III. Multivariate Statistical Inference

Why use a multivariate approach when conducting tests on p variables?

1. Type I error protection

   $\boxed{EX}$ $p = 10$ univariate tests at $\alpha = .05$

   If variables are independent,

   $$\begin{aligned} &\Pr\{\text{at least one rejection}\} \\ &\quad = 1 - \Pr\{\text{all 10 tests ``accept''}\} \\ &\quad = 1 - (.95)^{10} \cong .40 \end{aligned}$$

   In practice, the overall ("experimentwise") Type I error rate will fall in what range??

2. Power

   Multivariate test is more powerful in many cases.

   $\boxed{EX}$ All p univariate tests fail to reject, but multivariate test is significant due to combination of small effects on some variables.

3. Understanding variables acting in combination.

A.i. Hotelling's T^2

First consider univariate test of $H_0 : \mu = \mu_0$ vs. $H_1 : \mu \neq \mu_0$ when σ is known. (Consider only two-sided tests, since one-sided don't readily generalize for $p > 1$)

Test statistic using r.s. $(x_1, \ldots, x_n)$:

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1) \text{ under } H_0$$

or

$$\begin{aligned} z^2 &= \text{square of standardized distance} \\ &= n\left(\frac{\bar{x} - \mu_0}{\sigma}\right)^2 \sim \chi_1^2 \text{ under } H_0 \end{aligned}$$

Multivariate generalization ($\mathbf{\Sigma}$ is known):

$$H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0 = \begin{bmatrix} \mu_{01} \\ \mu_{02} \\ \vdots \\ \mu_{0p} \end{bmatrix} \quad \text{vs.} \quad \underbrace{H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0}_{\substack{\text{At least one } \mu_i \\ \text{is not equal to } \mu_{0i}}}$$

Test statistic using r.s. $(\underset{p\times 1}{\mathbf{x}_1}, \ldots, \mathbf{x}_n)$:

$$z^2 = n(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)'\mathbf{\Sigma}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0) \sim \chi^2_p \quad \text{under } H_0$$

More frequently in practice, $\mathbf{\Sigma}$ is unknown.

- Univariate test statistic using r.s. $(x_1, \ldots, x_n)$:

$$
\begin{aligned}
t^2 &= n\left(\frac{\bar{x}-\mu_0}{s}\right)^2 \sim t^2_{n-1} \quad \text{under } H_0 \\
&= \sqrt{n}(\bar{x}-\mu_0)(s^2)^{-1}(\bar{x}-\mu_0)\sqrt{n} \\
&= \left[N_1\left(0,\sigma^2\right)\right]\left[\frac{\text{scaled } \chi^2}{df}\right]^{-1}\left[N_1\left(0,\sigma^2\right)\right]
\end{aligned}
$$

- Multivariate generalization (Hotelling's T^2):

$$T^2 = n(\bar{\mathbf{x}}-\boldsymbol{\mu}_0)' \;\underset{\substack{\uparrow\\ \text{unbiased}\\ \text{estimate}\\ \text{of } \boldsymbol{\Sigma}}}{\mathbf{S}^{-1}}\; (\bar{\mathbf{x}}-\boldsymbol{\mu}_0) \sim T^2_{p,n-1} \quad \text{under } H_0$$

$$= \underset{\substack{\uparrow\\ \bar{\mathbf{x}}\text{ and }\mathbf{S}\\ \text{are indep.}\\ \text{since they}\\ \text{are based}\\ \text{on a r.s.}\\ \text{from MVN}}}{(\bar{\mathbf{x}}-\boldsymbol{\mu}_0)'} \underbrace{\left(\frac{\mathbf{S}}{n}\right)^{-1}}_{\substack{\text{inverse}\\ \text{sample}\\ \text{cov.}\\ \text{matrix}\\ \text{for } \bar{\mathbf{x}}}} (\bar{\mathbf{x}}-\boldsymbol{\mu}_0) \quad \leftarrow \text{“characteristic form”}$$

$$= \underbrace{\sqrt{n}(\bar{\mathbf{x}}-\boldsymbol{\mu}_0)'}_{\substack{N_p(\mathbf{0},\boldsymbol{\Sigma})\\ \text{random vector}}} \underbrace{\left[\frac{\sum_{i=1}^{n}(\mathbf{x}_i-\bar{\mathbf{x}})(\mathbf{x}_i-\bar{\mathbf{x}})'}{n-1}\right]^{-1}}_{\substack{W_p(n-1,\boldsymbol{\Sigma})\\ \text{random matrix}\\ \text{divided by d.f.}}} \underbrace{\sqrt{n}(\bar{\mathbf{x}}-\boldsymbol{\mu}_0)}_{\substack{N_p(\mathbf{0},\boldsymbol{\Sigma})\\ \text{random vector}}}$$

① $T^2_{\alpha,1,\nu} = t^2_{\alpha/2,\nu}$

Table A.7 Upper Percentage Points of Hotelling's T^2 Distribution

$\alpha = 0.05$

Degrees of Freedom, ν	$p = 1$	$p = 2$	$p = 3$	$p = 4$	$p = 5$	$p = 6$	$p = 7$	$p = 8$	$p = 9$	$p = 10$
2	18.513									
3	10.128	57.000								
4	7.709	25.472	114.986							
5	6.608	17.361	46.383	192.468						
6	5.987	13.887	29.661	72.937	289.446					
7	5.591	12.001	22.720	44.718	105.157	405.920				
8	5.318	10.828	19.028	33.230	62.561	143.050	541.890			
9	5.117	10.033	16.766	27.202	45.453	83.202	186.622	697.356		
10	4.965	9.459	15.248	23.545	36.561	59.403	106.649	235.873	872.317	
11	4.844	9.026	14.163	21.108	31.205	47.123	75.088	132.903	290.806	1066.774
12	4.747	8.689	13.350	19.376	27.656	39.764	58.893	92.512	161.967	351.421
13	4.667	8.418	12.719	18.086	25.145	34.911	49.232	71.878	111.676	193.842
14	4.600	8.197	12.216	17.089	23.281	31.488	42.881	59.612	86.079	132.582
15	4.543	8.012	11.806	16.296	21.845	28.955	38.415	51.572	70.907	101.499
16	4.494	7.856	11.465	15.651	20.706	27.008	35.117	45.932	60.986	83.121
17	4.451	7.722	11.177	15.117	19.782	25.467	32.588	41.775	54.041	71.127
18	4.414	7.606	10.931	14.667	19.017	24.219	30.590	38.592	48.930	62.746
19	4.381	7.504	10.719	14.283	18.375	23.189	28.975	36.082	45.023	56.587
20	4.351	7.415	10.533	13.952	17.828	22.324	27.642	34.054	41.946	51.884
21	4.325	7.335	10.370	13.663	17.356	21.588	26.525	32.384	39.463	48.184
22	4.301	7.264	10.225	13.409	16.945	20.954	25.576	30.985	37.419	45.202
23	4.279	7.200	10.095	13.184	16.585	20.403	24.759	29.798	35.709	42.750
24	4.260	7.142	9.979	12.983	16.265	19.920	24.049	28.777	34.258	40.699
25	4.242	7.089	9.874	12.803	15.981	19.492	23.427	27.891	33.013	38.961
26	4.225	7.041	9.779	12.641	15.726	19.112	22.878	27.114	31.932	37.469
27	4.210	6.997	9.692	12.493	15.496	18.770	22.388	26.428	30.985	36.176
28	4.196	6.957	9.612	12.359	15.287	18.463	21.950	25.818	30.149	35.043
29	4.183	6.919	9.539	12.236	15.097	18.184	21.555	25.272	29.407	34.044
30	4.171	6.885	9.471	12.123	14.924	17.931	21.198	24.781	28.742	33.156
35	4.121	6.744	9.200	11.674	14.240	16.944	19.823	22.913	26.252	29.881
40	4.085	6.642	9.005	11.356	13.762	16.264	18.890	21.668	24.624	27.783
45	4.057	6.564	8.859	11.118	13.409	15.767	18.217	20.781	23.477	26.326
50	4.034	6.503	8.744	10.934	13.138	15.388	17.709	20.117	22.627	25.256
55	4.016	6.454	8.652	10.787	12.923	15.090	17.311	19.600	21.972	24.437
60	4.001	6.413	8.577	10.668	12.748	14.850	16.992	19.188	21.451	23.790
70	3.978	6.350	8.460	10.484	12.482	14.485	16.510	18.571	20.676	22.834
80	3.960	6.303	8.375	10.350	12.289	14.222	16.165	18.130	20.127	22.162
90	3.947	6.267	8.309	10.248	12.142	14.022	15.905	17.801	19.718	21.663
100	3.936	6.239	8.257	10.167	12.027	13.867	15.702	17.544	19.401	21.279
110	3.927	6.216	8.215	10.102	11.934	13.741	15.540	17.340	19.149	20.973
120	3.920	6.196	8.181	10.048	11.858	13.639	15.407	17.172	18.943	20.725
150	3.904	6.155	8.105	9.931	11.693	13.417	15.121	16.814	18.504	20.196
200	3.888	6.113	8.031	9.817	11.531	13.202	14.845	16.469	18.083	19.692
400	3.865	6.052	7.922	9.650	11.297	12.890	14.447	15.975	17.484	18.976
1000	3.851	6.015	7.857	9.552	11.160	12.710	14.217	15.692	17.141	18.570
∞	3.841	5.991	7.815	9.488	11.070	12.592	14.067	15.507	16.919	18.307

④ As p increases, larger values of ν are required for T^2 to approach χ^2.

$\boxed{EX}$

$$\frac{T^2_{.05,1,30}}{T^2_{.05,1,\infty}} = 1.086$$

$$\frac{T^2_{.05,5,100}}{T^2_{.05,5,\infty}} = 1.086$$

$$\frac{T^2_{.05,10,200}}{T^2_{.05,10,\infty}} = 1.076$$

② $T^2_{p,\infty} = \chi^2_p$

That is, as $n \to \infty$, $S \to \Sigma$ and $T^2 \to Z^2$

③ Values increase along each row

$\Rightarrow$ T^2 increases with addition of variables, but so does the critical value

Important properties of T^2

1. Sometimes we refer to the subscripts for $T^2_{p,\nu}$ distribution as "dimension" and "df" (e.g., $T^2_{\text{dim},\text{df}}$)

2. Must have $n > p$
   - Otherwise $\mathbf{S}$ is singular and T^2 cannot be computed.

3. Degrees of freedom ν for T^2 is same as for analogous univariate t-test:
   - $\nu = n - 1$ for one-sample test
   - $\nu = n_1 + n_2 - 2$ for two-sample test

4. Alternative hypothesis is 2-sided (no such thing as "$H_1 : \boldsymbol{\mu} > \boldsymbol{\mu}_0$")
   - Critical region is one-tailed (reject for large values) since test statistic is *squared* distance

5. $\frac{\nu-p+1}{\nu p} T^2_{p,v} \overset{q}{\equiv} F_{p,\nu-p+1}$

[Note: "$\overset{q}{\equiv}$" is shorthand for the equivalence of the quantiles of two dist'ns]

- So, p-value for T^2 test is

$$p\text{-value} \;=\; \Pr\left\{F_{p,\nu-p+1} > \frac{\nu-p+1}{\nu p}T^2\right\}$$

- Critical value for T^2 test is

$$T^2_{\alpha,p,\nu} = \frac{\nu p}{\nu-p+1}F_{\alpha,p,\nu-p+1}\left(\text{or } \frac{(n-1)p}{n-p}F_{\alpha,p,n-p} \text{ when } \nu = n-1\right)$$

6. T^2 invariant under transformations of the form $\underset{p\times 1}{\ddot{\mathbf{x}}} = \underset{p\times p}{\mathbf{C}}\;\underset{p\times 1}{\mathbf{x}} + \mathbf{d}$, where $\mathbf{C}$ is nonsingular

7. T^2 is the likelihood ratio test (LRT) of $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$

- Under H_0 the likelihood is

$$L(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{np/2}|\boldsymbol{\Sigma}|^{n/2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(\mathbf{x}_i - \boldsymbol{\mu}_0)'\boldsymbol{\Sigma}^{-1}(\mathbf{x}_i - \boldsymbol{\mu}_0)\right\}$$

$$= \frac{1}{(2\pi)^{np/2}|\boldsymbol{\Sigma}|^{n/2}} \exp\left\{-\frac{1}{2}\mathrm{tr}\left[\boldsymbol{\Sigma}^{-1}(\sum_{i=1}^{n}(\mathbf{x}_i - \boldsymbol{\mu}_0)(\mathbf{x}_i - \boldsymbol{\mu}_0)')\right]\right\}$$

Using Result 4.10 (again), we obtain

$$\max_{\boldsymbol{\Sigma}} L(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{\frac{np}{2}}|\hat{\boldsymbol{\Sigma}}_0|^{\frac{n}{2}}} \exp\left\{\frac{-np}{2}\right\}$$

where $\hat{\boldsymbol{\Sigma}}_0 = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i - \boldsymbol{\mu}_0)(\mathbf{x}_i - \boldsymbol{\mu}_0)'$

- Recall

$$\max_{\boldsymbol{\mu},\boldsymbol{\Sigma}} L(\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{\frac{np}{2}} |\hat{\boldsymbol{\Sigma}}|^{\frac{n}{2}}} \exp\left\{\frac{-np}{2}\right\}$$

where $\hat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})'$

- Likelihood Ratio:

$$\lambda = \frac{\max_{\boldsymbol{\Sigma}} L(\boldsymbol{\mu}_0, \boldsymbol{\Sigma})}{\max_{(\boldsymbol{\mu},\boldsymbol{\Sigma})} L(\boldsymbol{\mu}, \boldsymbol{\Sigma})} = \left(\frac{|\hat{\boldsymbol{\Sigma}}|}{|\hat{\boldsymbol{\Sigma}}_0|}\right)^{\frac{n}{2}} < c_\alpha$$

$$\Lambda = \lambda^{\frac{2}{n}} = \frac{|\hat{\boldsymbol{\Sigma}}|}{|\hat{\boldsymbol{\Sigma}}_0|} = \frac{1}{1 + \frac{1}{n-1}T^2}$$

"Wilks' Lambda" is rejected for small Λ or large T^2

* $T^2 = (n-1)\frac{|\hat{\Sigma}_0|}{|\hat{\boldsymbol{\Sigma}}|} - (n-1)$

* $-2\ln\lambda \sim \chi^2_{\nu-\nu_0}$

where $\nu = \#$ of unrestricted parameters

and $\nu_0 = \#$ of parameters under H_0

$\boxed{\text{ex}}$ Turnips

A.ii. Confidence Regions

- Confidence region $R(\mathbf{X})$:

  Set of possible values of $\boldsymbol{\theta}$ in $\boldsymbol{\Theta}$ based on $\mathbf{X}$

- $R(\mathbf{X})$ is $100(1-\alpha)\%$ C.R. if, before the sample is selected

  $\Pr\{R(\mathbf{X})$ will cover the true $\boldsymbol{\theta}\} = 1-\alpha$.

- C.R. for $\boldsymbol{\mu}$ $[100(1-\alpha)\%]$

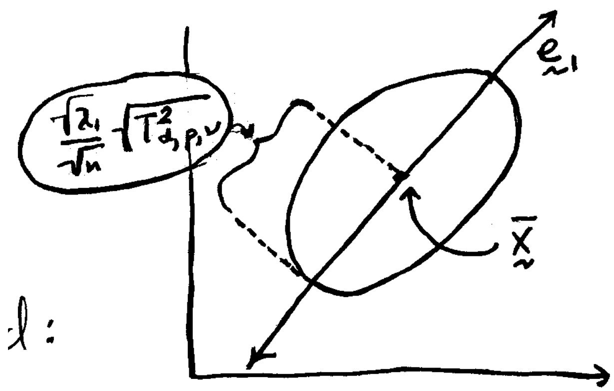
  $$\{\text{all } \boldsymbol{\mu} \ni \underbrace{n(\bar{\mathbf{x}}-\boldsymbol{\mu})'S^{-1}(\bar{\mathbf{x}}-\boldsymbol{\mu})}_{\text{squared mult. distance from } \bar{\mathbf{x}}} \le T^2_{\alpha,p,\nu}\}$$

  or $\{\text{all } \boldsymbol{\mu} \ni n(\bar{\mathbf{x}}-\boldsymbol{\mu})'S^{-1}(\bar{\mathbf{x}}-\boldsymbol{\mu}) \le \frac{\nu p}{\nu-p+1} F_{\alpha,p,\nu-p+1}\}$

- Axes of the ellipsoid (based on eigenvalues $\lambda_1, \ldots, \lambda_p$ and eigenvectors $\mathbf{e}_1, \ldots, \mathbf{e}_p$ of $\mathbf{S}$):

  $\frac{\pm\sqrt{\lambda_i}}{\sqrt{n}} \sqrt{T^2_{\alpha,p,\nu}}$ along $\mathbf{e}_i$

  Elongation of ellipsoid: $\frac{\sqrt{\lambda_1}}{\sqrt{\lambda_2}}$



Interest in C.I.'s for individual components of $\mathbf{x}$ or linear combination $\mathbf{a}'\mathbf{x}$.

- Define $z = \mathbf{a}'\mathbf{x}$

$$z \sim N_1(\mathbf{a}'\boldsymbol{\mu}, \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a}) = N_1(\mu_z, \sigma_z^2)$$

- Sample statistics:

$$\bar{z} = \mathbf{a}'\bar{\mathbf{x}}$$

$$s_z^2 = \mathbf{a}'\mathbf{S}\mathbf{a}$$

  Note: $\mathbf{a}_1 = [0, 1, 0, \ldots, 0]$ will yield $\mathbf{a}_1'\bar{\mathbf{x}} = \bar{x}_2$ and $\mathbf{a}_2 = [1, -1, 0, \ldots, 0]$ implies that $\mathbf{a}_2'\bar{\mathbf{x}} = \bar{x}_1 - \bar{x}_2$, etc.

- $100(1-\alpha)\%$ C.I. for μ_z is

$$\mathbf{a}'\bar{\mathbf{x}} \pm t_{\frac{\alpha}{2}, n-1}\sqrt{\frac{\mathbf{a}'\mathbf{S}\mathbf{a}}{n}} \qquad \text{“\underline{t-interval}”}$$

– Experimentwise Type I error rate (EER)

$$\begin{aligned} &\Pr\{\text{at least one C.I. "wrong"}\} \\ &\quad = 1 - \Pr\{\text{no C.I.'s are wrong}\} \\ &\quad = 1 - (1-\alpha)^p \qquad \text{assuming independence of C.I.'s} \end{aligned}$$

$\boxed{\text{ex}}$ $\alpha = .05$: EER for $p = 10$ is $1 - (.95)^{10} \cong .40$

- Rewrite t-interval as

$$\left\{ \text{all } \mathbf{a}'\boldsymbol{\mu} \ni n \frac{(\mathbf{a}'(\bar{\mathbf{x}} - \boldsymbol{\mu}))^2}{\mathbf{a}'\mathbf{S}\mathbf{a}} \le t^2_{n-1} \right\}$$

Is there a bound c^2 which can replace t^2_{n-1} and defines a C.R. that *simultaneously* contains $\mathbf{a}'\boldsymbol{\mu}$ for all $\mathbf{a}$??

