§9.4 2 sample Hotelling's T^2 test

10) 1 test assumes $\Sigma_{\underline{x}_1} = \Sigma_{\underline{x}_2}$, and is equivalent to a MANOVA test of ch 10, and is the multivariate analog of the pooled t test.

11) Another test does not assume $\Sigma_{\underline{x}_1} = \Sigma_{\underline{x}_2}$ and is the multivariate analog of the 2 sample t test.

There are 2 ind random samples $x_{1,1}, \ldots, x_{n_1,1}$ and $x_{1,2}, \ldots, x_{n_2,2}$ from populations with mean and covariance matrices $(\underline{\mu}_i, \Sigma_{\underline{x}_i})$ for $i=1,2$. For now, assume $n_1 = k\, n_2$ for some positive real number k.

By MCLT,
$$\begin{pmatrix} \sqrt{n_1}\,(\overline{\underline{x}}_1 - \underline{\mu}_1) \\ \sqrt{n_2}\,(\overline{\underline{x}}_2 - \underline{\mu}_2) \end{pmatrix} \xrightarrow{D} N_{2p}\left[\begin{pmatrix} \underline{0} \\ \underline{0} \end{pmatrix}, \begin{pmatrix} \Sigma_{\underline{x}_1} & 0 \\ 0 & \Sigma_{\underline{x}_2} \end{pmatrix}\right]$$

or
$$\begin{pmatrix} \sqrt{n_2}\,(\overline{\underline{x}}_1 - \underline{\mu}_1) \\ \sqrt{n_2}\,(\overline{\underline{x}}_2 - \underline{\mu}_2) \end{pmatrix} \xrightarrow{D} N_{2p}\left[\begin{pmatrix} \underline{0} \\ \underline{0} \end{pmatrix}, \begin{pmatrix} \frac{\Sigma_{\underline{x}_1}}{k} & 0 \\ 0 & \Sigma_{\underline{x}_2} \end{pmatrix}\right].$$

So
$$\sqrt{n_2}\left[(\overline{\underline{x}}_1 - \overline{\underline{x}}_2) - (\underline{\mu}_1 - \underline{\mu}_2)\right] \xrightarrow{D} N_p\left(\underline{0},\ \frac{\Sigma_{\underline{x}_1}}{k} + \Sigma_{\underline{x}_2}\right).$$

If $\underline{\mu}_1 = \underline{\mu}_2$, then
$$n_2\,(\overline{\underline{x}}_1 - \overline{\underline{x}}_2)^T \left(\frac{\Sigma_{\underline{x}_1}}{k} + \Sigma_{\underline{x}_2}\right)^{-1} (\overline{\underline{x}}_1 - \overline{\underline{x}}_2) \xrightarrow{D} \chi^2_p$$

or
$$(\overline{\underline{x}}_1 - \overline{\underline{x}}_2)^T \left(\frac{\Sigma_{\underline{x}_1}}{n_1} + \frac{\Sigma_{\underline{x}_2}}{n_2}\right)^{-1} (\overline{\underline{x}}_1 - \overline{\underline{x}}_2) \xrightarrow{D} \chi^2_p.$$

k disappears

37.5

So if $\underline{\mu}_1 = \underline{\mu}_2$, $T_0^2 = (\underline{\bar{X}}_1 - \underline{\bar{X}}_2)^T \left(\frac{S_1}{n_1} + \frac{S_2}{n_2}\right)^{-1} (\underline{\bar{X}}_1 - \underline{\bar{X}}_2) \xrightarrow{D} \chi^2_p$

12) Let $d_n = \min(n_1 - p,\ n_2 - p)$. Two sample Hotelling's T^2 test

i) H_0 $\underline{\mu}_1 = \underline{\mu}_2$ $\quad H_A$ $\underline{\mu}_1 \neq \underline{\mu}_2$

ii) Find $t_0 = \frac{T_0^2}{p}$

iii) pval $= P(t_0 < F_{p, d_n})$

iv) reject H_0 if pval $< \alpha$ and conclude $\underline{\mu}_1 \neq \underline{\mu}_2$

fail to reject H_0 if pval $\geq \alpha$ and conclude $\underline{\mu}_1 = \underline{\mu}_2$ or that there is not enough evidence to conclude that $\underline{\mu}_1 \neq \underline{\mu}_2$. If possible give a nontechnical sentence.

