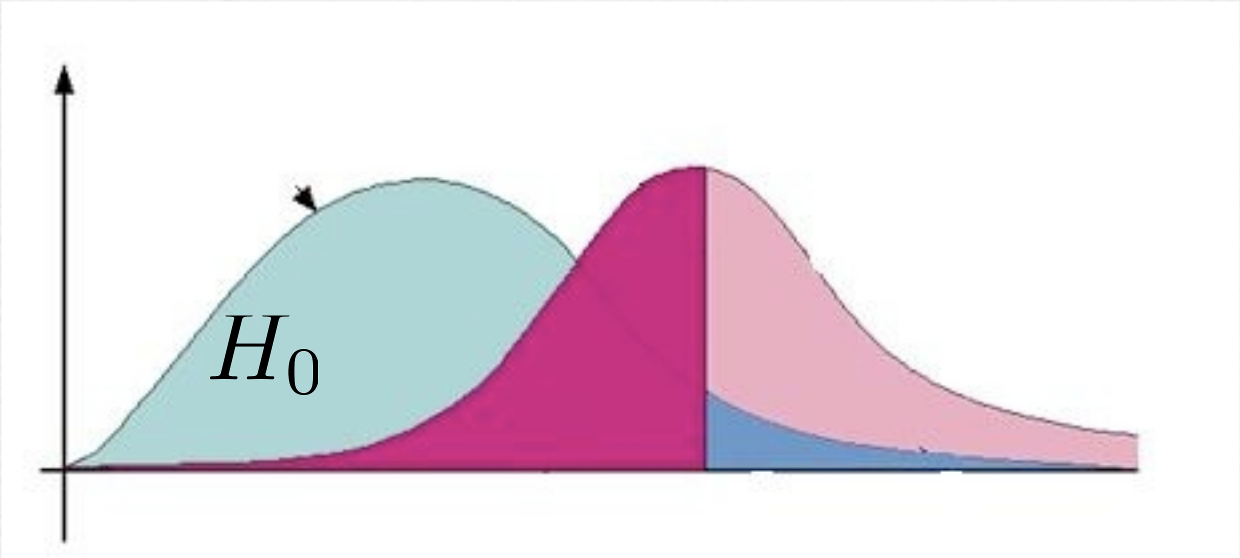
# 분산분석표(ANOVA table)  $n_1=n_2= \ldots =n_k$

| | d.f. | S.S. | M.S. | $F_0$ |
|---|---|---|---|---|
| | k-1 | SSB | MSB | |
| | N-k | SSE | MSE | |
| total | N-1 | SST | | |

$H_0$ :

If $H_0$ is true, $F_0 \sim F_{k-1,N-k}$

If $F_0 > F_{k-1,N-k,\alpha}$, reject $H_0$



# 다변량 일원배치법(MANOVA)

$$\begin{pmatrix} y_{ij,h} \\ y_{ij,w} \\ y_{ij,c} \end{pmatrix} = \begin{pmatrix} \mu_h \\ \mu_w \\ \mu_c \end{pmatrix} + \begin{pmatrix} \tau_{i,h} \\ \tau_{i,w} \\ \tau_{i,c} \end{pmatrix} + \begin{pmatrix} \epsilon_{ij,h} \\ \epsilon_{ij,w} \\ \epsilon_{ij,c} \end{pmatrix}$$

$$\sum_i^k \sum_j^{n_i} (\boldsymbol{y}_{ij} - \bar{\boldsymbol{y}})(\boldsymbol{y}_{ij} - \bar{\boldsymbol{y}})^T \longleftarrow \text{SSCP matrix}$$

$$= \sum_i^k \sum_j^{n_i} (\bar{\boldsymbol{y}_{i.}} - \bar{\boldsymbol{y}})(\bar{\boldsymbol{y}_{i.}} - \bar{\boldsymbol{y}})^T + \sum_i^k \sum_j^{n_i} (\boldsymbol{y}_{ij} - \bar{\boldsymbol{y}_{i.}})(\boldsymbol{y}_{ij} - \bar{\boldsymbol{y}_{i.}})^T$$

$$\boxed{T_{3\times 3} = B_{3\times 3} + W_{3\times 3}}$$

# MANOVA(Multivariate ANOVA)

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

$$\begin{pmatrix} \tau_{1,h} \\ \tau_{1,w} \\ \tau_{1,c} \end{pmatrix} = \begin{pmatrix} \tau_{2,h} \\ \tau_{2,w} \\ \tau_{2,c} \end{pmatrix} = \cdots = \begin{pmatrix} \tau_{k,h} \\ \tau_{k,w} \\ \tau_{k,c} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

SSCP 중에서 귀무가설 기각

$$\Lambda = \frac{det(W)}{det(B+W)} \longrightarrow$$ $H_0$ 기각 (Wilk's Lambda)

$$tr(B(B+E)^{-1}) \longrightarrow$$ $H_0$ 기각 (Pillai's trace)

$$tr(BE^{-1}) \longrightarrow$$ $H_0$ 기각 (Hotelling-Lawley's trace)

$$\max(\text{e-values of } BE^{-1}) \longrightarrow$$ $H_0$ 기각 (Roy's greatest root)