- Preliminary result (2–50, JW)

  For $\underset{p\times p}{\mathbf{B}}$ p.d. and $\mathbf{x}\neq\mathbf{0}$

$$\max_{\mathbf{x}\neq 0}\frac{(\mathbf{x}'\mathbf{d})^2}{\mathbf{x}'\mathbf{B}\mathbf{x}} = \mathbf{d}'\mathbf{B}^{-1}\mathbf{d}$$

  with maximum attained when $\mathbf{x}=c\mathbf{B}^{-1}\mathbf{d}, c\neq 0$

- So, $\max_{\mathbf{a}\neq 0}\frac{\left(\mathbf{a}'(\bar{\mathbf{x}}-\boldsymbol{\mu})\right)}{\mathbf{a}'\mathbf{S}\mathbf{a}} = n\left(\bar{\mathbf{x}}-\boldsymbol{\mu}\right)'\mathbf{S}^{-1}\left(\bar{\mathbf{x}}-\boldsymbol{\mu}\right) = T^2$

  with maximum at

$$\begin{aligned}\mathbf{a} &= c\,\mathbf{S}^{-1}\left(\bar{\mathbf{x}}-\boldsymbol{\mu}\right), \quad c\neq 0\\ &= \text{“}\underline{\text{discriminant function}}\text{”}\end{aligned}$$

$\Longrightarrow$ Simultaneously for all $\mathbf{a}$, the interval

$$\mathbf{a}'\bar{\mathbf{x}} \pm \sqrt{T^2_{\alpha,p,\nu}}\sqrt{\frac{\mathbf{a}'\mathbf{S}\mathbf{a}}{n}}$$

or

$$\mathbf{a}'\bar{\mathbf{x}} \pm \sqrt{\frac{\nu p}{\nu - p + 1} F_{\alpha,p,\nu-p+1} \frac{\mathbf{a}'\mathbf{S}\mathbf{a}}{n}}$$

$\leftarrow$ "<u>T^2 interval</u>"

or when $\nu = n - 1$

$$\mathbf{a}'\bar{\mathbf{x}} \pm \sqrt{\frac{(n-1)p}{n(n-p)} F_{\alpha,p,n-p} \mathbf{a}'\mathbf{S}\mathbf{a}}$$

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

☹ More conservative (wider) than t-interval

☺ Preserve EER$\leq \alpha$

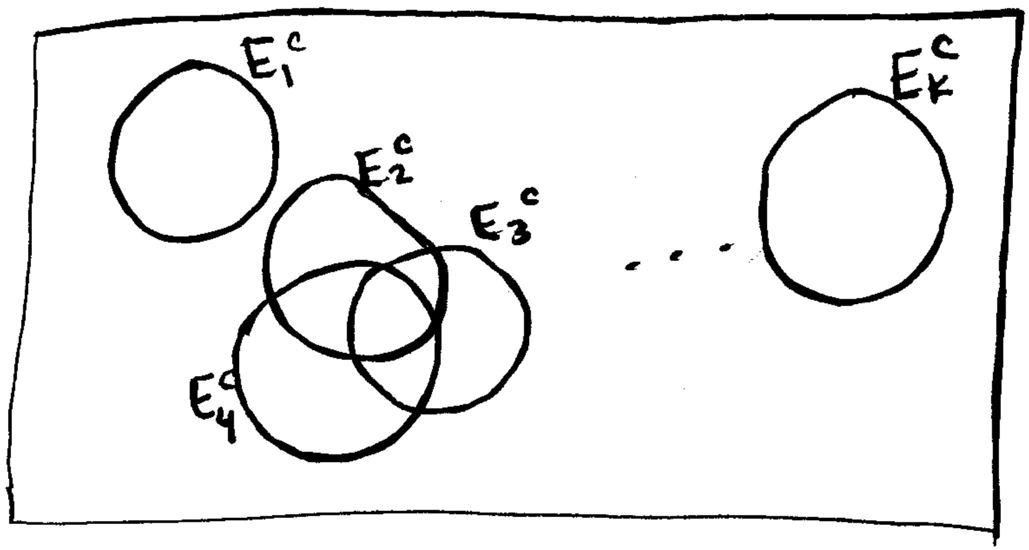
☺ Allows "data-snooping"

If we're willing to specify a few linear combinations $\mathbf{a}_1, \ldots, \mathbf{a}_k$ before collecting the data, we might consider using intervals based on the Bonferroni inequality which are narrower than T^2 intervals but still protect EER for a finite set of l.c.'s.

Given C.I.'s for k l.c.'s $\mathbf{a}_1'\boldsymbol{\mu}, \ldots, \mathbf{a}_k'\boldsymbol{\mu}$,

- E_i: event that i^{th} interval contains $\mathbf{a}_i'\boldsymbol{\mu}$
- $P\{E_i^c\} = \alpha_i$

$$\begin{aligned}
\Pr\{\text{all } E_i\} &= 1 - \Pr\{\text{at least one } E_i^c\} \\
&= 1 - \Pr\{E_1^c \cup E_2^c \cup \cdots \cup E_k^c\} \\
&\geq 1 - \sum_{i=1}^{k} \Pr\{E_i^c\} \\
&= 1 - \Sigma\alpha_i
\end{aligned}$$



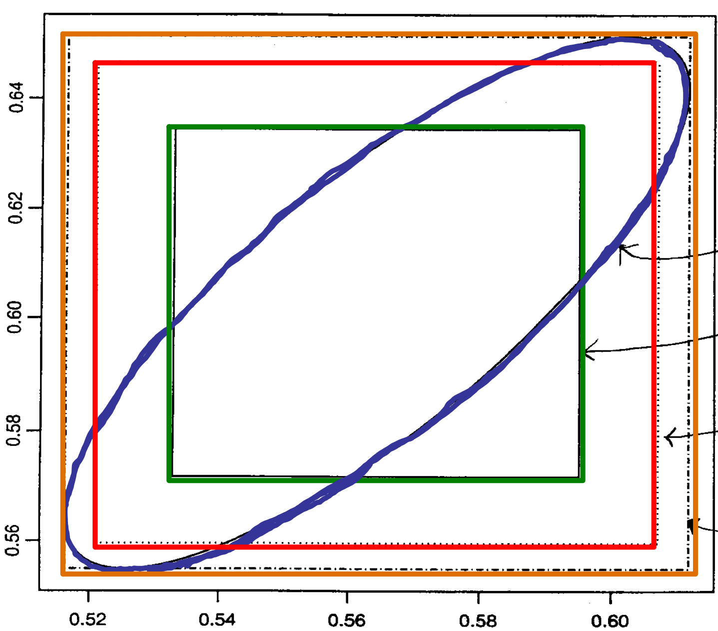
- Usually, specify $\alpha_i = \frac{\alpha}{k}$

So,

$$\mathbf{a}'\bar{\mathbf{x}} \pm t_{\frac{\alpha}{2k},n-1}\sqrt{\frac{\mathbf{a}'\mathbf{S}\mathbf{a}}{n}} \qquad \text{“\underline{Bonferroni Interval}”}$$

Critical values for 95% C.I.'s for $\mu_1, \ldots, \mu_p$

		$p=5$		$p=12$	
n	$t_{\alpha/2}$	$t_{\alpha/2p}$	$\sqrt{T_\alpha^2}$	$t_{\alpha/2p}$	$\sqrt{T_\alpha^2}$
15	2.14	2.98	4.83	3.42	22.13
25	2.06	2.80	4.03	3.17	7.59
40	2.02	2.71	3.72	3.04	5.95
100	1.98	2.63	3.47	2.93	5.02
200	1.97	2.60	3.40	2.90	4.79



$\boxed{\text{ex}}$ $p=2$

- Bivariate confidence region
- t intervals
- Bonferroni intervals
- T^2 intervals

Notes:

- Often useful to examine the discriminant function $\mathbf{a} = \mathbf{S}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu})$ in

$$\max_{\mathbf{a} \neq 0} \frac{n(\mathbf{a}'(\bar{\mathbf{x}} - \boldsymbol{\mu}))^2}{\mathbf{a}'\mathbf{S}\mathbf{a}} = T^2$$

- $\mathbf{a}$ indicates the relative contribution of the x's to the separation of the data from $\boldsymbol{\mu}_0$
  - Comparisons of $a_1, \ldots, a_p$ only informative when x's are commensurate (i.e., measured on the same scale with comparable variances)

– If x's are not commensurate, consider coefficients $a_1^*, \ldots, a_p^*$ that are applicable to standard variables.

– Discriminant function in terms of standardized variables

$$z = a_1^* \frac{x_1 - \bar{x}_1}{s_1} + \cdots + a_p^* \frac{x_p - \bar{x}_p}{s_p}$$

instead of

$$z = a_1 x_1 + \cdots + a_p x_p$$

OR

$$\mathbf{a}^* = \mathbf{D}^{\frac{1}{2}} \mathbf{a} \qquad \text{where } \mathbf{D} = \begin{bmatrix} s_{11} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & s_{pp} \end{bmatrix}$$

$\boxed{\text{ex}}$ Turnips

III.B. Comparison of Several Mean Vectors

III.B.i. Paired Observations

Let $\mathbf{x}_{1i}$ and $\mathbf{x}_{2i}$ be 2 p-variate responses for observation i ($i = 1, \ldots, n$)

$\boxed{\text{ex}}$ LaVerl's SAT pre-class test grades and post-class grades

Pre-class grades:

$$\mathbf{x}_{1i} = (\text{Quant} = 640, \text{Analyt} = 610, \text{Verbal} = 490)$$

Post-class grades:

$$\mathbf{x}_{2i} = (\text{Quant} = 680, \text{Analyt} = 620, \text{Verbal} = 560)$$

1. Calculate $\mathbf{d}_i = \mathbf{x}_{1i} - \mathbf{x}_{2i}$

2. Calculate

$$\bar{\mathbf{d}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{d}_i$$

and

$$\mathbf{S}_d = \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{d}_i - \bar{\mathbf{d}})(\mathbf{d}_i - \bar{\mathbf{d}})'$$

3. $T^2 = n\bar{\mathbf{d}}'\mathbf{S}_d^{-1}\bar{\mathbf{d}} \sim T^2_{p,\underbrace{n\,-\,1}_{\nu}} \overset{q}{\equiv} \frac{(n-1)p}{n-p} F_{p,\underbrace{n\,-\,p}_{\nu-p+1}}$

   [Note: "$\overset{q}{\equiv}$" is shorthand for the equivalence of the quantiles of two dist'ns]

⋆ Same follow-up analyses as in one-sample T^2 test/intervals apply here

- Confidence regions/intervals
- Discriminant functions

Alternatively, think of each observation

$$\underset{2p\times 1}{\mathbf{x}_i} = \begin{bmatrix} \mathbf{x}_{1i} \\ \mathbf{x}_{2i} \end{bmatrix} \begin{array}{l} \leftarrow \text{ pre-tests} \\ \leftarrow \text{ post-tests} \end{array}$$

$$\underset{2p\times 1}{\bar{\mathbf{x}}} = \begin{bmatrix} \bar{\mathbf{x}}_1 \\ \bar{\mathbf{x}}_2 \end{bmatrix}$$

$$\underset{2p\times 2p}{\mathbf{S}} = \begin{bmatrix} \mathbf{S}_{11} & \mathbf{S}_{12} \\ \mathbf{S}_{21} & \mathbf{S}_{22} \end{bmatrix}$$

Interest is in $\mathbf{C}\mathbf{x}_i$, where

$$\underset{p\times 2p}{\mathbf{C}} = \begin{bmatrix} 1 & & & \mathbf{0} & -1 & & & \mathbf{0} \\ & 1 & & & & -1 & & \\ & & \ddots & & & & \ddots & \\ \mathbf{0} & & & 1 & \mathbf{0} & & & -1 \end{bmatrix}$$

Note

$$\mathbf{d}_i = \mathbf{C}\mathbf{x}_i$$

$$\bar{\mathbf{d}} = \mathbf{C}\bar{\mathbf{x}}$$

$$\mathbf{S}_d = \mathbf{C}\mathbf{S}\mathbf{C}'$$

$$\text{and } T^2 = n\bar{\mathbf{x}}'\mathbf{C}'(\mathbf{C}\mathbf{S}\mathbf{C}')^{-1}\mathbf{C}\bar{\mathbf{x}} \sim T^2_{p,n-1}$$

$$\overset{q}{\equiv} \frac{(n-1)p}{(n-1-p+1)} F_{p,n-1-p+1}$$

$$\overset{q}{\equiv} \frac{(n-1)p}{n-p} F_{p,n-p}$$

An extension to a comparison of p treatments given to each subject over time

$$
\mathbf{x}_i = \begin{bmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{ip} \end{bmatrix}
\begin{array}{l} \leftarrow \text{ evaluation after day 1 dosage} \\ \leftarrow \text{ evaluation after day 2 dosage} \\ \\ \leftarrow \text{ evaluation after day p dosage} \end{array}
\qquad i = 1, \ldots, n
$$

Interest may lie in comparisons of treatment means

$$
\underset{(p-1)\times p}{\mathbf{C}}\boldsymbol{\mu} = \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & & -1 & 1 \end{bmatrix}
\begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{bmatrix}
= \begin{bmatrix} \mu_2 - \mu_1 \\ \mu_3 - \mu_2 \\ \vdots \\ \mu_p - \mu_{p-1} \end{bmatrix}
$$

$$
\begin{aligned}
T^2 = n(\mathbf{C}\bar{\mathbf{x}})'(\mathbf{C}\mathbf{S}\mathbf{C}')^{-1}\mathbf{C}\bar{\mathbf{x}} \sim T^2_{p-1,n-1} & \\
& \overset{q}{\equiv} \frac{(n-1)(p-1)}{(n-1-(p-1)+1)} F_{(p-1),n-1-(p-1)+1} \\
& \overset{q}{\equiv} \frac{(n-1)(p-1)}{n-p+1} F_{p-1,n-p+1}
\end{aligned}
$$

e.g., if comparing 3 days, we might use

$$
\underset{2\times 3}{\mathbf{C}} = \begin{bmatrix} -1 & 0 & 1 \\ 1 & -2 & 1 \end{bmatrix} \begin{array}{l} \leftarrow \text{linear increase/decrease in response} \\ \leftarrow \text{quadratic effect on response} \end{array}
$$

e.g., if comparing 4 days, we might use

$$
\underset{3\times 4}{\mathbf{C}} = \begin{bmatrix} -3 & -1 & 1 & 3 \\ 1 & -1 & -1 & 1 \\ 1 & -3 & 3 & -1 \end{bmatrix} \begin{array}{l} \leftarrow \text{linear} \\ \leftarrow \text{quadradic} \\ \leftarrow \text{cubic} \end{array}
$$

B.ii. Two-Sample Comparisons

Interest in $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2$ (difference in two population means).

Assumptions:

- $\mathbf{x}_{11}, \mathbf{x}_{12}, \ldots, \mathbf{x}_{1n_1}$ is a r.s. from $N_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$
- $\mathbf{x}_{21}, \mathbf{x}_{22}, \ldots, \mathbf{x}_{2n_2}$ is a r.s. from $N_p(\boldsymbol{\mu}_2, \boldsymbol{\Sigma})$
  - Note that $\boldsymbol{\Sigma}_1 = \boldsymbol{\Sigma}_2 = \boldsymbol{\Sigma}$
- The two samples are independent

In practice, we can relax these assumptions somewhat for large n.

Let
$$\bar{\mathbf{x}}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \mathbf{x}_{ij}, \qquad i = 1, 2$$

$$\mathbf{S}_i = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)(\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)', \qquad i = 1, 2$$

Since $\quad (n_1 - 1)\mathbf{S}_1 \sim W_p(n_1 - 1, \boldsymbol{\Sigma})$

and $\quad (n_2 - 1)\mathbf{S}_2 \sim W_p(n_2 - 1, \boldsymbol{\Sigma})$

$$\underbrace{(n_1 - 1)\mathbf{S}_1 + (n_2 - 1)\mathbf{S}_2}_{=(n_1+n_2-2)\mathbf{S}_{p\ell}} \sim W_p(n_1 + n_2 - 2, \boldsymbol{\Sigma})$$

$$\Longrightarrow E\{\mathbf{S}_{p\ell}\} = \boldsymbol{\Sigma}$$

Since the two samples are independent

$$(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) \sim N_p\left(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2, \frac{1}{n_1}\boldsymbol{\Sigma} + \frac{1}{n_2}\boldsymbol{\Sigma}\right)$$

and

$$\begin{aligned} T^2 &= [\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2 - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)]' \left[\left(\frac{1}{n_1} + \frac{1}{n_2}\right)\mathbf{S}_{p\ell}\right]^{-1} [\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2 - (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2)] \\ &\sim T^2_{p,\nu} = T^2_{p,n_1+n_2-2} \\ &\overset{q}{\equiv} \frac{(n_1 + n_2 - 2)p}{(n_1 + n_2 - 2) - p + 1} F_{p,(n_1+n_2-2)-p+1} \end{aligned}$$

$100(1-\alpha)\%$ C.R. for $\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 = \boldsymbol{\delta}$:

$$\left\{\text{all } \boldsymbol{\delta} \ni T^2 \leq T^2_{\alpha,p,\nu}\right\} \qquad \nu = n_1 + n_2 - 2$$

where T^2 is the squared mult. distance between $\bar{\mathbf{x}}_1$ and $\bar{\mathbf{x}}_2$

or

$$\left\{\text{all } \boldsymbol{\delta} \ni T^2 \leq \frac{(n_1+n_2-2)p}{(n_1+n_2-2)-p+1} F_{\alpha,p,(n_1+n_2-2)-p+1}\right\}$$

<u>Follow-up analyses</u>

- "t-interval":

$$\mathbf{a}'\bar{\mathbf{x}}_1 - \mathbf{a}'\bar{\mathbf{x}}_2 \pm t_{\frac{\alpha}{2},n_1+n_2-2}\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\mathbf{a}'\mathbf{S}_{p\ell}\mathbf{a}}$$

- "Bonferroni interval":

$$\mathbf{a}'\bar{\mathbf{x}}_1 - \mathbf{a}'\bar{\mathbf{x}}_2 \pm t_{\frac{\alpha}{2k},n_1+n_2-2}\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\mathbf{a}'\mathbf{S}_{p\ell}\mathbf{a}}$$

  - k is # of contrasts of interest

    $\boxed{\text{ex}}$ want intervals for each of p variables

    Then, $[\mathbf{a}_1, \ldots, \mathbf{a}_p] = \mathbf{I}_p$ and $k = p$

- "T^2-interval"

$$\mathbf{a}'\bar{\mathbf{x}}_1 - \mathbf{a}'\bar{\mathbf{x}}_2 \pm \sqrt{T^2_{\alpha,p,\nu}}\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\mathbf{a}'\mathbf{S}_{p\ell}\mathbf{a}}$$

  where $T^2_{\alpha,p,\nu} \equiv \frac{(n_1+n_2-2)p}{(n_1+n_2-2)-p+1}F_{\alpha,p,(n_1+n_2-2)-p+1}$

- Examine discriminant function

$$\mathbf{a} = \mathbf{S}_{p\ell}^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)$$

  for indication of contribution of the variables to separation of the groups

→ If x's are not commensurate consider standardized coefficients

$$\mathbf{a}^\star = \mathbf{D}_{p\ell}^{\frac{1}{2}}\mathbf{a}$$

where

$$\mathbf{D}_{p\ell} = \mathrm{diag}\{\mathbf{S}_{p\ell}\} = \begin{bmatrix} s_{11,p\ell} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & s_{pp,p\ell} \end{bmatrix}$$

$\boxed{\text{ex}}$ Duchenne muscular dystrophy

– Test $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ using $x_3, x_4, x_5, \& x_6$
  - ∗ Individual tests using $t_{\alpha/2}, t_{\alpha/2p}, \sqrt{T^2_{\alpha,p,\nu}}$ as critical values
  - ∗ Examine discriminant function coeff.
    - · Standardized coefficients

Testing $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ when $\boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_2$

Univariate case ("Behrens-Fisher Problem"):

$$t^* = \frac{\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \overset{\text{approx}}{\sim} t_\nu$$

where

$$\nu = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\left[\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1+1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2+1}\right]} \qquad \leftarrow \text{(Welch, 1937, 1947)}$$

- Hsu (1938) and Scheffe' (1959) argue that significance level for usual t-test is preserved when $n_1 = n_2$

Multivariate case:

$$T^{*^2} = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \left[\frac{\mathbf{S}_1}{n_1} + \frac{\mathbf{S}_2}{n_2} \right]^{-1} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) \to \chi_p^2$$

$$\text{as} \qquad (n_1 - p) \to \infty, \ (n_2 - p) \to \infty$$

- Significance level preserved for usual T^2 test when $n_1 = n_2$ and n_1 and n_2 are "very large" (Ito and Schull, 1964)

- "If sample sizes are equal the significance level [of usual T^2 test] is not affected" (Carter, Khatri, and Srivastava, 1979)

? But do these properties hold with small to moderate sample sizes ?