ex] (P 230) Samples of $n_1 = 45$, $n_2 = 55$ Wisconsin home owners with and without air conditioning

electrical consumption during July 1977

X_1 = on peak electrical consumption

X_2 = off peak

$\underline{\bar{X}}_1 = \begin{pmatrix} 204.4 \\ 556.6 \end{pmatrix}$, $\underline{\bar{X}}_2 = \begin{pmatrix} 130 \\ 355 \end{pmatrix}$, $T_0^2 = 15.66$. Test $\underline{\mu}_1 = \underline{\mu}_2$

i) H_0 $\underline{\mu}_1 = \underline{\mu}_2$ $\quad H_1$ $\underline{\mu}_1 \neq \underline{\mu}_2$

ii) $t_0 = T_0^2/p = 15.66/2 = 7.83$

iii) pval $= P(7.83 < F_{\underset{p}{2},\ \underset{\min(45-2,\ 55-2)}{43}}) < 0.05$

	2
30	3.22
∞	3.00

iv) reject H_0 the mean electrical consumptions for those with and without AC are different!

ex] (p185) $n=87$ $p=3$ $X_3 =$ CLEP Science score

$\overline{X_3} = 25.13$, $s_3^2 = 23.11$ Find

a 95% Simultaneous large sample CI for μ_3

Soln $F_{p, n-p, 0.95} = F_{3, 84, .95} \approx 2.61$

$k = 84 > 70$

	3
30	2.92 ← use for den df k with $30 \le k \le 70$
∞	2.61 ← use for den df $k > 70$

CI is $\overline{X_3} \pm \sqrt{\frac{p(n-1)}{n-p} F_{p, n-p, .95}} \sqrt{\frac{s_3^2}{n}}$

$= 25.13 \pm \sqrt{\frac{3(86)}{84} 2.61} \sqrt{\frac{23.11}{87}}$

$= 25.13 \pm \sqrt{8.016} \sqrt{\frac{23.11}{87}}$

$= 25.13 \pm 1.459 = \boxed{(23.671, 26.589)}$

Ch 10 MANOVA

$x = \mathbb{X}$

1] The (univariate) linear model

is $Y_i = \underset{\sim}{x}_i^T \underset{\sim}{\beta} + e_i = \underset{\sim}{\beta}^T \underset{\sim}{x}_i + \underset{\sim}{e}_i$, $i = 1, \ldots, n$.

In matrix form, the linear model is

$$\underset{n \times 1}{\underset{\sim}{Y}} = \underset{n \times p}{\mathbb{X}} \underset{p \times 1}{\underset{\sim}{\beta}} + \underset{n \times 1}{\underset{\sim}{e}}.$$

Data matrix $\begin{pmatrix} Y_1 & x_{11} & \cdots & x_{1p} \\ \vdots & & & \\ Y_n & x_{n1} & \cdots & x_{np} \end{pmatrix}$.

M584

Math 484 EDPSYCH 507, 508 MLR, DOE = ANOVA

2) Ch10,12 MANOVA multivariate linear regression

The multivariate linear model is

$$\underset{m\times 1}{\underset{\sim}{y}_i} = \underset{p\times m}{B^T}\,\underset{\sim}{x}_i + \underset{m\times 1}{\underset{\sim}{\varepsilon}_i}, \quad i=1,\dots,n,$$

$$\underset{\sim}{y}_i = \underbrace{(Y_{1i}, Y_{2i}, \dots, Y_{mi})^T}_{m \text{ response variables of interest}}, \quad \underset{\sim}{x}_i = \underbrace{(x_{1i}, \dots, x_{pi})^T}_{p \text{ explanatory variables}},$$

$$\underset{\sim}{\varepsilon}_i = (\varepsilon_{1i}, \dots, \varepsilon_{mi})^T, \quad B = (\underset{\sim}{\beta}_1, \dots, \underset{\sim}{\beta}_m).$$

In matrix form, $Z_1 = X B + E$.