# MANOVA TABLE

| Statistics | value | F | p-value |
|---|---|---|---|
| Wilk's Lambda | $\frac{det(W)}{det(B+W)}$ | 1.64 | 0.16 |
| Pillai's Trace | $tr(B(B+W)^{-1})$ | 1.53 | 0.19 |
| Hotelling-Lawley's Trace | $tr(BW^{-1})$ | 1.73 | 0.14 |
| Roy's Max. Root | $\max(e-values\ of\ BW^{-1})$ | 3.7 | 0.02 |

# MANOVA in SAS

```
proc glm;
  class group;


run;
```

# Linear Regression (선형회귀)

- y = dependent variable,   x=independent variable
- simple linear regression :   one  y  &  one  x (회귀1)
  (단순선형회귀)
- multiple linear regression :  one y  & multiple x's (회귀2)
  (중회귀/다중선형회귀)
- multivariate linear regression :  multiple y's & x's (다변량)
  (다변량 선형회귀)

# example (RDA)

- RDA(Recommended Dietary Allowances): dietary standards adequate to meet the nutritional needs for healthy people
- $y_I$ = Iron intake (a percentage of RDA)
- $y_P$= Protein intake (a percentage of RDA)
- $y_C$= Calcium intake (a percentage of RDA)
- $x_1$= poverty index
- $x_2$ = education level of a family head
- $x_3$ = single parent

# 단변량 회귀분석

$$y_{iI} \quad = \quad \beta_{0I} + \beta_{1I}x_{1i} + \beta_{2I}x_{2i} + \beta_{3I}x_{3i} + \epsilon_{iI}$$

$$\begin{pmatrix} y_{1I} \\ y_{2I} \\ \vdots \\ y_{nI} \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{21} & x_{31} \\ 1 & x_{12} & x_{22} & x_{32} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1n} & x_{2n} & x_{3n} \end{pmatrix} \begin{pmatrix} \beta_{0I} \\ \beta_{1I} \\ \beta_{2I} \\ \beta_{3I} \end{pmatrix} + \begin{pmatrix} \epsilon_{1I} \\ \epsilon_{2I} \\ \vdots \\ \epsilon_{nI} \end{pmatrix}$$

$$y_{iP} \quad = \quad \beta_{0P} + \beta_{1P}x_{1i} + \beta_{2P}x_{2i} + \beta_{3P}x_{3i} + \epsilon_{iP}$$

$$\begin{pmatrix} y_{1P} \\ y_{2P} \\ \vdots \\ y_{nP} \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{21} & x_{31} \\ 1 & x_{12} & x_{22} & x_{32} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1n} & x_{2n} & x_{3n} \end{pmatrix} \begin{pmatrix} \beta_{0P} \\ \beta_{1P} \\ \beta_{2P} \\ \beta_{3P} \end{pmatrix} + \begin{pmatrix} \epsilon_{1P} \\ \epsilon_{2P} \\ \vdots \\ \epsilon_{nP} \end{pmatrix}$$

# 다변량 회귀모형

$$
\begin{aligned}
y_{iI} &= \beta_{0I} + \beta_{1I}x_{1i} + \beta_{2I}x_{2i} + \beta_{3I}x_{3i} + \epsilon_{iI} \\
y_{iP} &= \beta_{0P} + \beta_{1P}x_{1i} + \beta_{2P}x_{2i} + \beta_{3P}x_{3i} + \epsilon_{iP} \\
y_{iC} &= \beta_{0C} + \beta_{1C}x_{1i} + \beta_{2C}x_{2i} + \beta_{3C}x_{3i} + \epsilon_{iC}
\end{aligned}
$$

$$
\left( \qquad \qquad \right) = X \left( \qquad \qquad \right) + \left( \boldsymbol{\epsilon}_I \quad \boldsymbol{\epsilon}_P \quad \boldsymbol{\epsilon}_C \right)
$$

$$
\begin{pmatrix} y_{1I} & y_{1P} & y_{1C} \\ y_{2I} & y_{2P} & y_{2C} \\ \vdots & \vdots & \vdots \\ y_{nI} & y_{nP} & y_{nC} \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{21} & x_{31} \\ 1 & x_{12} & x_{22} & x_{32} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1n} & x_{2n} & x_{3n} \end{pmatrix} \begin{pmatrix} \beta_{0I} & \beta_{0P} & \beta_{0C} \\ \beta_{1I} & \beta_{1P} & \beta_{1C} \\ \beta_{2I} & \beta_{2P} & \beta_{2C} \\ \beta_{3I} & \beta_{3P} & \beta_{3C} \end{pmatrix} + \begin{pmatrix} \epsilon_{1I} & \epsilon_{1P} & \epsilon_{1C} \\ \epsilon_{2I} & \epsilon_{2P} & \epsilon_{2C} \\ \vdots & \vdots & \vdots \\ \epsilon_{nI} & \epsilon_{nP} & \epsilon_{nC} \end{pmatrix}
$$