Simulation Study in Christensen & Rencher (1997)

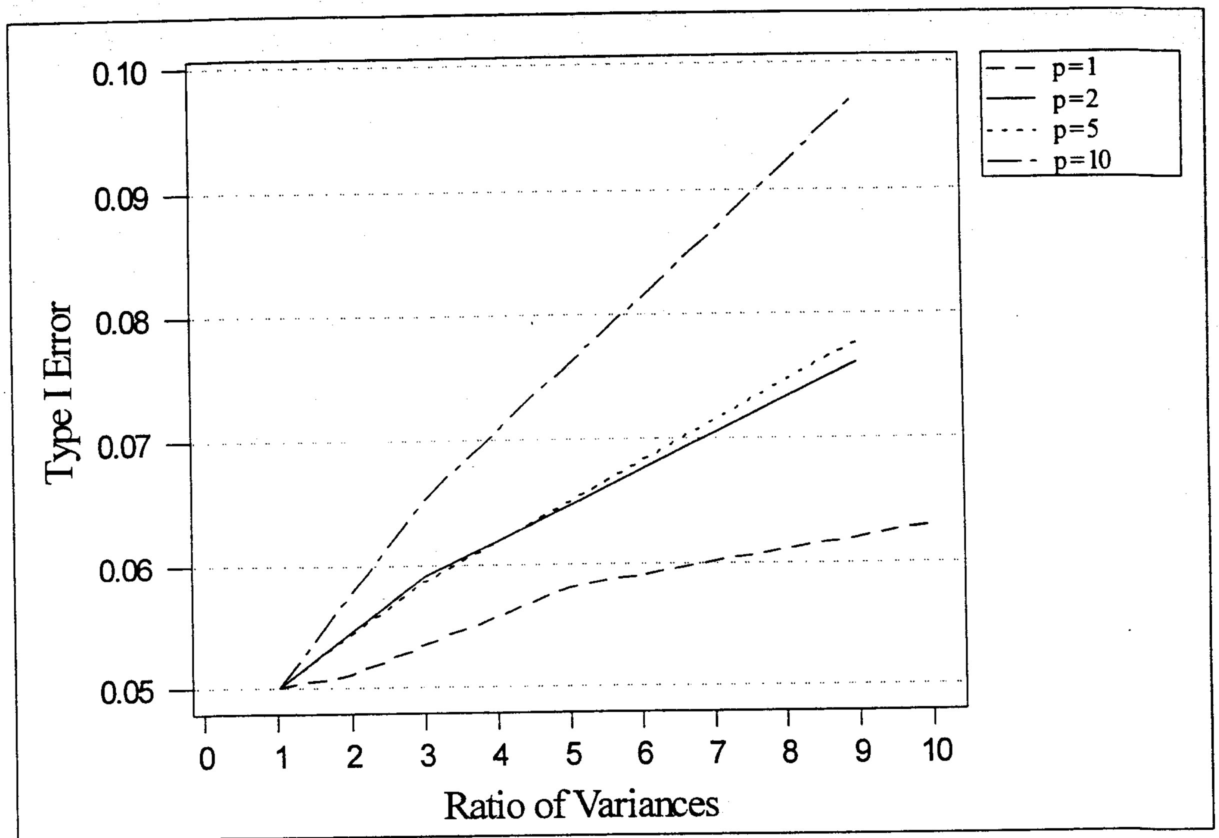


FIG. 5. Type I Error for Student's t (p=1) and Hotelling's T^2 (p=2, 5 and 10) for Different Variance Ratios ($\sigma_2:\sigma_1$ or $\Sigma_2:\Sigma_1$) when $n_1 = n_2$
(Christensen & Rencher, 1997)

For matrices of form $\mathbf{\Sigma}_2 = k\mathbf{\Sigma}_1$, equality of sample sizes ($n_1 = n_2$) is less able to protect Type I error rate as p increases

- (Study considered small to moderate $n_1, n_2 \in (2p, 10p)$)

For multivariate Behrens-Fisher problem, consider

$$T^{*^2} = (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)'\mathbf{S}_e^{-1}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)$$

as a statistic, where

$$\mathbf{S}_e = \frac{\mathbf{S}_1}{n_1} + \frac{\mathbf{S}_2}{n_2}$$

- There are several tests for $\boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ when $\boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_2$, and many of these use

$$T^{*^2} \overset{\text{approx}}{\sim} T^2_{p,\nu^*}$$

For example:

  - Yao (1965) test uses

$$\frac{1}{\nu^*} = \frac{1}{(T^{*^2})^2} \sum_{i=1}^{2} \frac{1}{n_i - 1} \left[(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{S}_e^{-1} \frac{\mathbf{S}_i}{n_i} \mathbf{S}_e^{-1} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2) \right]^2$$

    * Note: this is a multivariate extension of Welch's approach to univariate problem

  - Nel and Van der Merwe (1986) test uses

$$\nu^* = \frac{\operatorname{tr}\left\{\mathbf{S}_e^2\right\} + \left(\operatorname{tr}\{\mathbf{S}_e\}\right)^2}{\sum_{i=1}^{2} \frac{1}{n_i - 1} \left[\operatorname{tr}\left\{ \left(\frac{\mathbf{S}_i}{n_i} \right)^2 \right\} + \left(\operatorname{tr}\left\{ \frac{\mathbf{S}_i}{n_i} \right\} \right)^2 \right]}$$

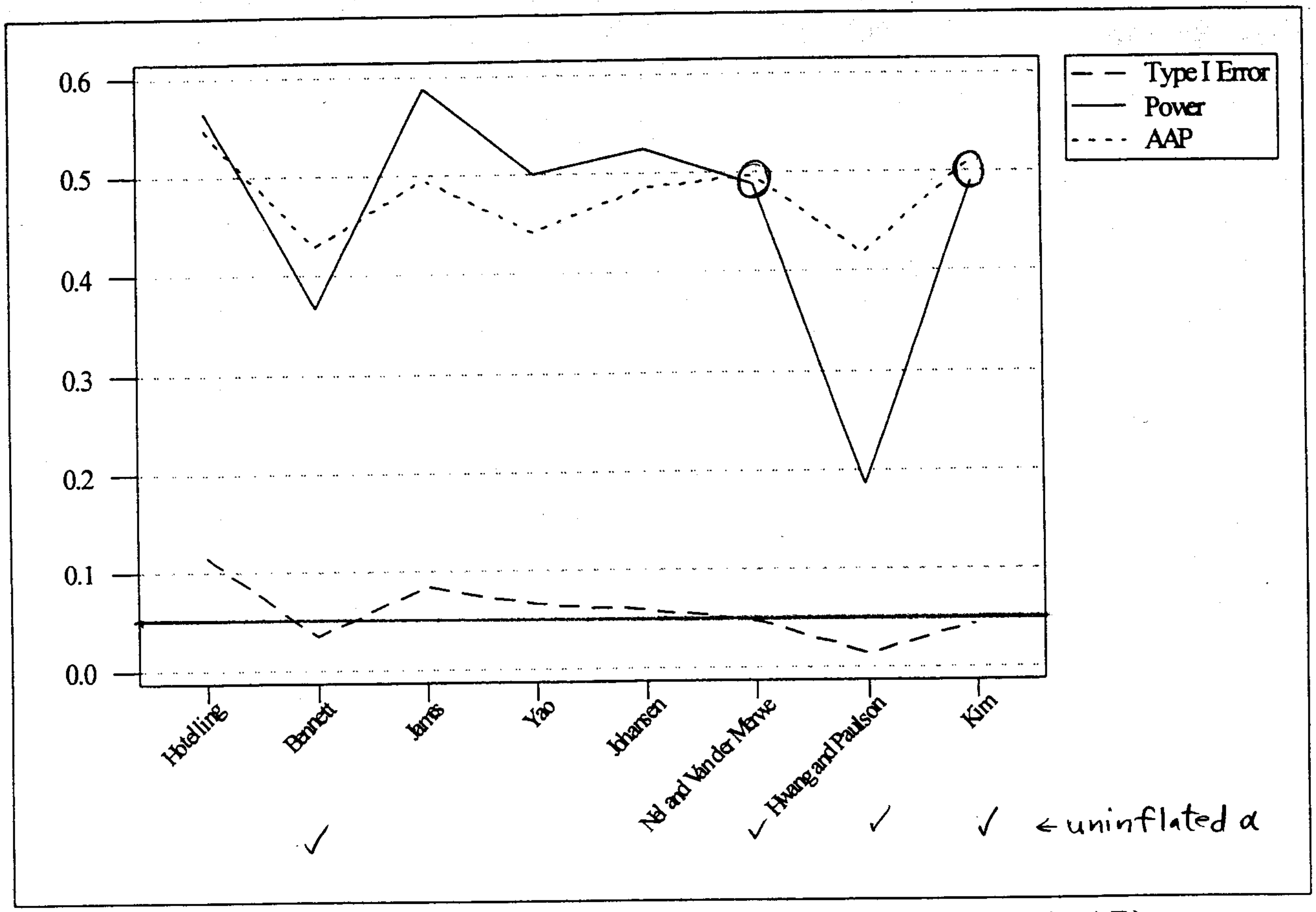


FIG. 1. Average Type I Error, Power and Alpha-adjusted Power (AAP)

(Christensen & Rencher, 1997)

- Simulation study: Nel and Van der Merwe (1986) and Kim (1992) have highest power among tests with uninflated Type I error rate

$\boxed{\text{ex}}$ Muscular Dystrophy

Tests for additional information

Let $\mathbf{x}_{1i} = \begin{pmatrix} \mathbf{y}_{1i} \\ \mathbf{z}_{1i} \end{pmatrix} \begin{matrix} \leftarrow & p \times 1 \\ \leftarrow & q \times 1 \end{matrix}$, $i = 1, \ldots, n_1$ be a r.s. from $N_{p+q}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma})$

and $\mathbf{x}_{2i} = \begin{pmatrix} \mathbf{y}_{2i} \\ \mathbf{z}_{2i} \end{pmatrix} \begin{matrix} \leftarrow & p \times 1 \\ \leftarrow & q \times 1 \end{matrix}$, $i = 1, \ldots, n_2$ be a r.s. from $N_{p+q}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma})$

- Start with $\mathbf{y}$ measurements
  - Will the $q \times 1$ subvector $\mathbf{z}$ measured in addition to $\mathbf{y}$ significantly increase the separation of the two samples (or is $\mathbf{z}$ redundant in presence of $\mathbf{y}$?)
- Sample means: $\bar{\mathbf{x}}_1 = \begin{pmatrix} \bar{\mathbf{y}}_1 \\ \bar{\mathbf{z}}_1 \end{pmatrix}$ and $\bar{\mathbf{x}}_2 = \begin{pmatrix} \bar{\mathbf{y}}_2 \\ \bar{\mathbf{z}}_2 \end{pmatrix}$

  Common sample covariance matrix: $\mathbf{S}_{p\ell} = \begin{bmatrix} \mathbf{S}_{yy} & \mathbf{S}_{yz} \\ \mathbf{S}_{zy} & \mathbf{S}_{zz} \end{bmatrix}$

- If $\mathbf{y}$ and $\mathbf{z}$ are independent:

$$T_{p+q}^2 = T_p^2 + T_q^2$$

- If not independent: Compare T_{p+q}^2 with T_p^2

$$T_{p+q}^2 = \frac{n_1 n_2}{n_1 + n_2} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \, \mathbf{S}_{p\ell}^{-1} \, (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)$$

$$T_p^2 = \frac{n_1 n_2}{n_1 + n_2} (\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)' \, \mathbf{S}_{yy}^{-1} \, (\bar{\mathbf{y}}_1 - \bar{\mathbf{y}}_2)$$

Then, we can show that

$$\begin{aligned} T_{\text{add}}^2 = (\nu - p)\frac{T_{p+q}^2 - T_p^2}{\nu + T_p^2} &\sim T_{q,\nu-p}^2 \\ &\overset{q}{\equiv} \frac{(\nu-p)q}{\nu - p - q + 1} F_{q,\nu-p-q+1} \end{aligned}$$

or

$$F_{\text{add}} = \left(\frac{\nu - p - q + 1}{q}\right) \frac{T^2_{p+q} - T^2_p}{\nu + T^2_p} \sim F_{q,\nu-p-q+1}$$

where $\nu = n_1 + n_2 - 2$

- If just checking the addition of one variable:

$$T^2_{\text{add}} \sim F_{1,\nu-p}$$

$\boxed{\text{ex}}$ Duchenne muscular dystrophy

– x_3 and x_4 are relatively inexpensive to measure compared to x_5 and x_6. Are x_5 and x_6 important above and beyond x_3 and x_4

– x_3, x_4, x_5, x_6 may depend on age and season. Are $x_1 =$ age and $x_2 =$ season important?

B.iii. MANOVA (one-way)

- Comparing means from g groups

$$\left.\begin{array}{l} \text{Sample from population 1: } \mathbf{x}_{11}, \mathbf{x}_{12}, \ldots, \mathbf{x}_{1n_1} \\ \text{Sample from population 2: } \mathbf{x}_{21}, \mathbf{x}_{22}, \ldots, \mathbf{x}_{2n_2} \\ \qquad\qquad\vdots \\ \text{Sample from population } g\text{: } \mathbf{x}_{g1}, \mathbf{x}_{g2}, \ldots, \mathbf{x}_{gn_g} \end{array}\right\} \begin{array}{c} \text{independent} \\ \text{random} \\ \text{samples} \end{array}$$

$\mathbf{x}_{\ell j} \sim N(\boldsymbol{\mu}_\ell, \mathbf{\Sigma})$ $\qquad \leftarrow \mathbf{\Sigma}$ is the common covariance matrix

- Instead of testing

$$H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \cdots = \boldsymbol{\mu}_g \text{ vs. } H_1\text{: at least two } \boldsymbol{\mu}\text{'s are unequal}$$

we usually reparameterize

$$\boldsymbol{\mu}_\ell = \boldsymbol{\mu} + \boldsymbol{\tau}_\ell \quad \leftarrow \text{ treatment effect}$$

Thus $\mathbf{x}_{\ell j} \sim N(\boldsymbol{\mu} + \boldsymbol{\tau}_\ell, \boldsymbol{\Sigma})$ and

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

- Our model:

$$\mathbf{x}_{\ell j} = \boldsymbol{\mu} + \boldsymbol{\tau}_\ell + \mathbf{e}_{\ell j}, \quad \ell = 1, \ldots, g, \quad j = 1, \ldots, n_\ell$$

  – For uniqueness (identifiability), we impose the constraint

$$\sum_{\ell=1}^{g} n_\ell \boldsymbol{\tau}_\ell = \mathbf{0}$$

- Decomposition of sample:

$$\underset{\substack{\uparrow \\ \text{observed}}}{\mathbf{x}_{\ell j}} = \underset{\substack{\uparrow \\ \text{overall} \\ \text{sample} \\ \text{mean} \\ \hat{\boldsymbol{\mu}}}}{\bar{\mathbf{x}}} + \underset{\substack{\uparrow \\ \text{estimated} \\ \text{treatment} \\ \text{effect} \\ \hat{\boldsymbol{\tau}}_\ell}}{(\bar{\mathbf{x}}_\ell - \bar{\mathbf{x}})} + \underset{\substack{\uparrow \\ \text{residual} \\ \hat{\boldsymbol{e}}_{\ell j}}}{(\mathbf{x}_{\ell j} - \bar{\mathbf{x}}_\ell)}$$

- Multivariate analog of total (corrected) sum of squares is

$$\underbrace{\sum_{\ell=1}^{g}\sum_{j=1}^{n_\ell}(\mathbf{x}_{\ell j}-\bar{\mathbf{x}})(\shortparallel)'}_{\substack{\text{total corrected}\\ \text{sum of squares}\\ \text{and cross}\\ \text{products matrix}}} = \underbrace{\sum_{\ell=1}^{g} n_\ell(\bar{\mathbf{x}}_\ell-\bar{\mathbf{x}})(\shortparallel)'}_{\substack{=\mathbf{H}\\ \text{``Between''}\\ \text{groups}\\ \text{matrix}}} + \underbrace{\sum_{\ell=1}^{g}\sum_{j=1}^{n_\ell}(\mathbf{x}_{\ell j}-\bar{\mathbf{x}}_\ell)(\shortparallel)'}_{\substack{=\mathbf{E}\\ \text{``Within'' groups}\\ \text{matrix}\\ =\sum_{\ell=1}^{g}(n_\ell-1)\mathbf{S}_\ell}}$$

Notes:

- Assuming no linear dependencies, $\text{rank}\{\mathbf{H}\} = \min(p, \nu_H)$
- $\mathbf{S}_\ell$ is the covariance matrix for the ℓ^{th} sample. So,

$$E\left\{\frac{1}{\left(\sum_{\ell=1}^{g} n_\ell\right)-g}\mathbf{E}\right\} = \mathbf{\Sigma}$$

  where $\text{rank}\{\mathbf{E}\} = \min(p, \nu_E)$

MANOVA TABLE (one-way)

Source	SS Matrix	d.f.
Treatment	$\mathbf{H}$	$\nu_H = g - 1$
Error	$\mathbf{E}$	$\nu_E = (\sum_{\ell=1}^{g} n_\ell) - g$
Total (corrected)	$\mathbf{H} + \mathbf{E}$	$(\sum_{\ell=1}^{g} n_\ell) - 1$

Wilks' Λ

The likelihood ratio test of $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \cdots = \boldsymbol{\mu}_g$ rejects H_0 when

$$\Lambda = \frac{|\mathbf{E}|}{|\mathbf{E} + \mathbf{H}|} \leq \Lambda_{\alpha,p,\nu_H,\nu_E}$$

- Note: Reject for small values of Λ. As in univariate anova F-test, we "accept" when total SS $(\mathbf{E} + \mathbf{H})$ is dominated by error $(\mathbf{E})$.
- Note: We sometimes refer to the subscripts of the Λ_{p,ν_H,ν_E} distribution as "dimension," "numerator df," and "denominator df" (e.g., $\Lambda_{\mathrm{dim},\mathrm{df}_{\mathrm{num}},\mathrm{df}_{\mathrm{den}}}$)

Properties of Wilk's Λ:

1. For statistic to be obtained, we need $\nu_E \geq p$.
2. Degrees of freedom ν_H and ν_E are the same as in analogous univariate case; e.g., one-way model: $\nu_H = g - 1$ and $\nu_E = \sum_{\ell=1}^{g} n_\ell - g$
3. Let $\lambda_1, \ldots, \lambda_s$ be the s non-zero eigenvalues of $\mathbf{E}^{-1}\mathbf{H}$, where $s = \min(p, \nu_H)$. Then $\Lambda = \prod_{i=1}^{s} \frac{1}{1+\lambda_i}$.
4. Critical value $\Lambda_{\alpha,p,\nu_H,\nu_E}$ decreases as p increases. Thus, adding variables decreases power unless variables contribute to separation.

5. When $\nu_H = 1$ or $\nu_H = 2$ or $p = 1$ or $p = 2$, Λ can be transformed to follow an F distribution.

- If $\nu_H = 1$

$$\frac{\nu_E - p + 1}{p} \quad \frac{1 - \Lambda}{\Lambda} \sim F_{p, \nu_E - p + 1}$$

- If $\nu_H = 2$

$$\frac{\nu_E - p + 1}{p} \quad \frac{1 - \sqrt{\Lambda}}{\sqrt{\Lambda}} \sim F_{2p, 2(\nu_E - p + 1)}$$

- If $p = 1$

$$\frac{\nu_E}{\nu_H} \quad \frac{1 - \Lambda}{\Lambda} \sim F_{\nu_H, \nu_E}$$

- If $p = 2$

$$\frac{(\nu_E - 1)}{\nu_H} \quad \frac{1 - \sqrt{\Lambda}}{\sqrt{\Lambda}} \sim F_{2\nu_H, 2(\nu_E - 1)}$$

6. Approximate tests

- For $p > 2$ or $\nu_H > 2$ and n large

$$\chi^2 = -\left[\nu_E - \frac{1}{2}(p - \nu_H + 1)\right] \ln \Lambda \overset{\text{approx}}{\sim} \chi^2_{p\,\nu_H}$$

  Approximately valid when $p^2 + \nu_H^2 \leq \frac{1}{3}\left[\nu_E - \frac{1}{2}(p - \nu_H + 1)\right]$

- More correct approximate distribution for Λ (exact when ν_H or p is 1 or 2):

$$F = \frac{1 - \Lambda^{1/t}}{\Lambda^{1/t}} \quad \frac{df_2}{df_1} \overset{\text{approx}}{\sim} F_{df_1, df_2}$$

  $df_1 = p\nu_H$

  $df_2 = wt - \frac{1}{2}(p\nu_H - 2)$

  $w = \nu_E + \nu_H - \frac{1}{2}(p + \nu_H + 1)$

$$t = \begin{cases} \sqrt{\frac{p^2\nu_H^2 - 4}{p^2 + \nu_H^2 - 5}} & \text{for } p^2 + \nu_H^2 - 5 > 0 \quad (\text{or} \quad p + \nu_H > 3) \\ 1 & \text{for } p^2 + \nu_H^2 - 5 \leq 0 \quad (\text{or} \quad p + \nu_H \leq 3) \end{cases}$$

Other MANOVA Tests

Let $(\lambda_1, \ldots, \lambda_s)$ be the ordered eigenvalues of $\mathbf{E}^{-1}\mathbf{H}$, where $s = \min(p, \nu_H) = \text{rank of } \mathbf{H}$

- Roy's Largest Root:

$$\theta = \lambda_1$$

  - Note: SAS and most authors denote Roy's Largest Root as λ_1 (the largest root of $\mathbf{E}^{-1}\mathbf{H}$). RC defines Roy's Largest Root as $\xi_1 = \frac{\lambda_1}{1+\lambda_1}$, which is the largest root of $(\mathbf{E}+\mathbf{H})^{-1}\mathbf{H}$.

– Approximate F-statistic (used by SAS):

$$F_\theta = \frac{(\nu_E - d + \nu_H)}{d}\lambda_1$$

is an upper bound for "true F" which is distributed

$$F_{d, \nu_E - d + \nu_H}$$

where $(d = \max(p, \nu_H))$

* Thus, F_θ-test is anti-conservative (yields lower bound on p-value)

– The eigenvector $\mathbf{a}_1$ corresponding to λ_1 comprises the discriminant function coefficients.