Data matrix

$$\begin{array}{c} \\ \left[\begin{array}{ccc} Y_{11} & \cdots & Y_{1m} \\ \vdots & & \\ Y_{n1} & \cdots & Y_{nm} \end{array}\right. \end{array}
\begin{array}{c} \text{often omitted if a column of 1s} \\ \left.\begin{array}{c} x_{11}, \dots, x_{1p} \\ \\ x_{n1}, \dots, x_{np} \end{array}\right] \end{array}
\begin{array}{l} \\ \text{1st case} \\ \\ \text{nth case} \end{array}$$

$$\underbrace{Y_{n1} \cdots Y_{nm}}_{Z_1} \qquad \underbrace{x_{n1}, \dots, x_{np}}_{X}$$

mth response variable X_1 X_p = pth predictor variable

$$Y_{ji} = \underset{\sim}{x}_i^T \underset{\sim}{\beta}_j + e_{ji}$$

$\underset{\sim}{Y}_j = X \underline{\beta_j} + \underset{\sim}{e}_j$ follows a univariate linear model

$j=1,\dots,m.$ $\quad \mathrm{Cov}(\underset{\sim}{\varepsilon}_i) = \underset{m\times m}{\Sigma_\varepsilon} = (\sigma_{ij})$;

$\mathrm{Cov}(\underline{e_i}, \underline{e_j}) = \sigma_{ij} I_n, \quad i,j = 1,\dots,m.$

3) The one way MANOVA model has $p \geq 2$ groups.

$Y_1, \ldots, Y_m$ are measured on each group.

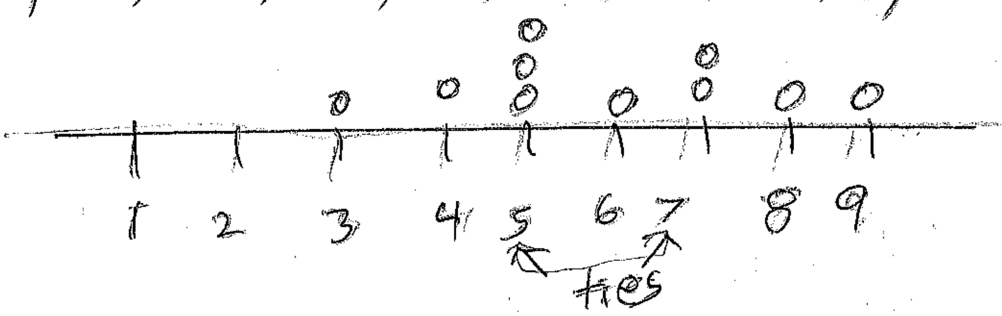
want to test H_0: $\underset{m\times 1}{\underline{\mu}_1} = \cdots = \underset{m\times 1}{\underline{\mu}_p}$.

The model assumes the p groups have the same covariance matrix $\underbrace{\Sigma_1 = \cdots = \Sigma_p = \Sigma}_{\text{bad assumption}}$ but possibly different means $\underline{\mu}_i$.

4) plots can be used to check linearity.

For a univariate sample, a dot plot is a plot of $y_1, \ldots, y_n$.

ex] 5, 5, 7, 9, 3, 4, 6, 8, 7, 5



5) $\underset{\sim}{Y}_k = X \underset{\sim}{\beta}_k + \underset{\sim}{e}_k$, $k = 1, \ldots, m$ each follow a

one way Anova model $\underbrace{\overset{\text{trt 1}}{\mu_{1k}}, \ldots, \overset{\text{trt p}}{\mu_{pk}}}_{\text{p groups}}$ k-th variable

$i = 1, \ldots, p$

write Y's as Y_{ijk} (group, case, resp variable) $j = 1, \ldots, n_i$, $k = 1, \ldots, m$