# 다변량회귀 in SAS

```
proc reg;
    model yI yP yC = x1 x2 x3;
    mtest x1, x2;
    mtest yI-yP, x1;
    mtest yI-yP, yP-yC ;
run;
```

$$H_0 : \begin{pmatrix} \beta_{1I} = \beta_{1P} = \beta_{1C} = 0 \\ \beta_{2I} = \beta_{2P} = \beta_{2C} = 0 \end{pmatrix}$$

$$H_0 : \beta_{1I} = \beta_{1P}$$

$$H_0 : \begin{pmatrix} \beta_{1I} = \beta_{1P} = \beta_{1C} \\ \beta_{2I} = \beta_{2P} = \beta_{2C} \\ \beta_{3I} = \beta_{3P} = \beta_{3C} \end{pmatrix}$$

$$\begin{pmatrix} y_{1I} & y_{1P} & y_{1C} \\ y_{2I} & y_{2P} & y_{2C} \\ \vdots & \vdots & \vdots \\ y_{nI} & y_{nP} & y_{nC} \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{21} & x_{31} \\ 1 & x_{12} & x_{22} & x_{32} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1n} & x_{2n} & x_{3n} \end{pmatrix} \begin{pmatrix} \beta_{0I} & \beta_{0P} & \beta_{0C} \\ \beta_{1I} & \beta_{1P} & \beta_{1C} \\ \beta_{2I} & \beta_{2P} & \beta_{2C} \\ \beta_{3I} & \beta_{3P} & \beta_{3C} \end{pmatrix} + \begin{pmatrix} \epsilon_{1I} & \epsilon_{1P} & \epsilon_{1C} \\ \epsilon_{2I} & \epsilon_{2P} & \epsilon_{2C} \\ \vdots & \vdots & \vdots \\ \epsilon_{nI} & \epsilon_{nP} & \epsilon_{nC} \end{pmatrix}$$

```
data a;  input yI yP yC x1  x2 x3;  cards;
56 63 75 85 3 0
85 73 69 65 3 0
34 56 87 65 1 0
23 14 55 28 2 1
37 44 29 54 1 1
21 17 23 74 2 1
12 32 10 78 1 1
88 78 23 48 3 0
58 57 63 42 1 0
23 52 31 78 1 0
;
proc reg ;
    model  yI  yP  yC = x1  x2  x3;
    mtest x1, x2, x3 ;
    mtest yI-yP, yP-yC;
 run;
```

$H_0 : \left( \qquad\qquad \right)$

$H_0 : \left( \qquad\qquad \right)$

$y_I$

Parameter Estimates

| Variable | DF | Parameter Estimate | Standard Error | t Value | Pr > \|t\| |
|---|---|---|---|---|---|
| Intercept | 1 | 53.56609 | 22.48920 | 2.38 | 0.0546 |
| x1 | 1 | -0.43615 | 0.28278 | -1.54 | 0.1739 |
| x2 | 1 | 15.80396 | 5.80497 | 2.72 | 0.0345 |
| x3 | 1 | -28.50747 | 10.44446 | -2.73 | 0.0342 |

$y_P$

Parameter Estimates

| Variable | DF | Parameter Estimate | Standard Error | t Value | Pr > \|t\| |
|---|---|---|---|---|---|
| Intercept | 1 | 55.34983 | 19.03019 | 2.91 | 0.0270 |
| x1 | 1 | 0.00381 | 0.23929 | 0.02 | 0.9878 |
| x2 | 1 | 3.78680 | 4.91212 | 0.77 | 0.4700 |
| x3 | 1 | -34.50294 | 8.83802 | -3.90 | 0.0079 |

$y_C$

Parameter Estimates

| Variable | DF | Parameter Estimate | Standard Error | t Value | Pr > \|t\| |
|---|---|---|---|---|---|
| Intercept | 1 | 73.59135 | 38.22958 | 1.92 | 0.1026 |
| x1 | 1 | -0.26648 | 0.48070 | -0.55 | 0.5994 |
| x2 | 1 | 0.70958 | 9.86792 | 0.07 | 0.9450 |
| x3 | 1 | -29.81646 | 17.75462 | -1.68 | 0.1441 |

$Y$

The REG Procedure
Model: MODEL1
Multivariate Test 1

Multivariate Statistics and F Approximations

S=3 M=-0.5 N=1

| Statistic | Value | F Value | Num DF | Den DF | Pr > F |
|---|---|---|---|---|---|
| Wilks' Lambda | 0.02417486 | 3.97 | 9 | 9.8856 | 0.0217 |
| Pillai's Trace | 1.72073659 | 2.69 | 9 | 18 | 0.0353 |
| Hotelling-Lawley Trace | 10.65843953 | 4.74 | 9 | 4 | 0.0742 |
| Roy's Greatest Root | 5.40157533 | 10.80 | 3 | 6 | 0.0078 |

NOTE: F Statistic for Roy's Greatest Root is an upper bound.