– For programs unable to obtain eigenvalues of nonsymmetric matrices, we can use the fact that λ_1 is a solution to both

$$(\mathbf{E}^{-1}\mathbf{H} - \lambda\mathbf{I})\mathbf{a} = \mathbf{0}$$

and

$$(\underbrace{\mathbf{E}^{-\frac{1}{2}}\mathbf{H}\mathbf{E}^{-\frac{1}{2}}}_{\text{“symmetric”}} - \lambda\mathbf{I}) \underbrace{\mathbf{E}^{\frac{1}{2}}\mathbf{a}}_{\substack{\text{“e'vector of} \\ \mathbf{E}^{-\frac{1}{2}}\mathbf{H}\mathbf{E}^{-\frac{1}{2}}\text{”}}} = \mathbf{0}$$

- Pillai's Trace:

$$V = \sum_{i=1}^{s} \frac{\lambda_i}{1+\lambda_i}$$
$$= \mathrm{tr}\left\{ (\mathbf{E}+\mathbf{H})^{-1}\,\mathbf{H} \right\} = \sum_{i=1}^{s} \xi_i$$

where $\xi_1, \ldots, \xi_s$ are the s ordered e'vals of $(\mathbf{E}+\mathbf{H})^{-1}\,\mathbf{H}$

- Note 1:

$$\left.\begin{array}{l} \mathbf{E}^{-1}\mathbf{H} \text{ is analagous to } \frac{\text{between SS}}{\text{within SS}} \\ (\mathbf{E}+\mathbf{H})^{-1}\mathbf{H} \text{ is analagous to } \frac{\text{between SS}}{\text{total SS}} \end{array}\right\} \begin{array}{c} \text{“Large Ratio”} \\ \Longrightarrow \\ \text{Reject } H_0 \end{array}$$

- Note 2:

$$\xi_i = \frac{\lambda_i}{1+\lambda_i} \text{ and } \lambda_i = \frac{\xi_i}{1-\xi_i}$$

Approximate F-statistic (used in SAS):

$$F_V = \frac{(2N+s+1)}{(2m+s+1)} \left(\frac{V}{s-V} \right) \sim F_{s(2m+s+1),s(2N+s+1)}$$

where

$$s = \min(\nu_H, p)$$
$$m = \tfrac{1}{2}\left(|\nu_H - p| - 1\right)$$
$$N = \tfrac{1}{2}(\nu_E - p - 1)$$

- Lawley-Hotelling Trace

$$
\begin{aligned}
U &= \sum_{i=1}^{s} \lambda_i \\
&= \mathrm{tr}\{\mathbf{E}^{-1}\mathbf{H}\}
\end{aligned}
$$

Approximate F-statistic (used in SAS):

$$F_u = \frac{2(sN+1)}{s^2(2m+s+1)} U \sim F_{s(2m+s+1),2(sN+1)}$$

$\rightarrow$ Also known as "Hotelling's generalized T^2"

Why four test statistics?

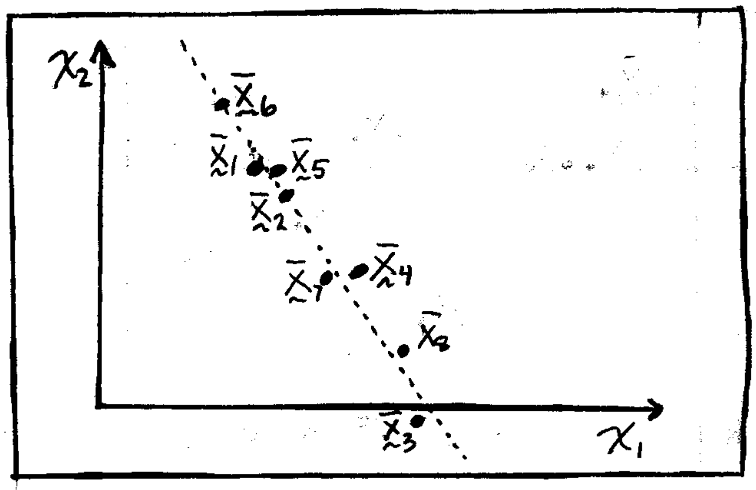
- All 4 are exact tests (i.e., have size α), but when H_0 not true they have different power
- For $p = 1, \mu_1, \ldots, \mu_k$ can be ordered along 1 dimension (line) and F-test is U.M.P.
- For $p > 1, \boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_k$ are points in $s = \min(p, \nu_H)$ dimensions. But means may in fact occupy only a subspace of the s dimensions; e.g., they may lie close to a line (1-D) or a plane (2-D).

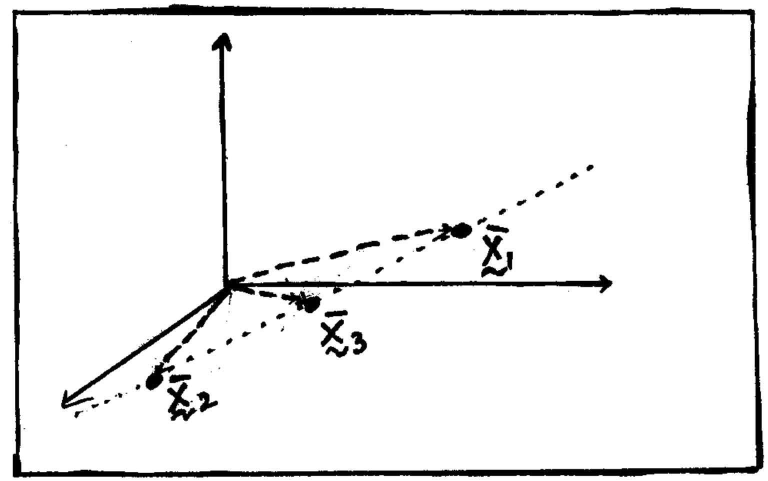
Examples:

$p = 2$ & $g = 8$ $(\nu_H = 7)$ $\qquad$ $p = 8$ & $g = 3$ $(\nu_H = 2)$

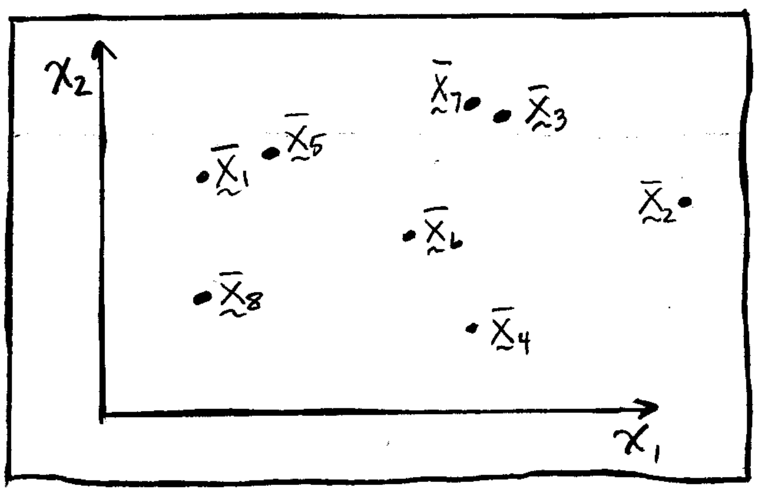
$s = 2$ $\qquad$ $s = 2$

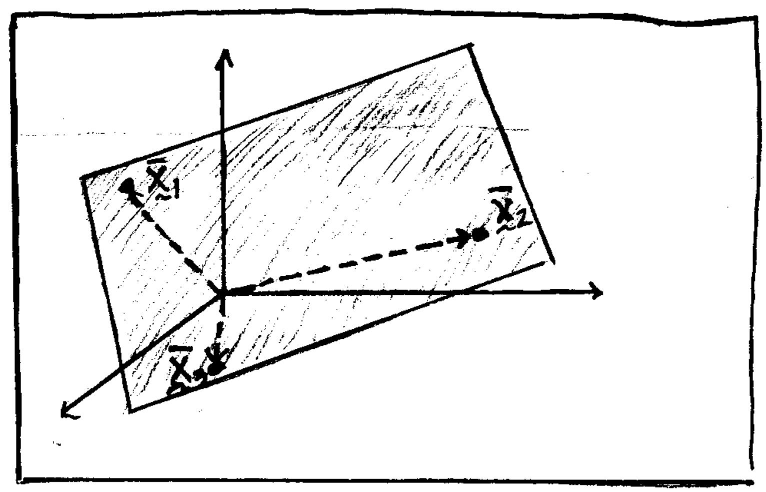
Collinear Means





Diffuse Means





Comparing Λ, Θ, V, & U:

Criterion	Ranking: Best ⟷ Worst
• Type I error rate under basic assumptions	All 4 statistics are same
• Power under basic assumptions & diffuse means	V Λ U Θ
• Power under basic assumptions & collinear means	Θ U Λ V
• Type I error rate with heterogeneous covariance matrices	V Λ U Θ

$\boxed{\text{ex}}$ Visual memory task

$x_1 = \%$ correct on positive stimulus questions

$x_2 = \%$ correct on negative stimulus questions

$g = 3$ (One healthy group and two impaired groups)

$\boxed{\text{ex}}$ Egyptian skulls

x_1 = maximum breadth of skull (mm)

x_2 = basibregmatic height of skull (mm)

x_3 = basialveolar length of skull (mm)

x_4 = nasal height of skull (mm)

$g = 3 \quad$ (4000 B.C., 3300 B.C., 1850 B.C.)

$\boxed{\text{ex}}$ Rootstock

x_1 = trunk girth at 4 years (mm $\times$ 100)

x_2 = extension growth at 4 years (m)

x_3 = trunk girth at 15 years (mm $\times$ 100)

x_4 = weight of tree above ground at 15 years (lb $\times$ 1000)

$g = 6$

Follow-up analyses

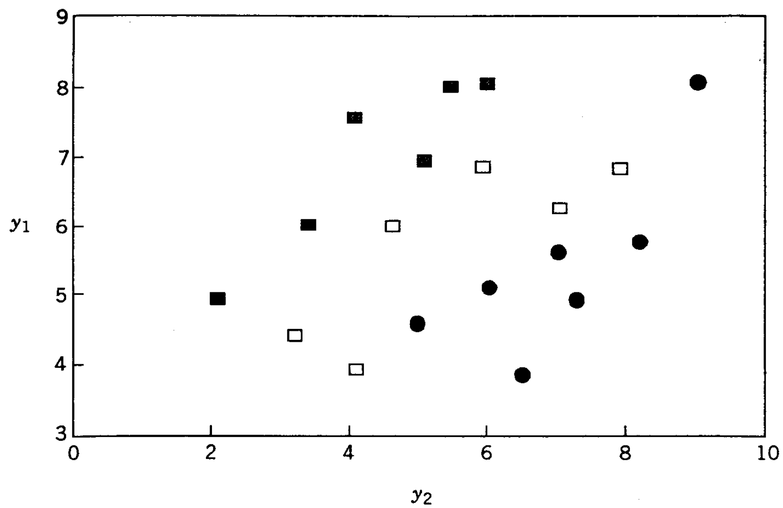


Figure 6.1 Three samples with significant Wilks' Λ but nonsignificant F's.
(Rencher, 1995)

Although only multivariate tests could detect group differences above, we still are often interested in follow-up analyses after conducting a multivariate analysis.

- Univariate hypothesis (F) tests
- Multivariate contrasts
- Confidence intervals/tests for $\mu_{ij} - \mu_{kj}$ (treatment differences for j^{th} variable)
- Analysis of discriminant function

Univariate F-tests

Often interested in univariate ANOVA for testing

$$H_{0,i}: \underset{\substack{\uparrow \\ \text{mean of} \\ i\text{th var.} \\ \text{for 1st} \\ \text{group}}}{\mu_{1i}} = \mu_{2i} = \cdots = \mu_{gi}, \qquad i = 1, \ldots, p$$

- Some advocate a "protected" univariate test approach:
  1. Conduct overall size α test of $H_0 : \boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_g$ using multivariate test (e.g. Λ)
  2. Test each of $H_{0,i}$ $(i = 1, ..., p)$ at level α only if multivariate test in step 1 rejects. [That is, when H_0 is "accepted" this approach automatically "accepts" $H_{0,1}, H_{0,2}, \ldots, H_{0,p}$.]

Defining our experiment by the p tests in step 2, the overall EER is (for independent variables when H_0 is true):

$$\begin{aligned}\Pr\{\text{at least one } H_{0,i} \text{ rejects}\} &= (\alpha)\,(1-(1-\alpha)^p) \\ &\leq \alpha\end{aligned}$$

What about properties of individual tests when H_0 is false??

Suppose:

$$\boldsymbol{\mu}_1 = \begin{bmatrix}\mu_{11} \\ \mu_{12} \\ \vdots \\ \mu_{1p}\end{bmatrix} = \underset{p\times 1}{\boldsymbol{\mu}} + \begin{bmatrix}\delta \\ 0 \\ \vdots \\ 0\end{bmatrix} \text{ and } \boldsymbol{\mu}_i = \begin{bmatrix}\mu_{i1} \\ \mu_{i2} \\ \vdots \\ \mu_{ip}\end{bmatrix} = \boldsymbol{\mu}, i = 2, \ldots, g$$

Let δ be some value such that our test of $H_0 : \boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_p$ using Λ has power $= .50$.

Consider the "partial" experiment defined by the $p-1$ tests of $H_{0,2}, H_{0,3}, \ldots, H_{0,p}$.

The partial EER for this scenario (assuming independence) is

$$\Pr \underbrace{\{\text{at least one rejection among } H_{0,2}, \ldots, H_{0,p}\}}_{\text{“}A\text{”}}$$

$$\begin{aligned} &= \Pr\{\text{“}A\text{”} | \Lambda \text{ rejects}\} \cdot \Pr\{\Lambda \text{ rejects}\} \\ &= \left[1-(1-\alpha)^{p-1}\right] \cdot (.50) \end{aligned}$$

Thus, the partial EER can be dramatically larger than α

$\boxed{\text{ex}}$ $p = 10, \alpha = .05 \Rightarrow$ partial EER $\cong .20$

Conclusion:

"Protected F test" approach protects overall EER, but may have poor properties for other inferences

$\rightarrow$ Consider tests at $\frac{\alpha}{p}$ level

Contrasts (multivariate)

- Already considered contrasts of the type $\underset{q\times p}{\mathbf{C}}\underset{p\times 1}{\boldsymbol{\mu}}$ for testing

  $H_0 : \mathbf{C}\boldsymbol{\mu} = 0$, where each row of $\mathbf{C}$ sums to 0

  $\boxed{\text{ex}}$ Linear trend among 4 observations?

$$H_0 : \begin{bmatrix} -3 & -1 & 1 & 3 \end{bmatrix} \begin{bmatrix} \mu_{\text{day1}} \\ \mu_{\text{day2}} \\ \mu_{\text{day3}} \\ \mu_{\text{day4}} \end{bmatrix}$$

- Here consider contrasts of the type

$$\boldsymbol{\delta} = c_1\boldsymbol{\mu}_1 + c_2\boldsymbol{\mu}_2 + \cdots + c_g\boldsymbol{\mu}_g = \mathbf{Mc}$$

  where $\mathbf{M} = \underbrace{\begin{bmatrix} \boldsymbol{\mu}_1 & \boldsymbol{\mu}_2 & \cdots & \boldsymbol{\mu}_g \end{bmatrix}}_{p\times g}$

$$\hat{\boldsymbol{\delta}} = c_1 \bar{\mathbf{x}}_1 + c_2 \bar{\mathbf{x}}_2 + \cdots + c_g \bar{\mathbf{x}}_g$$

$$\mathrm{var}\{\hat{\boldsymbol{\delta}}\} = \sum_{i=1}^{g} {c_i}^2 \frac{\boldsymbol{\Sigma}}{n_i} = \left(\sum_{i=1}^{g} \frac{{c_i}^2}{n_i} \right) \boldsymbol{\Sigma}$$

$$\widehat{\mathrm{var}}\{\hat{\boldsymbol{\delta}}\} = \left(\sum_{i=1}^{g} \frac{{c_i}^2}{n_i} \right) \mathbf{S}_{p\ell}$$

where $\mathbf{S}_{p\ell} = \frac{1}{\nu_E}\mathbf{E}$ and $\nu_E = \sum_{i=1}^{g}(n_i - 1)$.

So, our test is based on

$$T^2 = \hat{\boldsymbol{\delta}}' \left[\widehat{\mathrm{var}} \left\{ \hat{\boldsymbol{\delta}} \right\} \right]^{-1} \hat{\boldsymbol{\delta}} \sim T^2_{p,\nu_E}$$

or define

$$\mathbf{H}_1 = \frac{1}{\sum_{i=1}^{g} \frac{c_i^2}{n_i}} \hat{\boldsymbol{\delta}} \hat{\boldsymbol{\delta}}' \text{ and } \Lambda = \frac{|\mathbf{E}|}{|\mathbf{E} + \mathbf{H}_1|} \sim \Lambda_{p,1,\nu_E}$$

$\boxed{\text{ex}}$ Rootstock

Confidence intervals for treatment differences $(\mu_{ij} - \mu_{kj})$

Interest in components of difference between groups

$$\boldsymbol{\mu}_i - \boldsymbol{\mu}_k = \boldsymbol{\tau}_i - \boldsymbol{\tau}_k$$

Specifically, interested in the jth component of this difference vector

$$\mu_{ij} - \mu_{kj} = \tau_{ij} - \tau_{kj}$$

which is estimated by $\bar{x}_{ij} - \bar{x}_{kj}$

Because we often want to obtain confidence intervals for all $g(g-1)/2$ pairwise comparisons for each of p variables simultaneously, we use a Bonferroni adjustment to protect overall EER.

$100(1-\alpha)\%$ (Simultaneous) Confidence Interval for $\mu_{ij} - \mu_{kj}$ is:

$$(\bar{x}_{ij} - \bar{x}_{kj}) \pm t_{\underbrace{[\alpha/(pg(g-1))]}_{\substack{\frac{\alpha}{2} \text{ divided by} \\ \# \text{ of comparisons} \\ = pg(g-1)/2}}, [\sum_{i=1}^{g}(n_i-1)]} \sqrt{s_{pl,jj}\left(\frac{1}{n_i} + \frac{1}{n_k}\right)}$$

where $s_{pl,jj}$ is the j^{th} diagonal element of $\mathbf{S}_{p\ell} = \mathbf{E}/\left(\sum_{i=1}^{g}(n_i-1)\right)$

Warning for SAS implementation

SAS uses upper $\frac{\alpha}{g(g-1)}$ quantile instead of upper $\frac{\alpha}{pg(g-1)}$ quantile of t distribution

(Bonferroni intervals are part of univariate output.)

$\rightarrow$ Adjust by specifying ALPHA in MEANS statement

$\boxed{\text{ex}}$ $p = 3, g = 5$, desired overall EER $= .05$

```
proc glm;
   class group;
   model y1 y2 y3 = group;
   means group/bon alpha = .016667   <=   .05/p ;
run;
```

$\boxed{\text{ex}}$ Rootstock

Analysis of discriminant function

(More detail to come in Section V of the course)

$g = 2$ case:

Choose $\mathbf{a}$ to maximize (for $\mathbf{a} \neq \mathbf{0}$):

$$\frac{\left[\mathbf{a}'\left(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2\right)\right]^2}{\mathbf{a}'\mathbf{S}_{p\ell}\mathbf{a}} = \frac{\mathbf{a}'\left(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2\right)\left(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2\right)'\mathbf{a}}{\mathbf{a}'\mathbf{S}_{p\ell}\mathbf{a}}$$

$\Longrightarrow \mathbf{a} = \mathbf{S}_{p\ell}^{-1}\left(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2\right)$

$g > 2$ case:

Choose $\mathbf{a}$ to maximize (for $\mathbf{a} \neq \mathbf{0}$):

$$\lambda_1 = \frac{\mathbf{a}'\mathbf{H}\mathbf{a}}{\mathbf{a}'\mathbf{E}\mathbf{a}}$$

$\Longrightarrow \lambda_1 =$ largest e'value of $\mathbf{E}^{-1}\mathbf{H}$ and $\mathbf{a}_1$ is corresp. e'vec

- Relative importance of 1^{st} disc fcn $= \frac{\lambda_1}{\sum_{i=1}^{s} \lambda_i}$

$\boxed{\text{ex}}$ Rootstock

	Girth4	Growth	Girth15	Weight
Univariate F	1.93	2.91	11.97	12.16

$$\mathbf{a} = [.4703 \quad -.2627 \quad .6532 \quad -.0738]$$

Recall that $a_i^* = a_i\sqrt{s_{p\ell,ii}} = a_i \times \sqrt{\frac{1}{\nu_E}e_{ii}}$, where e_{ii} is the ith diagonal element of $\mathbf{E}$, so

$$\begin{aligned}\mathbf{a}^* &= \frac{1}{\sqrt{42}}\left[.4703\sqrt{.3200} \quad -.2627\sqrt{12.1428} \quad .6532\sqrt{4.2908} \quad -.0738\sqrt{1.712}\right] \\ &= [.0411 \quad -.1413 \quad .2088 \quad -.0149]\end{aligned}$$

Test for statistical significance of final m discriminant functions:

$$\Lambda_m = \prod_{i=m}^{s} \frac{1}{1+\lambda_i} \sim \Lambda_{p-m+1,\nu_H-m+1,\nu_E-m+1}$$

Tests for additional information

Let $\mathbf{x}_{\ell j} = \begin{bmatrix} \mathbf{y}_{\ell j} \\ \mathbf{z}_{\ell j} \end{bmatrix} \begin{matrix} \leftarrow & p \times 1 \\ \leftarrow & q \times 1 \end{matrix}$, $j = 1, \ldots, n_\ell$

be $(p+q)$-variate observations from the ℓ^{th} group

- Wish to determine if $\mathbf{z}$ makes a significant contribution beyond $\mathbf{y}$ in detecting separation of groups