$\bar{Y}_{1ok}, \ldots, \bar{Y}_{pok}$ } sample of size n_i

ex) $m=2$, $p=4$ groups: fat for brands 1, 2, 3, 4; cholesterol "

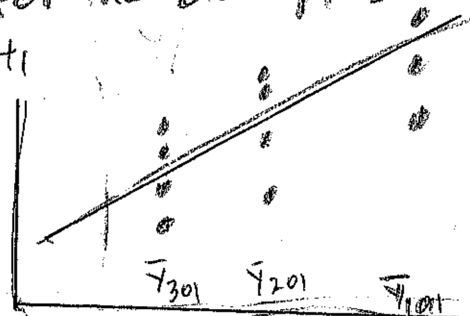
6) Make DD plot of residuals $\hat{\underset{\sim}{\epsilon}}_i$ $(m \times 1)$, $i=1,\ldots,n$ — $\underset{\sim}{\epsilon}_i$ might be MVN

7) For each response variable $\underset{\sim}{Y}_1, \ldots, \underset{\sim}{Y}_m$, make a response plot and a residual plot. The response plot is a plot of $\hat{Y}_{ijk}$ vs Y_{ijk}, $j=1,\ldots,n_i$, where $\hat{Y}_{ijk} = \overline{Y}_{i0k}$ if Y_{ijk} is in the ith group. Hence the response plot has m dot plots with n_j cases at $\overline{Y}_{i0k}$, $k=1,\ldots,m$.

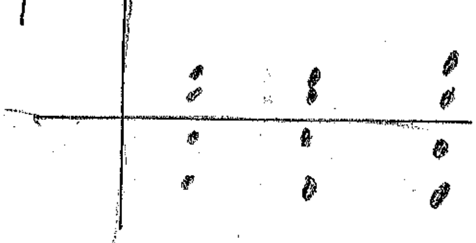
These dot plots should scatter about the identity line. The residual plot is a plot of $\hat{Y}_{ijk} = \overline{Y}_{i0k}$ vs the residuals $r_{ijk} = Y_{ijk} - \overline{Y}_{i0k}$, which will give p dot plots scattering about the $r=0$ line. want $\frac{\text{max range}}{\text{min range}} < 2$ for the dot plots.

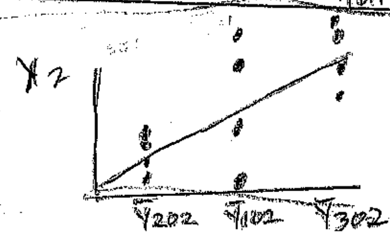
resp variable: data m

$\underset{\sim}{Y}_1$	$\ldots$	$\underset{\sim}{Y}_m$	group
Y_{111}		Y_{11m}	1
$\vdots$		$\vdots$	
Y_{1n_11}		Y_{1n_1m}	
Y_{211}		Y_{21m}	2
$\vdots$		$\vdots$	
Y_{2n_21}		Y_{2n_2m}	
$\vdots$		$\vdots$	
Y_{p11}		Y_{p1m}	p
$\vdots$		$\vdots$	
Y_{pn_p1}		Y_{pn_pm}	

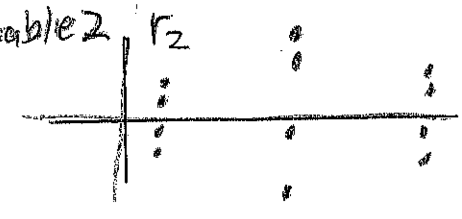


variable 1, r_1





variable 2, r_2



p234 Forch12

MU 140

8] Want to test $LB = 0$

where L is of full rank $r \times p$.

Ch12 not

Usual test statistic uses

$$H = \hat{B}^T L^T \left(L (X^T X)^{-1} L^T \right)^{-1} L \hat{B}$$

and $W_e = \hat{E}^T \hat{E}$ where $\frac{W_e}{n-p} = \hat{\Sigma}_\epsilon$.