Calculate:

$$\underset{(p+q)\times(p+q)}{\mathbf{E}} = \begin{bmatrix} \mathbf{E}_{yy} & \mathbf{E}_{yz} \\ \mathbf{E}_{zy} & \mathbf{E}_{zz} \end{bmatrix} \text{ and } \underset{(p+q)\times(p+q)}{\mathbf{H}} = \begin{bmatrix} \mathbf{H}_{yy} & \mathbf{H}_{yz} \\ \mathbf{H}_{zy} & \mathbf{H}_{zz} \end{bmatrix}$$

$$\Lambda_{yz} = \frac{|\mathbf{E}|}{|\mathbf{E} + \mathbf{H}|}$$

$$\Lambda_y = \frac{|\mathbf{E}_{yy}|}{|\mathbf{E}_{yy} + \mathbf{H}_{yy}|}$$

Test of additional info:

$$\underset{\substack{\uparrow \\ \text{“partial } \Lambda \\ \text{statistic”}}}{\Lambda_{z|y}} = \frac{\Lambda_{yz}}{\Lambda_y} \sim \Lambda_{\underset{\substack{\uparrow \\ \text{\# of} \\ \text{vars} \\ \text{in } \mathbf{z}}}{q},\, \nu_H,\, \nu_E - \underset{\substack{\uparrow \\ \text{\# of} \\ \text{vars} \\ \text{in } \mathbf{y}}}{p}}$$

Two-Way MANOVA (fixed-effects)

Model:

$$\mathbf{x}_{ijk} = \boldsymbol{\mu} + \boldsymbol{\alpha}_i + \boldsymbol{\beta}_j + \boldsymbol{\gamma}_{ij} + \mathbf{e}_{ijk}$$

$i = 1, \ldots, a$

$j = 1, \ldots, b$

$k = 1, \ldots, n$ (for simplicity, assume $n_{ij} = n \;\forall\; i, j$)

- $\sum_{i=1}^{a} \boldsymbol{\alpha}_i = \sum_{j=1}^{b} \boldsymbol{\beta}_i = \sum_{i=1}^{a} \boldsymbol{\gamma}_{ij} = \sum_{j=1}^{b} \boldsymbol{\gamma}_{ij} = \mathbf{0}$
- Assume $\mathbf{e}_{ijk} \sim N_p(\mathbf{0}, \boldsymbol{\Sigma})$

Source	Sum of Sqrares and Products	df
A	$\underset{\sim}{H}_A = nb\sum_{i=1}^{a}(\bar{\underset{\sim}{x}}_{i\cdot} - \bar{\underset{\sim}{x}})(\bar{\underset{\sim}{x}}_{i\cdot} - \bar{\underset{\sim}{x}})'$	$a-1$
B	$\underset{\sim}{H}_B = na\sum_{j=1}^{b}(\bar{\underset{\sim}{x}}_{\cdot j} - \bar{\underset{\sim}{x}})(\bar{\underset{\sim}{x}}_{\cdot j} - \bar{\underset{\sim}{x}})'$	$b-1$
AB (interaction)	$\underset{\sim}{H}_{AB} = n\sum_{i=1}^{a}\sum_{j=1}^{b}(\bar{\underset{\sim}{x}}_{ij} - \bar{\underset{\sim}{x}}_{i\cdot} - \bar{\underset{\sim}{x}}_{\cdot j} + \bar{\underset{\sim}{x}})(\;''\;)'$	$(a-1)(b-1)$
Error	$\underset{\sim}{E} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(\underset{\sim}{x}_{ijk} - \bar{\underset{\sim}{x}}_{ij})(\underset{\sim}{x}_{ijk} - \bar{\underset{\sim}{x}}_{ij})'$	$ab(n-1)$
Total (corrected)	$\underset{\sim}{H}_A + \underset{\sim}{H}_B + \underset{\sim}{H}_{AB} + \underset{\sim}{E} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{n}(\underset{\sim}{x}_{ijk} - \bar{\underset{\sim}{x}})(\;''\;)'$	$abn-1$

- $\bar{\mathbf{x}}_{i.}$ = average over ith level of factor A
- $\bar{\mathbf{x}}_{.j}$ = average over jth level of factor B
- $\bar{\mathbf{x}}_{ij}$ = average over ith level of A and jth level of B

Test for A, B, and AB (interaction):

$$\Lambda_A = \frac{|\mathbf{E}|}{|\mathbf{E}+\mathbf{H}_A|} \sim \Lambda_{p,a-1,ab(n-1)}$$

$$\Lambda_B = \frac{|\mathbf{E}|}{|\mathbf{E}+\mathbf{H}_B|} \sim \Lambda_{p,b-1,ab(n-1)}$$

$$\Lambda_{AB} = \frac{|\mathbf{E}|}{|\mathbf{E}+\mathbf{H}_{AB}|} \sim \Lambda_{p,(a-1)(b-1),ab(n-1)}$$

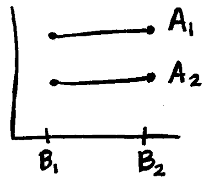
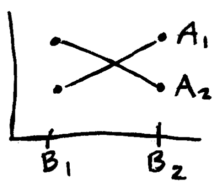
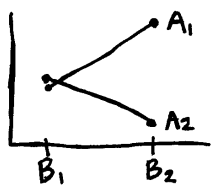
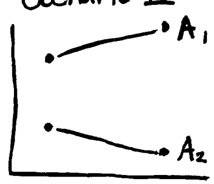
Follow-up analyses

- Individual F-tests (univariate anova's)
  - As You Might Expect (AYME)
- Contrasts
  - AYME
- C.I.'s for treatment effects
  - AYME
- Analysis of discriminant function
  - AYME
    - For analyzing contribution of p variables to separation of levels of A use first dicrim. function (e'vector) of $\mathbf{E}^{-1}\mathbf{H}_A$
    - Analyzing levels of $B \Rightarrow$ use $\mathbf{E}^{-1}\mathbf{H}_B$
    - Analyzing levels of $AB \Rightarrow$ use $\mathbf{E}^{-1}\mathbf{H}_{AB}$

An interlude about interactions...

Suppose we have two levels of factor A and two levels of factor B:

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk}, \qquad i = 1, 2,\ j = 1, 2$$

	A significant?	B sig.?	Interax sig.?
Scenario I 	Yes	No	No
Scenario II 	No	No	Yes
Scenario III 	Yes	No	Yes
Scenario IV 	Yes	No	Yes

Is Main Effect for A interpretable in Scenarios II, III, and IV?

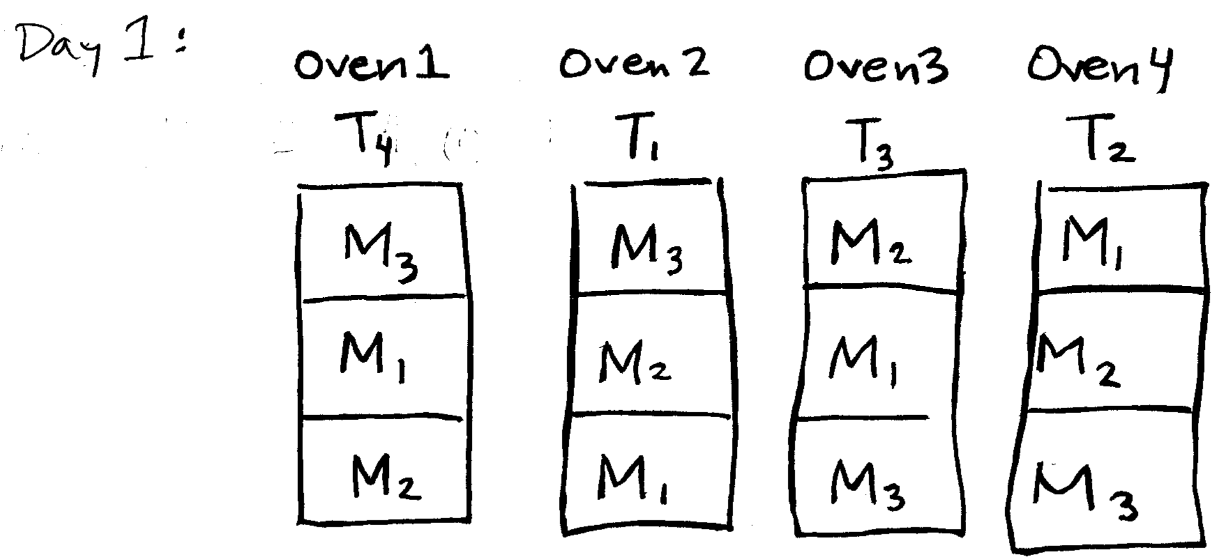
- Yes, if "significance" simply refers to size of effect $\alpha_1 - \alpha_2$ (i.e., effect of A *averaged over levels of B*).
  - "Significant" doesn't mean "one level is best"
  - "Significance of Main Effect for A" is affected by number of levels of B and sample size for each level

Mixed Model MANOVA (Split-plot)

$\boxed{\text{ex}}$

4	temperatures	$(T_i, i = 1, \ldots, 4)$	$t = 4$
3	days	$(D_j, j = 1, \ldots, 3)$	$d = 3$
3	metal alloys	$(M_k, k = 1, \ldots, 3)$	$m = 3$

$\mathbf{x}_{ijk}$ is a p-variate response of metal strength



Source	df	Tests
T	$t-1=3$	$\lvert \underset{\sim}{H}_D \rvert / \lvert \underset{\sim}{H}_D + \underset{\sim}{H}_T \rvert$
D	$t(d-1)=8$	
M	$m-1=2$	$\lvert \underset{\sim}{E} \rvert / \lvert \underset{\sim}{E} + \underset{\sim}{H}_M \rvert$
TM (interaction)	$(t-1)(m-1)=6$	$\lvert \underset{\sim}{E} \rvert / \lvert \underset{\sim}{E} + \underset{\sim}{H}_{TM} \rvert$
Error	$t(d-1)(m-1)=16$	

- Need $t(d-1)(m-1) \geq p$ to make test
- Need $t(d-1) \geq p$ to make test
  - → alternative: Conduct p univariate anova's testing each variable at α/p

III.B.iv. Profile Analysis

p-variate response consists of tests, questions, etc. measured on members of g groups.

$\boxed{\text{ex}}$ Guinea pigs on three diets

- Weights measured at ends of week 1, 3, 4, 5, 6, & 7

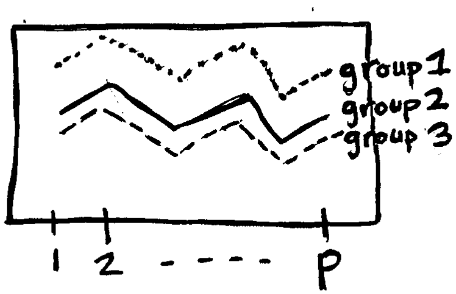
Break hypothesis:

$$H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \cdots = \boldsymbol{\mu}_g$$

into three more specific hypotheses:

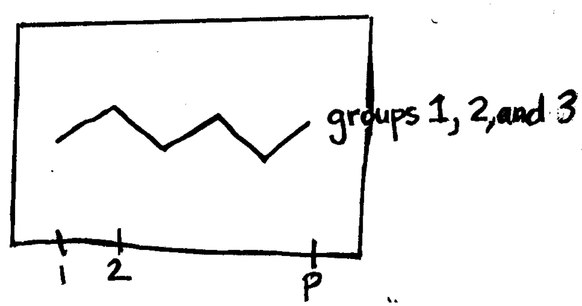
$$H_{01} : \text{"The } g \text{ profiles are parallel"}$$

$\boxed{\text{ex}}$ $H_{0,1}$ true might yield a profile plot like:



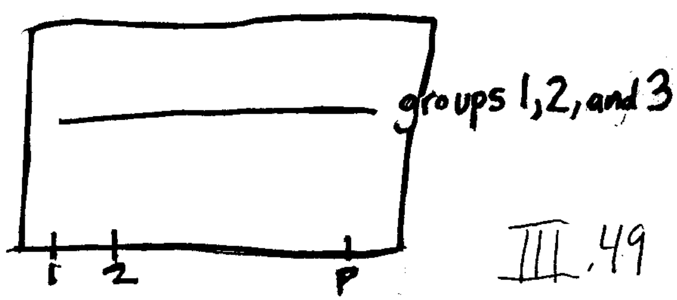
H_{02} : "The g profiles are at same level"

$\boxed{\text{ex}}$ $H_{0,1}$ and $H_{0,2}$ true might yield a profile plot like:



H_{03} : "The g profiles are flat"

$\boxed{\text{ex}}$ $H_{0,1}$, $H_{0,2}$, and $H_{0,3}$ yields the profile plot:



III.49

Formalizing the null hypotheses

- "Parallelism": Difference in responses between any time points is the same for all groups.

$$H_{01}: \quad \mu_{1j}-\mu_{1(j-1)} = \mu_{2j}-\mu_{2(j-1)} = \cdots = \mu_{gj}-\mu_{g(j-1)} \text{ for } j = 2, \ldots, p$$

OR

$$\mathbf{C}\boldsymbol{\mu}_1 = \mathbf{C}\boldsymbol{\mu}_2 = \cdots = \mathbf{C}\boldsymbol{\mu}_g$$

$$\text{where } \underset{(p-1)\times p}{\mathbf{C}} = \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & & -1 & 1 \end{bmatrix}$$

or $\mathbf{C}$ can be any other full row rank $(p-1)\times p$ matrix such that $\mathbf{C}\mathbf{1} = \mathbf{0}$

- "Same level": Total (or average) response (over time) is the same for all groups.

$$H_{02} : \mathbf{1}'\boldsymbol{\mu}_\mathbf{1} = \mathbf{1}'\boldsymbol{\mu}_\mathbf{2} = \cdots = \mathbf{1}'\boldsymbol{\mu}_\mathbf{g}$$

  – Note: If H_{01} holds, we can also refer to H_{02} as the hypothesis of "coincident profiles" and H_{02} can be written:

$$H_{02} : \mu_{1j} = \mu_{2j} = \cdots = \mu_{gj} \text{ for } j = 1, \ldots, p$$

- "Flatness": No change in response (over time) for the profiles — responses at each time (averaged across groups) are the same.

$$H_{03} : \frac{\mu_{11} + \mu_{21} + \cdots + \mu_{g1}}{g} = \cdots = \frac{\mu_{1p} + \mu_{2p} + \cdots + \mu_{gp}}{g}$$

OR

$$\mathbf{C} \underbrace{\left(\frac{\boldsymbol{\mu}_1 + \cdots + \boldsymbol{\mu}_g}{g} \right)}_{\text{“}\bar{\boldsymbol{\mu}}\text{”} \leftarrow \text{ average profile}} = \mathbf{O} \text{ or } \mathbf{C}\bar{\boldsymbol{\mu}} = \mathbf{O}$$

  - Note: If H_{01} and H_{02} hold, H_{03} can also be written $H_{03} : \mu_{11} = \mu_{12} = \cdots = \mu_{1p} = \mu_{21} = \cdots = \mu_{gp}$ that is, all pg response means are equal.

Tests

Test for H_{01}:

$$\Lambda = \frac{|\mathbf{CEC}'|}{|\mathbf{C}(\mathbf{E}+\mathbf{H})\mathbf{C}'|} \sim \Lambda_{p-1,\nu_H,\nu_E}$$

Test for H_{02}:

$$\Lambda = \frac{\mathbf{1}'\mathbf{E}\mathbf{1}}{\mathbf{1}'\mathbf{E}\mathbf{1}+\mathbf{1}'\mathbf{H}\mathbf{1}} \sim \Lambda_{1,\nu_H,\nu_E}$$

$$\Rightarrow \frac{1-\Lambda}{\Lambda}\frac{\nu_E}{\nu_H} \sim F_{\nu_H,\nu_E}$$

Test for H_{03}:

$$T^2 = \left(\sum_{\ell=1}^{g} n_\ell\right)(\mathbf{C}\bar{\mathbf{x}})'\left(\frac{1}{\nu_e}\mathbf{CEC}'\right)^{-1}\mathbf{C}\bar{\mathbf{x}} \sim T^2_{p-1,\nu_E}$$

$$\Rightarrow \frac{\nu_E-(p-1)+1}{\nu_E(p-1)}T^2 \sim F_{p-1,\nu_E-(p-1)+1}$$

$\boxed{\text{ex}}$ Guinea Pigs

H_{01}: parallel?

H_{02}: same level?

H_{03}: flat?

III.B.v. Repeated Measures

- Similarities to “profile analysis”
- Each subject measured under several treatments or time points
- Comparing means of treatments applied to each subject: within-subjects tests
- Comparing levels of factors assigned to groups of subjects: between-subjects tests

Structure of g-groups R.M. experiment

Factor B (between-subjects)	Subjects	Factor A (within-subjects) A_1	A_2	$\cdots$	A_p		
B_1	S_{11}	$(x_{111}$	x_{112}	$\cdots$	$x_{11p})$	$=$	$\underset{\sim}{X}_{11}'$
	$\vdots$						$\vdots$
	S_{1n}	$(x_{1n1}$	x_{1n2}	$\cdots$	$x_{1np})$	$=$	$\underset{\sim}{X}_{1n}'$
$\vdots$	$\vdots$						$\vdots$
B_g	S_{g1}	$(x_{g11}$	x_{g12}	$\cdots$	$x_{g1p})$	$=$	$\underset{\sim}{X}_{g1}'$
	$\vdots$						
	S_{gn}	$(x_{gn1}$	x_{gn2}	$\cdots$	$x_{gnp})$	$=$	$\underset{\sim}{X}_{gn}'$

Univariate model: (split-plot):

$$x_{ijr} = \mu + \underbrace{B_i}_{\text{between}} + S_{(i)j} + \underbrace{A_r}_{\text{within}} + \underbrace{BA_{ir}}_{\text{interaction}} + \varepsilon_{ijr}$$

$\boxed{\text{ANOVA Table}}$

Source	df	MS
B (between)	$g-1$	MSB
S (subjects)	$g(n-1)$	MSS
A (within)	$p-1$	MSA
BA	$(g-1)(p-1)$	MSBA
Error (SA interax)	$g(n-1)(p-1)$	MSE

Are these F-tests valid when the p within-subject observations $x_{ij1}, \ldots, x_{ijp}$ are correlated?

- Standard univariate assumption:

$$\text{var}\{\mathbf{x}_{ij}\} = \boldsymbol{\Sigma} = \sigma^2 \mathbf{I}_p \;\; \forall\, i, j \qquad \leftarrow \text{ "sphericity"}$$

Univariate F-tests still valid as long as

$$\underset{\substack{\uparrow \\ (p-1)\times p \\ \text{orthonormal} \\ \text{contrast} \\ \text{matrix}}}{\mathbf{C}}\boldsymbol{\Sigma}\mathbf{C}' = \sigma^2 \mathbf{I}$$

This condition is often called "sphericity" (but we'll say "generalized sphericity" for clarity)

$\boxed{\text{ex}}$ For $p = 4$, we could use

$$\mathbf{C} = \begin{bmatrix} 3/\sqrt{12} & -1/\sqrt{12} & -1/\sqrt{12} & -1/\sqrt{12} \\ 0 & 2/\sqrt{6} & -1/\sqrt{6} & -1/\sqrt{6} \\ 0 & 0 & 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}$$

Special case of $\mathbf{C}\boldsymbol{\Sigma}\mathbf{C}' = \sigma^2\mathbf{I}$:

$$\boldsymbol{\Sigma} = \sigma^2 \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \ddots & \ddots & \vdots \\ \rho & \rho & \cdots & 1 \end{bmatrix} = \sigma^2 \left[(1-\rho)\mathbf{I} + \rho\mathbf{1}\mathbf{1}'\right]$$

– This case is called "compound symmetry"

Univariate strategies

1. Assume "generalized sphericity"

   Fehlberg (1980): Use $H_0 : \mathbf{C\Sigma C}' = \sigma^2\mathbf{I}$ preliminary test using $\alpha = .40$. [This test to be discussed later in the course]

   If hypothesis is "accepted," use standard F-tests …

   … for A: $F = \dfrac{MSA}{MSE} \sim F_{p-1,g(n-1)(p-1)}$

   … for AB: $F = \dfrac{MSAB}{MSE} \sim F_{(g-1)(p-1),g(n-1)(p-1)}$

   BUT, even mild departures from $\mathbf{C\Sigma C}' = \sigma^2\mathbf{I}$ can seriously inflate Type I error (Boik, *Psychometrika*, 1981).

2. Conservative test:

   … for A: $F = \dfrac{MSA}{MSE} \dot\sim F_{1,g(n-1)}$

   … for AB: $F = \dfrac{MSAB}{MSE} \dot\sim F_{(g-1),g(n-1)}$

- Too conservative

3. Adjusted F-tests

- A compromise between approaches 1 and 2 when sphericity violated.
- Greenhouse and Geisser (1959) recommend approximate F-tests involving within-subjects factor which reduce numerator and denominator d.f. by a factor of

$$\varepsilon = \frac{\left[\mathrm{tr}\left(\boldsymbol{\Sigma} - \frac{1}{p}\mathbf{1}\mathbf{1}'\boldsymbol{\Sigma}\right)\right]^2}{(p-1)\mathrm{tr}\left[\left(\boldsymbol{\Sigma} - \frac{1}{p}\mathbf{1}\mathbf{1}'\boldsymbol{\Sigma}\right)^2\right]} \qquad \boxed{\text{SAS: “G – G } \varepsilon\text{”}}$$

  – To estimate ε, use $\hat{\boldsymbol{\Sigma}} = \frac{\mathbf{E}}{\nu_E}$

  – F-tests ...

  ... for A: $F = \frac{MSA}{MSE} \sim F_{\hat{\varepsilon}(p-1),\hat{\varepsilon}g(n-1)(p-1)}$

  ... for AB: $F = \frac{MSAB}{MSE} \sim F_{\hat{\varepsilon}(g-1)(p-1),\hat{\varepsilon}g(n-1)(p-1)}$

– ε and $\hat{\varepsilon} \in \Big(\underset{\substack{\uparrow \\ \text{general} \\ \text{(non-spherical)} \\ \boldsymbol{\Sigma}}}{\frac{1}{p-1}} \quad , \quad \underset{\substack{\uparrow \\ \text{sphericity} \\ \text{holds}}}{1} \Big)$

– Approach is generally too conservative, especially for small n

- Huynh and Feldt (1976) give another expression for ε $\boxed{\text{SAS: “H – F } \varepsilon\text{”}}$
  – Less conservative
  – “H – F ε” can exceed 1 $\Rightarrow$ set equal to 1

Multivariate Model:

$\mathbf{x}_{ij} = \boldsymbol{\mu} + \boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_{ij}$

- Notes about $\boldsymbol{\beta}_i$
  - $\boldsymbol{\beta}_i$ is a p-vector of main effects for group i
  - Tests on factor A (within subjects) and AB interaction constructed with contrasts of $\boldsymbol{\beta}_i$ (as in profile analysis)
- Standard multivariate assumption:

$$\text{var}\{\mathbf{x}_{ij}\} = \boldsymbol{\Sigma} \quad \forall\ i, j$$

  Note: $\boldsymbol{\Sigma}$ is completely unrestricted (no sphericity requirement, etc.)
- Several similarities with g groups profile analysis

Contrast Matrices in SAS Proc GLM (“**`repeated`**” statement)

– Assume p=5 times/variables

`contrast` or **`contrast(5)`**:

$$\begin{bmatrix} 1 & 0 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 1 & -1 \end{bmatrix}$$

`contrast(2)`:

$$\begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 & 1 \end{bmatrix}$$

`polynomial`

$$\begin{bmatrix} -2 & -1 & 0 & 1 & 2 \\ 2 & -1 & -2 & -1 & 2 \\ 1 & -2 & 0 & 2 & -1 \\ 1 & -4 & 6 & -4 & 1 \end{bmatrix}$$

`‘‘repeated time 5 (1 2 5 10 20) polynomial’’`:

$$\begin{bmatrix} -.43 & -.36 & -.17 & .15 & .80 \\ .43 & .21 & -.33 & -.71 & .39 \\ -.43 & .14 & .73 & -.51 & .08 \\ .49 & -.78 & .37 & -.09 & .01 \end{bmatrix}$$

`helmert`:

$$\begin{bmatrix} 1 & -.25 & -.25 & -.25 & -.25 \\ 0 & 1 & -.33 & -.33 & -.33 \\ 0 & 0 & 1 & -.5 & -.5 \\ 0 & 0 & 0 & 1 & -1 \end{bmatrix}$$

$\leftarrow$ The helmert contrast matrix identifies the time at which the treatments cease to change or plateau

`mean` or `mean(5)`

$$\begin{bmatrix} 1 & -.25 & -.25 & -.25 & -.25 \\ -.25 & 1 & -.25 & -.25 & -.25 \\ -.25 & -.25 & 1 & -.25 & -.25 \\ -.25 & -.25 & -.25 & 1 & -.25 \end{bmatrix}$$

`profile`:

$$\begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{bmatrix}$$

- Test for A (within subjects)
  - Analogous to "flatness" test in profile analysis
  - Want to compare means for $x_1, \ldots, x_p$ averaged across levels of B

$$\text{Let } \bar{\boldsymbol{\mu}} = \sum_{i=1}^{g} \boldsymbol{\mu}_i / g = (\mu_{\cdot 1}, \ldots, \mu_{\cdot p})'$$

$$H_0 : \mu_{\cdot 1} = \ldots = \mu_{\cdot p} \text{ or } \mathbf{C}\bar{\boldsymbol{\mu}} = \mathbf{O}$$

$$\boxed{\text{ex}} \quad \mathbf{C} = \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{bmatrix} \begin{array}{l} \text{or some similar} \\ \text{contrast matrix} \end{array}$$

Test statistic for A:

$$T^2 = \underset{\substack{\uparrow \\ \sum_{i=1}^{g} n_i}}{N} (\mathbf{C} \underset{\substack{\uparrow \\ \text{grand} \\ \text{mean}}}{\bar{\mathbf{x}}})'(\mathbf{C} \underbrace{\mathbf{S}_{p\ell}}_{\frac{\mathbf{E}}{\nu_E}} \mathbf{C}')^{-1}(\underbrace{\mathbf{C}}_{(p-1)\times p} \bar{\mathbf{x}}) \sim T^2_{p-1,\nu_E}$$

$$\frac{\nu_E - (p-1) + 1}{\nu_E(p-1)} T^2 \sim F_{p-1,\nu_E-(p-1)+1}$$

OR

$\Lambda = \frac{|\mathbf{CEC}'|}{|\mathbf{C}(\mathbf{E}+\mathbf{H}^*)\mathbf{C}'|} \sim \Lambda_{p-1,1,\nu_E}$
where $\mathbf{H}^* = N\bar{\mathbf{x}}\bar{\mathbf{x}}'$ is from the partitioning

$$\sum_{i=1}^{g} \sum_{j=1}^{n_1} \mathbf{x}_{ij}\mathbf{x}'_{ij} = \mathbf{E} + \mathbf{H} + N\bar{\mathbf{x}}\bar{\mathbf{x}}'$$

- Test for B (between subjects)
  - Analogous to "same level" test in profile analysis
  - Want to compare group means (averaging over p levels of A)

$$H_0 : \mathbf{1}'\boldsymbol{\mu}_1 = \cdots = \mathbf{1}'\boldsymbol{\mu}_g \text{ or } \underbrace{\frac{1}{\sqrt{p}}\mathbf{1}'\boldsymbol{\mu}_1 = \cdots = \frac{1}{\sqrt{p}}\mathbf{1}'\boldsymbol{\mu}_g}_{\text{SAS}}$$

    * That is, we can just conduct one-way ANOVA on $z_{ij} = \mathbf{1}'\mathbf{x}_{ij}$, so the test statistic for B is

$$\Lambda = \frac{\mathbf{1}'\mathbf{E}\mathbf{1}}{\mathbf{1}'\mathbf{E}\mathbf{1} + \mathbf{1}'\mathbf{H}\mathbf{1}} \sim \Lambda_{1,\nu_H,\nu_E}$$

- Test for AB interaction
  - Analogous to "parallelism" in profile analysis

$$H_0 : \mathbf{C}\boldsymbol{\mu}_1 = \cdots = \mathbf{C}\boldsymbol{\mu}_g$$

Test statistic for AB:

$$\Lambda = \frac{|\mathbf{C}\mathbf{E}\mathbf{C}'|}{|\mathbf{C}(\mathbf{E}+\mathbf{H})\mathbf{C}'|} \sim \Lambda_{p-1,\nu_H,\nu_E}$$

$\boxed{\text{ex}}$ Wear of fabrics

- Measured in 3 periods (within-subjects factor)
  - 1^{st} 1000 revolutions
  - 2^{nd} 1000 revolutions
  - 3^{rd} 1000 revolutions
- 2 abrasive surfaces (between subjects factor #1)
- 2 fillers (between subjects factor #2)
- 3 levels of “proportion of filler” (between subjects factor #3)
  - 25% filler
  - 50% filler
  - 75% filler

? Linear or Quadratic trend in proportion of filler?

? Linear or Quadratic trend in periods

? How do univariate and multivariate tests compare?

Repeated Measures with 2 Within-Subjects Factors

Between-Subjects Factor	Subjects		Within-Subjects Factors									
			A_1			A_2			$\cdots$	A_4		
			B_1	B_2	B_3	B_1	B_2	B_3	$\cdots$	B_1	B_2	B_3
C_1	S_{11}	$\underset{\sim}{X}'_{11} =$	$(x_{111}$	x_{112}	x_{113}	x_{114}	x_{115}	x_{116}	$\cdots$	$x_{1,1,10}$	$x_{1,1,11}$	$x_{1,1,12})$
	$\vdots$											
	S_{1n_1}	$\underset{\sim}{X}'_{1n_1} =$	$(x_{1n_1 1}$	$x_{1n_1 2}$	$x_{1n_1 3}$	$x_{1n_1 4}$	$x_{1n_1 5}$	$x_{1n_1 6}$	$\cdots$	$x_{1,n_1,10}$	$x_{1,n_1,11}$	$x_{1,n_1,12})$
C_2	S_{21}						$\underset{\sim}{X}'_{21}$					
	$\vdots$						$\vdots$					
	S_{2n_2}						$\underset{\sim}{X}'_{2n_2}$					
C_3	S_{31}						$\underset{\sim}{X}'_{31}$					
	$\vdots$						$\vdots$					
	S_{3n_3}						$\underset{\sim}{X}'_{3n_3}$					

Model:

$$\mathbf{x}_{ij} = \boldsymbol{\mu} + \boldsymbol{\gamma}_i + \boldsymbol{\varepsilon}_{ij}$$

- $\boldsymbol{\gamma}_i$ is a p-vector of main effects for group i
- effects for A, B, AB, AC, BC, ABC assessed with contrasts
- Denote $a =$ number of levels for factor A
- Denote $b =$ number of levels for factor B
- To test factors A, B, and AB, specify contrast matrices with $(a-1), (b-1)$, and $(a-1)(b-1)$ linearly independent rows, respectively.

$\boxed{\text{ex}}$ Blood data

– Compare 4 different reagents used in blood testing. (Reagent 1 is standard and reagents 2, 3, 4 are inexpensive alternatives.)

$$\mathbf{A}^* = \begin{bmatrix} -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{bmatrix}$$

– Measuring 3 blood counts (white blood, red blood, hemoglobin)

$$\mathbf{B}^* = \begin{bmatrix} -1 & 0 & 1 \\ 1 & -2 & 1 \end{bmatrix}$$

– 2 groups of 10 subjects with potentially different blood properties — each subject's sample has 12 measures matrix for "reagent"

$$\begin{array}{cccccccccccc}
\multicolumn{3}{c}{\text{Reagent 1}} & \multicolumn{3}{c}{\text{Reagent 2}} & \multicolumn{3}{c}{\text{Reagent 3}} & \multicolumn{3}{c}{\text{Reagent 4}} \\
\text{BC1} & \text{BC2} & \text{BC3} & \text{BC1} & \text{BC2} & \text{BC3} & \text{BC1} & \text{BC2} & \text{BC3} & \text{BC1} & \text{BC2} & \text{BC3}
\end{array}$$

$$\underset{\sim}{A} = \left[\begin{array}{cccccccccccc}
-1 & -1 & -1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
-1 & -1 & -1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\
-1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1
\end{array}\right]
\begin{array}{l}
\leftarrow \underset{\sim}{a}_1' \\
\leftarrow \underset{\sim}{a}_2' \\
\leftarrow \underset{\sim}{a}_3'
\end{array}$$

matrix for "blood counts"

$$\underset{\sim}{B} = \left[\begin{array}{cccccccccccc}
-1 & 0 & 1 & -1 & 0 & 1 & -1 & 0 & 1 & -1 & 0 & 1 \\
1 & -2 & 1 & 1 & -2 & 1 & 1 & -2 & 1 & 1 & -2 & 1
\end{array}\right]
\begin{array}{l}
\leftarrow \underset{\sim}{b}_1' \\
\leftarrow \underset{\sim}{b}_2'
\end{array}$$

Note: $\mathbf{A} = \mathbf{A}^* \otimes \mathbf{1}_b'$ and $\mathbf{B} = \mathbf{1}_a' \otimes \mathbf{B}^*$, where $a = 4$ and $b = 3$

$\underset{1\times 12}{\mathbf{a}_1'} \Rightarrow$ R1 vs. R2

$\mathbf{a}_2' \Rightarrow$ R1 vs. R3

$\mathbf{a}_3' \Rightarrow$ R1 vs. R4

$\underset{1\times 12}{\mathbf{b}_1'} \Rightarrow$ white vs. hemoglobin (or linear in bc's)

$\mathbf{b}_2' \Rightarrow$ red vs. $\frac{\text{white+hemo}}{2}$ (or quadratic in bc's)

$$\underset{6\times 12}{\mathbf{G}} = \begin{bmatrix} \mathbf{a}_1 * \mathbf{b}_1 \\ \mathbf{a}_1 * \mathbf{b}_2 \\ \vdots \\ \mathbf{a}_3 * \mathbf{b}_2 \end{bmatrix} = \mathbf{A}^* \otimes \mathbf{B}^*$$

where "$*$" is an element-wise product

(So, first row of $\mathbf{G}$ is [1 0 -1 -1 0 1 0 0 0 0 0 0])

Test for A (Reagents):

- $T^2 = N(\underset{\substack{\uparrow \\ \text{grand} \\ \text{mean}}}{\mathbf{A}\,\bar{\mathbf{x}}})'(\mathbf{A}\,\underset{\substack{\uparrow \\ \frac{1}{\nu_E}\mathbf{E}}}{\mathbf{S}_{p\ell}}\,\mathbf{A}')^{-1}\mathbf{A}\bar{\mathbf{x}} \sim T^2_{a-1,\ \underset{\substack{\uparrow \\ \sum_{i=1}^{g}(n_i-1) \\ \text{when only} \\ \text{one between} \\ \text{subjects} \\ \text{factor is used}}}{\nu_E}}$

  $\boxed{\text{ex}}$ $T^2_{\text{Reagent}} \sim T^2_{3,18}$

  OR

- $\Lambda = \frac{|\mathbf{AEA}'|}{|\mathbf{A}(\mathbf{E}+\mathbf{H}^*)\mathbf{A}'|} \sim \Lambda_{a-1,1,\nu_E}$

  where $\mathbf{H}^* = N\bar{\mathbf{x}}\bar{\mathbf{x}}'$ is from the partitioning

$$\sum_{i=1}^{g}\sum_{j=1}^{n_1}\mathbf{x}_{ij}\mathbf{x}'_{ij} = \mathbf{E} + \mathbf{H} + N\bar{\mathbf{x}}\bar{\mathbf{x}}'$$

Test for B (Blood counts):

- $T^2 = N(\mathbf{B}\bar{\mathbf{x}})'(\mathbf{B}\mathbf{S}_{p\ell}\mathbf{B}')^{-1}\mathbf{B}\bar{\mathbf{x}} \sim T^2_{b-1,\nu_E}$

  OR

- $\Lambda = \frac{|\mathbf{BEB}'|}{|\mathbf{B}(\mathbf{E}+\mathbf{H}^*)\mathbf{B}'|} \sim \Lambda_{b-1,1,\nu_E}$

Test of AB interaction:

- $T^2 = N(\mathbf{G}\bar{\mathbf{x}})'(\mathbf{G}\mathbf{S}_{p\ell}\mathbf{G}')^{-1}\mathbf{G}\bar{\mathbf{x}} \sim T^2_{(a-1)(b-1),\nu_E}$

  OR

- $\Lambda = \frac{|\mathbf{GEG}'|}{|\mathbf{G}(\mathbf{E}+\mathbf{H}^*)\mathbf{G}'|} \sim \Lambda_{(a-1)(b-1),1,\nu_E}$

Test for C (groups):

- Conduct ANOVA test (F-test) using
$$z_{ij} = \mathbf{1}'\mathbf{x}_{ij},\ i = 1, \ldots, g, j = 1, \ldots, n_i$$
OR
- $\Lambda = \frac{\mathbf{1}'\mathbf{E}\mathbf{1}}{\mathbf{1}'\mathbf{E}\mathbf{1}+\mathbf{1}'\mathbf{H}\mathbf{1}} \sim \Lambda_{1,\nu_H,\nu_E}$

Tests for AC, BC, ABC interactions:

- $\Lambda = \frac{|\mathbf{AEA}'|}{|\mathbf{A}(\mathbf{E}+\mathbf{H})\mathbf{A}'|} \sim \Lambda_{a-1,\nu_H,\nu_E}$
- $\Lambda = \frac{|\mathbf{BEB}'|}{|\mathbf{B}(\mathbf{E}+\mathbf{H})\mathbf{B}'|} \sim \Lambda_{b-1,\nu_H,\nu_E}$
- $\Lambda = \frac{|\mathbf{GEG}'|}{|\mathbf{G}(\mathbf{E}+\mathbf{H})\mathbf{G}'|} \sim \Lambda_{(a-1)(b-1),\nu_H,\nu_E}$

Note: Between subjects effects (e.g., C) and associated interactions (e.g., AC, BC, ABC) use $\mathbf{H}$ (not $\mathbf{H}^*$)

$\boxed{\text{ex}}$ Blood data in SAS

III.B.vii. Tests on Covariance Matrices

(Reference: RC, Ch. 7)

- $H_0 : \mathbf{\Sigma} = \mathbf{\Sigma}_0$ vs. $H_1 : \mathbf{\Sigma} \neq \mathbf{\Sigma}_0$ (assuming MVN)

$$u = \nu \left[\ln |\mathbf{\Sigma}_0| - \ln |\mathbf{S}| + \mathrm{tr}\{\mathbf{S}\mathbf{\Sigma}_0^{-1}\} - p \right]$$

  is a modification of the likelihood ratio with $\nu =$ degrees of freedom for $\mathbf{S}$.

  – ν large:

$$u \mathrel{\dot{\sim}} \chi^2_{\frac{1}{2}p(p+1)}$$

  – ν small to moderate:

$$\left[1 - \frac{1}{6\nu - 1} \left(2p + 1 - \frac{2}{p+1} \right) \right] u \mathrel{\dot{\sim}} \chi^2_{\frac{1}{2}p(p+1)}$$

- $H_0 : \boldsymbol{\Sigma} = \sigma^2 \mathbf{I}$ ("sphericity" ... assuming MVN)

  Likelihood ratio test:

$$\lambda = \left[\frac{|\mathbf{S}|}{(\mathrm{tr}\{\mathbf{S}\}/p)} \right]^{\frac{n}{2}}$$

$$-2 \ln \lambda = -n \ln u$$

$$\text{where } u = \lambda^{\frac{2}{n}} = \frac{p^p |\mathbf{S}|}{(\mathrm{tr}\{\mathbf{S}\})^p} = \frac{p^p \prod_{i=1}^p \lambda_i}{(\sum_{i=1}^p \lambda_i)^p}$$

  and $\lambda_1, \ldots, \lambda_p$ are the e'vals of $\mathbf{S}$

  - ν large: $-n \ln u \stackrel{\cdot}{\sim} \chi^2_{\frac{1}{2}p(p+1)-1}$
  - ν small to moderate: $-\left(\nu - \frac{2p^2+p+2}{6p}\right) \ln u \stackrel{\cdot}{\sim} \chi^2_{\frac{1}{2}p(p+1)-1}$

– Note: Testing $\mathbf{C\Sigma C'} = \sigma^2\mathbf{I}$

use $\mathbf{CSC'}$ in place of $\mathbf{S}$ in the test, i.e.,

$$-n\ln\left(\frac{(p-1)^{p-1}|\mathbf{CSC'}|}{(\text{tr}\{\mathbf{CSC'}\})^{p-1}}\right) \sim \chi^2_{\frac{1}{2}(p-1)(p)-1}$$

where $\underset{(p-1)\times p}{\mathbf{C}}$ has orthonormal contrasts as its rows

$\boxed{\text{ex}}$ $p = 4$

$$\mathbf{C} = \begin{bmatrix} 3/\sqrt{12} & -1/\sqrt{12} & -1/\sqrt{12} & -1/\sqrt{12} \\ 0 & 2/\sqrt{6} & -1/\sqrt{6} & -1/\sqrt{6} \\ 0 & 0 & 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$$

– Often called "Mauchly's test"

  * Calculated by SAS with "PRINTE" option of "REPEATED" statement in PROC GLM.

– Fehlberg (1980) recommends a preliminary test of $\mathbf{\Sigma} = \sigma^2\mathbf{I}$ at $\alpha = .40$ before using standard univariate F-tests in r.m. analysis.

- $H_0 : \mathbf{\Sigma}_1 = \mathbf{\Sigma}_2 = \cdots = \mathbf{\Sigma}_g$ (assuming MVN for all groups)

$$\text{“Box's M”} = \frac{|\mathbf{S}_1|^{\frac{\nu_1}{2}} |\mathbf{S}_2|^{\frac{\nu_2}{2}} \cdots |\mathbf{S}_g|^{\frac{\nu_g}{2}}}{|\mathbf{S}_{p\ell}|^{\sum_i \frac{\nu_i}{2}}}$$

where $\nu_i = n_i - 1, i = 1, \ldots, g$ and $\mathbf{S}_{p\ell} = \frac{\sum_{i=1}^{g} \nu_i \mathbf{S}_i}{\sum_{i=1}^{g} \nu_i}$

  - M near 0 $\Rightarrow$ “reject H_0”
  - M near 1 $\Rightarrow$ “accept H_0”
  - Note: $M = \prod_{i=1}^{g} \left(\frac{|\mathbf{S}_i|}{|\mathbf{S}_{p\ell}|} \right)^{\frac{\nu_i}{2}}$

    ... is maximized at 1 when $\mathbf{S}_1 = \cdots = \mathbf{S}_g$

    ... approaches 0 when one or more $|\mathbf{S}_i|$ is very small (with other $|\mathbf{S}_i|$ large)
  - $u = -2(1 - c_1) \ln M \mathrel{\dot\sim} \chi^2_{\left[\frac{1}{2}(g-1)p(p+1)\right]}$

    where $c_1 = \left(\sum_{i=1}^{g} \frac{1}{\nu_i} - \frac{1}{\sum_{i=1}^{g} \nu_i} \right) \frac{2p^2+3p-1}{6(p+1)(g-1)}$

- Note: M-test *not* recommended pre-test before T^2 or MANOVA tests
  * Sensitive to nonnormality (often of little concern) and innocuous forms of heterogeneity (e.g., varying amounts of kurtosis)
- Note: A better approximation is $u \dot{\sim} F_{a_1,a_2}$. See RC for details.