Let $\lambda_1 \geq \cdots \geq \lambda_m$ be the eigenvalues of $W_e^{-1} H$.

lambda

MANOVA has similar tests, but need to be careful about which B matrix is used.

i) Wilk's Λ statistic $\Lambda(L) = |(H + W_e)^{-1} W_e|$

$$= |W_e^{-1} H + I|^{-1} = \prod_{i=1}^{m} (1+\lambda_i)^{-1}$$

ii) Pillai's trace statistic $V(L) = tr[(H+W_e)^{-1} H]$

$$= \sum_{i=1}^{m} \frac{\lambda_i}{1+\lambda_i}$$

iii) Hotelling-Lawley trace statistic $U(L) = tr(W_e^{-1} H)$

$$= \sum_{i=1}^{m} \lambda_i$$

poor

iv) Roy's maximum root statistic $\lambda_{max}(L) = \lambda_1$.

9] Use $L = \begin{array}{c} \quad\quad\;\; \text{jth position} \\ [0 \cdots 0\ 1\ 0 \cdots 0] \end{array}$ to test if X_j is needed in the model given the other $p-1$ predictors are in the model.

Use $L = [0 \;\; I_k]$ to test if the last k predictors are needed in the model given the 1st $p-k$ predictors are in the model.

use $L = \begin{bmatrix} 0 & 1 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix}$ to test if X_2 and X_p are needed in the model given the remaining $p-2$ predictors are in the model.

10) The Hotelling-Lawley test uses

$$\begin{pmatrix} \hat{\underset{\sim}{\beta}}_1 \\ \vdots \\ \hat{\underset{\sim}{\beta}}_m \end{pmatrix} - \begin{pmatrix} \underset{\sim}{\beta}_1 \\ \vdots \\ \underset{\sim}{\beta}_m \end{pmatrix} \approx N_{pm}\left(\underset{\sim}{0}, \begin{pmatrix} D_{11} & \cdots & D_{1m} \\ \vdots & & \\ D_{m1} & \cdots & D_{mm} \end{pmatrix} \right)$$

where $D_{ij} = \sigma_{ij} (\underset{\sim}{X}^T \underset{\sim}{X})^{-1}$.

$$\underbrace{\begin{bmatrix} L & & \\ & \ddots & \\ & & L \end{bmatrix}}_{rm \times mp} \left[\begin{pmatrix} \hat{\underset{\sim}{\beta}}_1 \\ \vdots \\ \hat{\underset{\sim}{\beta}}_m \end{pmatrix} - \begin{pmatrix} \underset{\sim}{\beta}_1 \\ \vdots \\ \underset{\sim}{\beta}_m \end{pmatrix} \right] \sim N_{rm}\left[\underset{\sim}{0}, \underbrace{\begin{pmatrix} L D_{11} L^T & \cdots & L D_{1m} L^T \\ \vdots & & \\ L D_{m1} L^T & \cdots & L D_{mm} L^T \end{pmatrix}}_{W} \right]$$

$$LHS = \underbrace{\begin{pmatrix} L\hat{\underset{\sim}{\beta}}_1 \\ \vdots \\ L\hat{\underset{\sim}{\beta}}_m \end{pmatrix}}_{rm \times 1} - \underbrace{\begin{pmatrix} L\beta_1 \\ \vdots \\ L\beta_m \end{pmatrix}}_{\underset{\sim}{0} \text{ under } H_0}$$

$$\text{so } T = (L\hat{\underset{\sim}{\beta}}_1 \cdots L\hat{\underset{\sim}{\beta}}_m)^T \hat{W}^{-1} \begin{pmatrix} L\hat{\underset{\sim}{\beta}}_1 \\ \vdots \\ L\hat{\underset{\sim}{\beta}}_m \end{pmatrix} \xrightarrow{D} \chi^2_{rm}.$$

Also all subsets of MVN are MVN, so figure out the subset corresponding to $\begin{pmatrix} L\hat{\underset{\sim}{\beta}}_1 \\ \vdots \\ L\hat{\underset{\sim}{\beta}}_m \end{pmatrix}$ and the corresponding covariance matrix.