III.C.i. Multivariate Multiple Regression

A short review of vec and Kronecker notation

Let $\underset{m\times n}{\mathbf{A}} = \begin{bmatrix} \mathbf{a}'_{1\cdot} \\ \mathbf{a}'_{2\cdot} \\ \vdots \\ \mathbf{a}'_{m\cdot} \end{bmatrix} = \begin{bmatrix} \underset{\substack{\uparrow \\ \text{an } m\text{-vector}}}{\mathbf{a}_{\cdot 1}} & \mathbf{a}_{\cdot 2} & \cdots & \mathbf{a}_{\cdot n} \end{bmatrix} = (a_{ij})$

- vec $\mathbf{A} = \begin{bmatrix} \mathbf{a}_{\cdot 1} \\ \mathbf{a}_{\cdot 2} \\ \vdots \\ \mathbf{a}_{\cdot n} \end{bmatrix}$ R: "$c(A)$" gives vec $\mathbf{A}$

Let $\underset{p\times q}{\mathbf{B}} = (b_{ij})$

- $$\underset{m\times n}{\mathbf{A}} \otimes \underset{p\times q}{\mathbf{B}} = \underbrace{\begin{bmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \cdots & a_{1n}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \cdots & a_{2n}\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}\mathbf{B} & a_{m2}\mathbf{B} & \cdots & a_{mn}\mathbf{B} \end{bmatrix}}_{mp\times nq}$$

R: "kronecker(A, B)" gives $\mathbf{A} \otimes \mathbf{B}$

Some properties (without proof)

Assuming that all dimensions are appropriate for matrix multiplication...

✓(a) $(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{AC}) \otimes (\mathbf{BD})$

✓(b) $\text{vec}\ (\mathbf{ABC}) = (\mathbf{C}' \otimes \mathbf{A})\text{vec}\ \mathbf{B}$

(c) $\text{tr}\{\mathbf{AB}\} = (\text{vec}\ \mathbf{A}')'\ \text{vec}\ \mathbf{B} = (\text{vec}\ \mathbf{A})'\ \text{vec}\ \mathbf{B}'$

(d) $\text{tr}\{\mathbf{ABCD}\} = (\text{vec}\ \mathbf{A}')'\ (\mathbf{D}' \otimes \mathbf{B})\ \text{vec}\ \mathbf{C} = (\text{vec}\ \mathbf{A})'\ (\mathbf{B} \otimes \mathbf{D}')\ \text{vec}\ \mathbf{C}'$

✓(e) $(\mathbf{A} \otimes \mathbf{B})' = \mathbf{A}' \otimes \mathbf{B}'$

✓(f) $(\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1}$

Univariate Multiple Regression:

$$\underset{n\times 1}{\mathbf{y}} = \underset{n\times r}{\mathbf{X}}\underset{r\times 1}{\boldsymbol{\beta}} + \underset{n\times 1}{\mathbf{e}}$$

- Assume $E\{\mathbf{e}\} = \mathbf{0}$ and $\text{var}\{\mathbf{e}\} = \sigma^2\mathbf{I}_n$. Then

$$\underset{\substack{\uparrow \\ \text{O.L.S.} \\ \text{estimator}}}{\hat{\boldsymbol{\beta}}} \quad = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

  is B.L.U.E. for $\boldsymbol{\beta}$.

- Note: we'll use q to denote the # of xs and $r = q + 1$ to denote the # of columns in the $\mathbf{X}$ matrix when using an intercept

Multivariate Multiple Regression:

$$\mathbf{Y} = \mathbf{X}\mathbf{B} + \boldsymbol{\Xi}$$

where

$$\mathbf{Y} = \begin{bmatrix} \mathbf{y}'_{1\cdot} \\ \vdots \\ \mathbf{y}'_{n\cdot} \end{bmatrix} = \begin{bmatrix} \mathbf{y}_{\cdot 1} & \cdots & \mathbf{y}_{\cdot p} \end{bmatrix}$$

$$\mathbf{B} = \begin{bmatrix} \boldsymbol{\beta}_{\cdot 1} & \boldsymbol{\beta}_{\cdot 2} & \cdots & \boldsymbol{\beta}_{\cdot p} \end{bmatrix}$$

$$\boldsymbol{\Xi} = \begin{bmatrix} \mathbf{e}'_{1\cdot} \\ \vdots \\ \mathbf{e}'_{n\cdot} \end{bmatrix} = \begin{bmatrix} \mathbf{e}_{\cdot 1} & \cdots & \mathbf{e}_{\cdot p} \end{bmatrix}$$

- Note that

$$\underset{n\times 1}{\mathbf{y}_{\cdot j}} = \underset{n\times r}{\mathbf{X}}\,\underset{r\times 1}{\boldsymbol{\beta}_{\cdot j}} + \underset{n\times 1}{\mathbf{e}_{\cdot j}}$$

- Assume $E\{\boldsymbol{\Xi}\} = \mathbf{0}$, $\mathrm{var}\{\mathbf{e}_{i\cdot}\} = \underset{p\times p}{\boldsymbol{\Sigma}}$, and $\mathrm{cov}\{\mathbf{e}_{i\cdot}, \mathbf{e}_{k\cdot}\} = \underset{p\times p}{\mathbf{0}}$ for all $i \neq k$

- Question: Is $\underset{r\times p}{\hat{\mathbf{B}}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ a B.L.U.E.?

Rewrite model:

$$
\begin{aligned}
\text{vec } \mathbf{Y} &= \text{vec } (\mathbf{XB}) + \text{vec } (\boldsymbol{\Xi}) \\
&= \underbrace{(\mathbf{I}_p \otimes \mathbf{X})}_{\substack{\uparrow \\ \text{rank } = pr \\ \text{when rank}(\mathbf{X})=r}} \quad \underbrace{\text{vec } \mathbf{B}}_{\equiv \underset{pr\times 1}{\boldsymbol{\beta}}} + \underbrace{\text{vec } (\boldsymbol{\Xi})}_{\equiv \underset{np\times 1}{\mathbf{e}}}
\end{aligned}
$$

Note: $E\{\mathbf{e}\} = \mathbf{0}$

and

$$
\text{var}\{\mathbf{e}\} = \text{var}\left\{\begin{pmatrix} \mathbf{e}_{\cdot 1} \\ \vdots \\ \mathbf{e}_{\cdot p} \end{pmatrix}\right\} = \begin{bmatrix} \sigma_{11}\mathbf{I}_n & \sigma_{12}\mathbf{I}_n & \cdots & \sigma_{1p}\mathbf{I}_n \\ \sigma_{21}\mathbf{I}_n & \sigma_{22}\mathbf{I}_n & \cdots & \sigma_{2p}\mathbf{I}_n \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1}\mathbf{I}_n & \sigma_{p2}\mathbf{I}_n & \cdots & \sigma_{pp}\mathbf{I}_n \end{bmatrix} = \underset{p\times p}{\boldsymbol{\Sigma}} \otimes \mathbf{I}_n
$$

- Since var$\{\underset{np\times 1}{\mathbf{e}}\}$ does *not* take the form $\sigma^2\mathbf{I}_{np}$, the B.L.U.E. for $\boldsymbol{\beta}$ will be the G.L.S. estimator for $\boldsymbol{\beta}$ (which depends on the unknown $\boldsymbol{\Sigma}$) BUT...

$$\hat{\boldsymbol{\beta}} = \left[\underbrace{\left(\mathbf{I}_p \otimes \underset{n \times p}{\mathbf{X}} \right)'}_{(\mathbf{I}_p \otimes \mathbf{X}')} \underbrace{(\boldsymbol{\Sigma} \otimes \mathbf{I}_n)^{-1}}_{\left(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_n\right)} \left(\mathbf{I}_p \otimes \underset{n \times p}{\mathbf{X}} \right) \right]^{-1} (\mathbf{I}_p \otimes \mathbf{X})' \, (\boldsymbol{\Sigma} \otimes \mathbf{I}_n)^{-1} \operatorname{vec} \mathbf{Y}$$

$$= \left[\left(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{X}' \right) (\mathbf{I}_p \otimes \mathbf{X}) \right]^{-1} (\mathbf{I}_p \otimes \mathbf{X}') \left(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_n \right) \ \operatorname{vec} \mathbf{Y}$$

[by prop's (a),(e),(f)]

$$= \left[\boldsymbol{\Sigma}^{-1} \otimes (\mathbf{X}'\mathbf{X}) \right]^{-1} \left(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{X}' \right) \ \operatorname{vec} \mathbf{Y} \qquad \text{[by prop (a)]}$$

$$= \left(\boldsymbol{\Sigma} \otimes (\mathbf{X}'\mathbf{X})^{-1} \right) \left(\boldsymbol{\Sigma}^{-1} \otimes \mathbf{X}' \right) \ \operatorname{vec} \mathbf{Y} \qquad \text{[by prop (f)]}$$

$$= \left(\mathbf{I}_p \otimes (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \right) \ \operatorname{vec} \mathbf{Y} \qquad \text{[by prop (a)]}$$

$$\Rightarrow \hat{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \qquad \text{[by prop (b)]}$$

- O.L.S. = G.L.S. is BLUE!

  (Even when $\boldsymbol{\Sigma}$ is unknown)

- Despite the fact that the p variables $y_{i1,\dots,y_{ip}}$ are correlated, all the info needed to estimate $\underset{r\times 1}{\boldsymbol{\beta}_{\cdot i}}$ is found in $\mathbf{y}_{\cdot i}$ only. That is, multivariate regression coefficient matrix $\underset{r\times p}{\hat{\mathbf{B}}}$ can be formed by pasting together the p columns from p separate univariate regressions (as long as each regression uses the same predictors $\underset{n\times r}{\mathbf{X}}$)

- But all $\hat{\boldsymbol{\beta}}_{ij}$ in $\mathbf{B}$ are intercorrelated ... must take multivariate approach to inference

Assumptions for Multivariate Multiple Regression:

Model is: $\underset{n\times p}{\mathbf{Y}} = \underset{n\times r}{\mathbf{X}}\underset{r\times p}{\mathbf{B}} + \underset{n\times p}{\boldsymbol{\Xi}}$ or vec $\mathbf{Y} = (\mathbf{I}\otimes\mathbf{X})\underbrace{\text{vec }\mathbf{B}}_{=``\boldsymbol{\beta}''} + \text{vec }\boldsymbol{\Xi}$

Assumptions:

1. $E\{\mathbf{Y}\} = \mathbf{XB}$ or $E\{\boldsymbol{\Xi}\} = \mathbf{0}$
2. $\text{var}\{\text{vec }\mathbf{Y}\} = \text{var}\{\text{vec }\boldsymbol{\Xi}\} = \boldsymbol{\Sigma}\otimes\mathbf{I}_n$

   (That is, $\text{var}\{\mathbf{y}_{i\cdot}\} = \boldsymbol{\Sigma}$ for all $i = 1,\ldots,n$ and $\text{cov}\{\mathbf{y}_{i\cdot},\mathbf{y}_{j\cdot}\} = \underset{p\times p}{\mathbf{0}}$ for all $i\neq j$)

Some properties of $\hat{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$

1. $\hat{\mathbf{B}}$ is called the "least squares estimator" because it "minimizes" $\underset{p\times p}{\mathbf{E}} = \hat{\boldsymbol{\Xi}}'\hat{\boldsymbol{\Xi}} = (\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}})'(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}})$ (where $\mathbf{E}$ is an "error matrix" analogous to the $\mathbf{E}$ matrix in MANOVA). Matrix is "minimized" in several senses:
   (a) Let $\tilde{\mathbf{B}}$ be some other estimate of $\mathbf{B}$.
   Then,
   $$(\mathbf{Y} - \mathbf{X}\tilde{\mathbf{B}})'(\mathbf{Y} - \mathbf{X}\tilde{\mathbf{B}}) = (\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}})'(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}}) + \mathbf{A}$$
   where $\mathbf{A}$ is a positive definite matrix
   (b) $\mathbf{B} = \hat{\mathbf{B}}$ minimizes $\mathrm{tr}\{(\mathbf{Y} - \mathbf{X}\mathbf{B})'(\mathbf{Y} - \mathbf{X}\mathbf{B})$
   (c) $\mathbf{B} = \hat{\mathbf{B}}$ minimizes $|(\mathbf{Y} - \mathbf{X}\mathbf{B})'(\mathbf{Y} - \mathbf{X}\mathbf{B})|$

2. Let $\hat{\mathbf{Y}} = \mathbf{X}\hat{\mathbf{B}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ be predicted values and $\hat{\mathbf{\Xi}} = \mathbf{Y} - \hat{\mathbf{Y}} = (\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{Y}$ be residuals

   Then

   (a) Residuals are perpendicular to the columns of $\mathbf{X}$

$$\rightarrow \mathbf{X}'\hat{\mathbf{\Xi}} = \mathbf{X}'(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{Y} = \underset{r\times p}{\mathbf{0}}$$

   (b) Residuals are perpendicular to the columns of $\hat{\mathbf{Y}}$

$$\rightarrow \hat{\mathbf{Y}}'\hat{\mathbf{\Xi}} = \hat{\mathbf{B}}'\mathbf{X}'(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{Y} = \underset{p\times p}{\mathbf{0}}$$

   (c) Total sum of squares and cross products ("Total SS and CP") can be partitioned as:

$$\mathbf{Y}'\mathbf{Y} = (\hat{\mathbf{Y}} + \hat{\mathbf{\Xi}})'(\hat{\mathbf{Y}} + \hat{\mathbf{\Xi}})$$

$$\underbrace{\mathbf{Y}'\mathbf{Y}}_{\substack{\uparrow \\ \text{total} \\ \text{SS\&CP} \\ \text{matrix}}} = \underbrace{\hat{\mathbf{Y}}'\hat{\mathbf{Y}}}_{\substack{\uparrow \\ \text{predicted} \\ \text{SS\&CP} \\ \text{matrix}}} + \underbrace{\hat{\mathbf{\Xi}}'\hat{\mathbf{\Xi}}}_{\substack{\uparrow \\ \text{error} \\ \text{SS\&CP} \\ \text{matrix}}}$$

3. $\hat{\mathbf{B}}$ is B.L.U.E. for $\mathbf{B}$

   - Minimum variance estimator among all unbiased estimators
   - If columns of $\mathbf{\Xi}$ are normal, $\hat{\mathbf{B}}$ is B.U.E.

4. Elements of $\hat{\mathbf{B}}$ are intercorrelated

$$\underset{r \times p}{\mathbf{B}} = \begin{bmatrix} \hat{\beta}_{01} & \hat{\beta}_{02} & \cdots & \hat{\beta}_{0p} \\ \hat{\beta}_{11} & \hat{\beta}_{12} & \cdots & \hat{\beta}_{1p} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\beta}_{q1} & \hat{\beta}_{q2} & \cdots & \hat{\beta}_{qp} \end{bmatrix}$$

   - $\hat{\beta}$s in each row are correlated due to correlation in $\mathbf{y}$
   - $\hat{\beta}$s in each column are correlated due to correlation in $\mathbf{x}$

5. Unbiased estimate of $\text{var}(\mathbf{y}_{i\cdot}) = \text{var}(\mathbf{e}_{i\cdot}) = \boldsymbol{\Sigma}$.

$$\mathbf{S} = \frac{\mathbf{E}}{n-q-1} = \frac{(\mathbf{Y}-\mathbf{X}\hat{\mathbf{B}})'(\mathbf{Y}-\mathbf{X}\hat{\mathbf{B}})}{n-q-1} = \frac{\hat{\boldsymbol{\Xi}}'\hat{\boldsymbol{\Xi}}}{n-q-1}$$
$$= \frac{1}{n-q-1}(\mathbf{Y}'\mathbf{Y} - \hat{\mathbf{B}}'\mathbf{X}'\mathbf{Y})$$

Proof:

$$\begin{aligned}
E\{\underset{n\times p}{\hat{\boldsymbol{\Xi}}}\} &= E\{\mathbf{Y}-\mathbf{X}\hat{\mathbf{B}}\} \\
&= E\{(\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{Y}\} \\
&= (\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\ E\{\mathbf{X}\mathbf{B}+\boldsymbol{\Xi}\} \\
&= (\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\ E\{\boldsymbol{\Xi}\} \\
&= \underset{n\times p}{\mathbf{0}}
\end{aligned}$$

$$
\begin{aligned}
E\{\hat{\mathbf{e}}'_{\cdot i}\hat{\mathbf{e}}_{\cdot j}\} &= E\left\{\left[(\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{y}_{\cdot i}\right]'\left[(\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{y}_{\cdot j}\right]\right\} \\
&= E\left\{\left[(\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{e}_{\cdot i}\right]'\left[(\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{e}_{\cdot j}\right]\right\} \\
&= E\{\mathbf{e}'_{\cdot i}\underbrace{\left(\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right)\left(\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right)}_{\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'}\mathbf{e}_{\cdot j}\} \\
&= E\left\{\mathrm{tr}\left\{\left(\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right)\mathbf{e}_{\cdot j}\mathbf{e}'_{\cdot i}\right\}\right\} \\
&= \mathrm{tr}\{\left(\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right)\underbrace{E\left\{\mathbf{e}_{\cdot j}\mathbf{e}'_{\cdot i}\right\}}_{\sigma_{ij}\mathbf{I}_n}\} \\
&= \sigma_{ij}\mathrm{tr}\left\{\left(\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right)\right\} \\
&= \sigma_{ij}\left(\mathrm{tr}\left\{\mathbf{I}_n\right\} - \mathrm{tr}\left\{\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\right\}\right) \\
&= \sigma_{ij}\left(n - (q+1)\right)
\end{aligned}
$$

$$
\therefore E\left\{\tfrac{1}{n-q-1}\hat{\mathbf{\Xi}}'\hat{\mathbf{\Xi}}\right\} = \mathbf{\Sigma} = (\sigma_{ij})
$$

Note: If $\mathbf{X}$ is not full rank, we can obtain similar results based on
$\hat{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{Y}$ ... we'll leave that discussion for "linear models"!

Another note: $\widehat{\text{var}}\left\{\underset{np\times 1}{\mathbf{e}}\right\} = \left(\frac{1}{n-q-1}\mathbf{E}\right) \otimes \mathbf{I}_n$

and $E\left\{\left(\frac{1}{n-q-1}\mathbf{E}\right) \otimes \mathbf{I}_n\right\} = \mathbf{\Sigma} \otimes \mathbf{I}_n$

6. Variance of $\hat{\boldsymbol{\beta}}$ (i.e., var$\{\text{vec } \hat{\mathbf{B}}\}$)

$$
\begin{aligned}
\text{var}\left\{\underset{rp\times 1}{\hat{\boldsymbol{\beta}}}\right\} &= \text{var}\left\{\left[(\mathbf{I}\otimes\mathbf{X})'(\mathbf{I}\otimes\mathbf{X})\right]^{-1}(\mathbf{I}\otimes\mathbf{X})' \text{vec } \mathbf{Y}\right\} \\
&= \left[\mathbf{I}_p \otimes (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right] \text{var}\{\text{vec } \mathbf{Y}\} \left[\mathbf{I}_p \otimes (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right]' \\
&= \left[\mathbf{I}_p \otimes (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right] \text{var}\{\text{vec } \mathbf{\Xi}\} \left[\mathbf{I}_p \otimes (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right]' \\
&= \left[\mathbf{I}_p \otimes (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right] (\mathbf{\Sigma}\otimes\mathbf{I}_n) \left[\mathbf{I}_p \otimes \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\right] \\
&= \Sigma \otimes \left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{I}_n\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\right] \\
&= \mathbf{\Sigma} \otimes (\mathbf{X}'\mathbf{X})^{-1}
\end{aligned}
$$

Notes:

(a) $$\widehat{\text{var}}\{\hat{\boldsymbol{\beta}}\} = \overbrace{\left(\frac{1}{n-q-1}\mathbf{E}\right)}^{\text{“}\mathbf{S}\text{”}} \otimes (\mathbf{X}'\mathbf{X})^{-1}$$

(b) $$\text{cov}\{\hat{\boldsymbol{\beta}}_{\cdot i}, \hat{\boldsymbol{\beta}}_{\cdot j}\} = \sigma_{ij}(\mathbf{X}'\mathbf{X})^{-1}$$

(c) $$\text{cov}\{\underset{r\times 1}{\hat{\boldsymbol{\beta}}_{\cdot i}}, \underset{n\times 1}{\hat{\mathbf{e}}_{\cdot j}}\} =$$

$$\begin{aligned}
&\text{cov}\left\{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}_{\cdot i}, \left(\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right)\mathbf{y}_{\cdot j}\right\} \\
&= \text{cov}\left\{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{e}_{\cdot i}, \left(\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right)\mathbf{e}_{\cdot j}\right\} \\
&= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\sigma_{ij}\mathbf{I}_n\left(\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right) \\
&= \sigma_{ij}\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\right] \\
&= \underset{r\times n}{\mathbf{0}}
\end{aligned}$$

(d) Estimating mean of $\underset{1\times r}{\mathbf{x}_0'}\underset{r\times p}{\mathbf{B}}$

- $\mathbf{x}_0'\hat{\mathbf{B}}$ is an unbiased estimator of $\mathbf{x}_0'\mathbf{B}$
- $\text{var}\{\mathbf{x}_0'\hat{\mathbf{B}}\} = \mathbf{\Sigma}(\underbrace{\mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0}_{\text{scalar}})$

(e) Estimating a new observation $\mathbf{y}_0$ using $\mathbf{x}_0$

$\mathbf{y}_0' = \mathbf{x}_0'\mathbf{B} + \mathbf{e}_0'$

- $\mathbf{x}_0'\hat{\mathbf{B}}$ is an unbiased estimator of $\mathbf{y}_0$
- $\text{var}\{\mathbf{y}_0' - \mathbf{x}_0'\hat{\mathbf{B}}\}$ $\leftarrow$ "forecast error variance"
  - Note that

$$\begin{aligned}
\text{cov}\{\mathbf{y}_0', \mathbf{x}_0'\hat{\mathbf{B}}\} &= \text{cov}\left\{\mathbf{e}_0', \mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\left(\mathbf{X}\mathbf{B} + \mathbf{\Xi}\right)\right\} \\
&= \text{cov}\left\{\mathbf{e}_0', \mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Xi}\right\} \\
&= \underset{p\times p}{\mathbf{0}} \quad \text{since } \mathbf{e}_0' \text{ is indep. of } \mathbf{\Xi} = \begin{bmatrix} \mathbf{e}_{1\cdot}' \\ \vdots \\ \mathbf{e}_{n\cdot}' \end{bmatrix}
\end{aligned}$$

– So

$$
\begin{aligned}
\text{var}\{\mathbf{y}_0' - \mathbf{x}_0'\hat{\mathbf{B}}\} &= \text{var}\{\mathbf{y}_0'\} + \text{var}\{\mathbf{x}_0'\hat{\mathbf{B}}\} - 2\ \text{cov}\{\mathbf{y}_0', \mathbf{x}_0'\hat{\mathbf{B}}\} \\
&= \mathbf{\Sigma} + \mathbf{\Sigma} \cdot \left(\mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0\right) + \mathbf{0} \\
&= \mathbf{\Sigma} \cdot \left[1 + \mathbf{x}_0'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_0\right]
\end{aligned}
$$

7. MLE's of $\underset{r\times p}{\mathbf{B}}$ and $\boldsymbol{\Sigma}$

Thus far, we have assumed $E\{\underset{\substack{\uparrow\\ \text{vec } \boldsymbol{\Xi}}}{\mathbf{e}}\} = \mathbf{0}$ and $\text{var}\{\mathbf{e}\} = \boldsymbol{\Sigma} \otimes \mathbf{I}_n$

If we assume:

$$\underset{np\times 1}{\mathbf{e}} \sim N_{np}\left(\mathbf{0}, \boldsymbol{\Sigma} \otimes \mathbf{I}_n\right)$$

then MLE's of $\mathbf{B}$ and $\boldsymbol{\Sigma}$ are

$$\hat{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

and

$$\underset{p\times p}{\hat{\boldsymbol{\Sigma}}} = \frac{1}{n}\hat{\boldsymbol{\Xi}}'\hat{\boldsymbol{\Xi}} = \frac{1}{n}\mathbf{E}$$

where

$$\mathbf{E} \sim W_p(n-q-1, \boldsymbol{\Sigma})$$

Proof: omitted.

8. Model Corrected for Means Rewrite

$$\mathbf{Y} = \underset{n\times r}{\mathbf{X}}\,\mathbf{B} + \boldsymbol{\Xi}$$

as

$$\underset{n\times p}{\mathbf{Y}_c} = \underset{n\times q}{\mathbf{X}_c}\,\underset{q\times p}{\mathbf{B}_c} + \boldsymbol{\Xi} \text{ where } q = \#\text{ of predictors} = r-1$$

and

$$\mathbf{Y}_c = \begin{bmatrix} y_{11}-\bar{y}_{\cdot 1} & y_{12}-\bar{y}_{\cdot 2} & \cdots & y_{1p}-\bar{y}_{\cdot p} \\ \vdots & & & \vdots \\ y_{n1}-\bar{y}_{\cdot 1} & y_{n2}-\bar{y}_{\cdot 2} & \cdots & y_{np}-\bar{y}_{\cdot p} \end{bmatrix}$$

$$\mathbf{X}_c = \begin{bmatrix} x_{11}-\bar{x}_{\cdot 1} & \cdots & x_{1q}-\bar{x}_{\cdot q} \\ \vdots & & \vdots \\ x_{n1}-\bar{x}_{\cdot 1} & \cdots & x_{nq}-\bar{x}_{\cdot q} \end{bmatrix}$$

Then $\hat{\mathbf{B}}_c = \mathbf{S}_{xx}^{-1}\mathbf{S}_{xy}$

where $\mathbf{S} = \begin{bmatrix} \mathbf{S}_{yy} & \mathbf{S}_{yx} \\ \mathbf{S}_{xy} & \mathbf{S}_{xx} \end{bmatrix}$ is the sample covariance matrix of the $p+q$ variables $(y_1, \ldots, y_p,\ x_1, \ldots, x_q)$

$$\hat{\mathbf{Y}} = \begin{bmatrix} \bar{y}_{\cdot 1} & \mathbf{1}_n & \cdots & \bar{y}_{\cdot p} & \mathbf{1}_n \end{bmatrix} + \underset{n\times q}{\mathbf{X}_c}\,\underset{q\times p}{\hat{\mathbf{B}}_c}$$

Hypothesis Tests (assuming $\mathbf{e} \sim N_{np}\{\mathbf{0}, \boldsymbol{\Sigma} \otimes \mathbf{I}_r\}$)

$$\boxed{H_0 : \underset{q\times p}{\mathbf{B}_1} = \mathbf{0} \quad \text{(Test of overall regression)}}$$

$$\text{where } \underset{r\times p}{\mathbf{B}} = \begin{bmatrix} \boldsymbol{\beta}_0' \\ \mathbf{B}_1 \end{bmatrix} \begin{matrix} \leftarrow 1\times p \;\; \text{vector of intercepts} \\ \leftarrow q\times p \end{matrix}$$

Partition the total SS and CP matrix:

$$\underset{p\ n}{\mathbf{Y}'}\underset{n\ p}{\mathbf{Y}} = \underbrace{\left(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}}\right)'\left(\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}}\right)}_{=\mathbf{Y}'\mathbf{Y} - \hat{\mathbf{B}}'\mathbf{X}'\mathbf{Y} \;=\; \mathbf{E}} + \hat{\mathbf{B}}'\mathbf{X}'\mathbf{Y}$$

To avoid inclusion of $\boldsymbol{\beta}_0' = \mathbf{0}'$ as part of the null hypothesis, we subtract $n\bar{\mathbf{y}}\bar{\mathbf{y}}'$:

$$\underbrace{\mathbf{Y}'\mathbf{Y} - n\bar{\mathbf{y}}\bar{\mathbf{y}}'}_{\text{corrected total SS \& CP}} = \underbrace{\mathbf{Y}'\mathbf{Y} - \hat{\mathbf{B}}'\mathbf{X}'\mathbf{Y}}_{=\underset{p\ p}{\mathbf{E}}} + \underbrace{\hat{\mathbf{B}}'\mathbf{X}'\mathbf{Y} - n\bar{\mathbf{y}}\bar{\mathbf{y}}'}_{=\underset{p\times p}{\mathbf{H}}}$$

$$\Lambda = \frac{|\mathbf{E}|}{|\mathbf{E}+\mathbf{H}|} = \frac{|\mathbf{Y}'\mathbf{Y} - \hat{\mathbf{B}}'\mathbf{X}'\mathbf{Y}|}{|\mathbf{Y}'\mathbf{Y} - n\bar{\mathbf{y}}\bar{\mathbf{y}}'|} \sim \Lambda_{p,\underbrace{q}_{r-1},\underbrace{n-q-1}_{n-r}}$$

- $\mathbf{H}$ is "large" when $\hat{\boldsymbol{\beta}}_{ij}$'s are large
- The 4 MANOVA statistics can be calculated as functions of the eigenvalues of $\mathbf{E}^{-1}\mathbf{H}, (\lambda_1, \ldots, \lambda_s)$:
  - Wilks': $\Lambda = \prod_{i=1}^{s} \frac{1}{1+\lambda_i}$
  - Roy's: $\theta = \lambda_1$
  - Pillai's Trace: $V = \sum_{i=1}^{s} \frac{\lambda_i}{1+\lambda_i}$
  - Lawley-Hotelling Trace: $U = \sum_{i=1}^{s} \lambda_i$

$\rightarrow$ Critical values (and p-values) based on approximate F-distributions given on the MANOVA pages on these notes

. . . use:

$s = \min(p, q)$
$m = \frac{1}{2}(|q-p| - 1)$
$N = \frac{1}{2}(n - q - p - 2)$

- Essential dimensionality of $\underbrace{\mathbf{E}^{-1}\mathbf{H}}_{p\times p}$ is the essential dimensionality of $\underset{q\times p}{\mathbf{B}_1}$. For example, a single non-zero eigenvalue (i.e., rank of B_1 is 1) could be due to several causes:

  1. $\mathbf{B}_1$ has only one nonzero row
     $\Rightarrow$ only one of the x's predicts the y's
  2. $\mathbf{B}_1$ has only one nonzero column
     $\Rightarrow$ only one of the y's is predicted by the x's
  3. All of the rows of $\mathbf{B}_1$ are linear combinations of each other
     $\Rightarrow$ x's act alike in predicting y's
     [or, in other words]
     All of the columns of $\mathbf{B}_1$ are linear combinations of each other
     $\Rightarrow$ only one dimension in the y's as they relate to x's

- "Essential dimensionality" of $\mathbf{E}^{-1}\mathbf{H}$ is number of substantially non-zero eigenvalues and takes value less than or equal to $s = \min(p, q)$

Essential dimensionality

$1 \longrightarrow s$

θ is m.p. test $\qquad$ $V^{(s)}$ is m.p. test

- Λ can also be calculated from the partitioned sample covariance matrix of $(y_1, \ldots, y_p, x_1, \ldots, x_q)$

$$\underset{(p+q)\times(p+q)}{\mathbf{S}} = \begin{bmatrix} \underset{p\times p}{\mathbf{S}_{yy}} & \underset{p\times q}{\mathbf{S}_{yx}} \\ \underset{q\times p}{\mathbf{S}_{xy}} & \underset{q\times q}{\mathbf{S}_{xx}} \end{bmatrix}$$

using

$$\frac{|\mathbf{S}|}{|\mathbf{S}_{xx}||\mathbf{S}_{yy}|}$$

which is essentially a test of independence between $\mathbf{y}$ and $\mathbf{x}$ since

$$\text{Independence of } \mathbf{y} \text{ and } \mathbf{x} \;\Rightarrow |\mathbf{S}| = |\mathbf{S}_{yy}||\mathbf{S}_{xx}|$$

$\boxed{H_0 : \mathbf{B}_a = \mathbf{0} \qquad \text{Tests on a subset of the } x\text{'s}}$

Hypothesis states that the y's do not depend on the last h of the x's. That is,

$$H_0 : \mathbf{B}_{\text{add}} = \mathbf{0}$$

where

$$\underset{r \times p}{\mathbf{B}} = \begin{bmatrix} \mathbf{B}_{\text{red}} \\ \mathbf{B}_{\text{add}} \end{bmatrix} \begin{array}{l} \leftarrow (r-h) \times p \\ \leftarrow h \times p \end{array}$$

Compare SS and CP matrix for full and reduced models:

$$\underset{p \times p}{\mathbf{H}_{\text{diff}}} = \hat{\mathbf{B}}'\mathbf{X}'\mathbf{Y} - \hat{\mathbf{B}}_r'\mathbf{X}_r'\mathbf{Y} \quad \leftarrow \text{difference in regression SS and CP}$$

and

$$\underset{p \times p}{\mathbf{E}_{\text{full}}} = \mathbf{Y}'\mathbf{Y} - \hat{\mathbf{B}}'\mathbf{X}'\mathbf{Y} \quad \leftarrow \mathbf{E} \text{ matrix based on full model}$$

Then

$$\Lambda_{x_{q-h+1},\cdots,x_q|x_1,\cdots,x_{q-h}} = \frac{|\mathbf{E}_{\text{full}}|}{|\mathbf{E}_{\text{full}} + \mathbf{H}_{\text{diff}}|} = \frac{|\mathbf{E}_{\text{full}}|}{|\mathbf{E}_{\text{red}}|}$$

$$= \frac{|\mathbf{Y}'\mathbf{Y} - \hat{\mathbf{B}}'\mathbf{X}'\mathbf{Y}|}{|\mathbf{Y}'\mathbf{Y} - \hat{\mathbf{B}}'_{\text{red}}\mathbf{X}'_{\text{red}}\mathbf{Y}|}$$

$$\sim \Lambda_{p,\underset{\substack{\uparrow \\ \#\text{ of } x\text{s}}}{h},n-q-1}$$

- Note:

$$\Lambda_{x_{q-h+1},\cdots,x_q|x_1,\cdots,x_{q-h}} = \frac{\left(\frac{|\mathbf{Y}'\mathbf{Y}-\hat{\mathbf{B}}'\mathbf{X}'\mathbf{Y}|}{|\mathbf{Y}'\mathbf{Y}-n\bar{\mathbf{y}}\bar{\mathbf{y}}'|}\right)}{\left(\frac{|\mathbf{Y}'\mathbf{Y}-\hat{\mathbf{B}}'_{\text{red}}\mathbf{X}'_{\text{red}}\mathbf{Y}|}{|\mathbf{Y}'\mathbf{Y}-n\bar{\mathbf{y}}\bar{\mathbf{y}}'|}\right)} = \frac{\Lambda_{\text{full}}}{\Lambda_{\text{red}}}$$

$\rightarrow$ makes full vs. reduced testing simple to carry out

- Note: θ, V, and U can be calculated from eigenvalues of $\mathbf{E}_{\text{full}}^{-1}\mathbf{H}_{\text{diff}}$ with

$$s = \min(p, h)$$
$$m = \frac{1}{2}(|h - p| - 1)$$
$$N = \frac{1}{2}(n - h - p - 2)$$

Subset Selection

Finding a subset of the x's to include in a model

- Forward Selection

  (Step 1)

  Start with

$$\hat{\mathbf{B}}_i = \begin{bmatrix} \hat{\boldsymbol{\beta}}_{01} & \hat{\boldsymbol{\beta}}_{02} & \cdots & \hat{\boldsymbol{\beta}}_{0p} \\ \hat{\boldsymbol{\beta}}_{i1} & \hat{\boldsymbol{\beta}}_{i2} & \cdots & \hat{\boldsymbol{\beta}}_{ip} \end{bmatrix}$$

  and calculate

$$\Lambda_{x_i} = \frac{|\mathbf{Y}'\mathbf{Y} - \hat{\mathbf{B}}_i'\mathbf{X}'\mathbf{Y}|}{|\mathbf{Y}'\mathbf{Y} - n\bar{\mathbf{y}}\bar{\mathbf{y}}'|} \sim \Lambda_{p,1,n-2}$$

  for $i = 1, \ldots, q$. Add the x_i that minimizes Λ_{x_i} (as long as $\Lambda_{x_i} < \Lambda_{\alpha,p,1,n-2}$ — stop otherwise)

(Step $j+1, j=1,2,\ldots$)

Let $x_1,\ldots,x_j$ be the variables added in previous steps. Calculate

$$\Lambda_{x_i|x_1,\ldots,x_j} \sim \Lambda_{p,1,n-j-1}$$

for all x_i among the $q-j$ remaining candidate variables. For the x_i that minimizes $\Lambda_{x_i|x_1,\ldots,x_j}$:

– add x_i if $\Lambda_{x_i|x_1,\ldots,x_j} < \Lambda_{\alpha,p,1,n-j-1}$

– stop the procedure if $\Lambda_{x_i|x_1,\ldots,x_j} > \Lambda_{\alpha,p,1,n-j-1}$

- Backward Elimination

Start with all x's and delete one at a time until the least valuable remaining x is significant. For the m remaining x's after a given step, find the x_i maximizing

$$\Lambda_{x_i|x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_m} \sim \Lambda_{p,1,n-m-1}$$

– drop x_i if $\Lambda_{x_i|x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_m} > \Lambda_{\alpha,p,1,n-m-1}$

– stop the procedure if x_i if $\Lambda_{x_i|x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_m} < \Lambda_{\alpha,p,1,n-m-1}$

- Stepwise
  - Add most significant candidate x_i if partial Λ is less than critical value
  - Then, remove least significant selected x_i if partial Λ is greater than critical value
- Best Subsets

  Choose "best" subset of size ℓ, for $\ell = 1, \ldots, q$, with respect to some criterion (e.g., a multivariate extension of Mallow's C_p, or $\text{tr}\{\mathbf{S}\}$, etc.)

After selecting a subset of the x's, subset of y's may be selected using "stepwise discriminant" approach ... to be discussed later.

$\boxed{\text{ex}}$ chemical reaction data

How are the responses (y_1, y_2, and y_3) affected by the inputs (x_1, x_2, and x_3)?

y_1 = % of unchanged starting material

y_2 = % converted to the desired product

y_3 = % of unwanted by-product

x_1 = temperature

x_2 = concentration

x_3 = time

- Regress $\mathbf{y}$ on $\mathbf{x}$ to obtain $\hat{\mathbf{B}}$ and test $\mathbf{B}_1 = 0$, where $\mathbf{B} = \begin{bmatrix} \boldsymbol{\beta}_0' \\ \mathbf{B}_1 \end{bmatrix}$.
- Determine what the eigenvalues of $\mathbf{E}^{-1}\mathbf{H}$ reveal about the essential rank of $\hat{\mathbf{B}}_1$ and the power of the 4 ("MANOVA") statistics.

  NOTE: "MTEST/PRINT DETAILS" gives eigenvalues of $(\mathbf{E}+\mathbf{H})^{-1}\mathbf{H}$ $(\xi_1, \ldots, \xi_s)$ and $\xi_i = \frac{\lambda_i}{1+\lambda_i}$, and $\lambda_i = \frac{\xi_i}{1-\xi_i}$

- Check the significance of x_1x_2, x_1x_3, x_2x_3, x_1^2, x_2^2, and x_3^2 adjusted for x_1, x_2, and x_3

III.C.ii Seemingly Unrelated Regressions (SUR)

- Standard multivariate regression:

$$\mathbf{Y} = \mathbf{X}\mathbf{B} + \boldsymbol{\Xi}$$

or

$$\underbrace{\begin{bmatrix} \mathbf{y}_{\cdot 1} \\ \vdots \\ \mathbf{y}_{\cdot p} \end{bmatrix}}_{\text{vec } \mathbf{Y}} = \underbrace{\begin{bmatrix} \mathbf{X} & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{X} \end{bmatrix}}_{\underset{p\times p}{\mathbf{I}} \otimes \underset{n\times p}{\mathbf{X}}} \underbrace{\begin{bmatrix} \boldsymbol{\beta}_1 \\ \vdots \\ \boldsymbol{\beta}_p \end{bmatrix}}_{\text{vec } \mathbf{B}} + \underbrace{\begin{bmatrix} \mathbf{e}_{\cdot 1} \\ \vdots \\ \mathbf{e}_{\cdot p} \end{bmatrix}}_{\text{vec } \boldsymbol{\Xi}}$$

  - Note: each $\mathbf{y}_{\cdot 1}$ uses the same regressors $\underset{n\times r}{\mathbf{X}}$
  - vec $\hat{\mathbf{B}} = (\mathbf{I} \otimes (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\text{vec } \mathbf{Y}$

    or

    $\hat{\mathbf{B}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ is B.L.U.E. even though $\text{var}\{\text{vec } \mathbf{Y}\} = \boldsymbol{\Sigma} \otimes \mathbf{I}$
  - What if each $\mathbf{y}_{\cdot j}$ uses different regressors $\underset{n\times r_j}{\mathbf{X}_j}$?

- SUR Model

$$\underset{n\times 1}{\mathbf{y}_{\cdot j}} = \underset{n\times r_j}{\mathbf{X}_j}\underset{r_j\times 1}{\boldsymbol{\beta}_j} + \underset{n\times 1}{\mathbf{e}_j} \qquad j = 1, \ldots, p$$

$\boxed{\text{ex}}$ $\mathbf{y}_{\cdot j}$ is the jth economic outcome for n regions and $\mathbf{X}_j$ is the matrix of economic indicators (unemployment, housing starts, etc.) and the indicators used in the model are potentially different for each outcome—that is, $\mathbf{X}_j \neq \mathbf{X}_{j'}$

- Model assumptions:

$$\underbrace{\begin{bmatrix} \mathbf{y}_{\cdot 1} \\ \mathbf{y}_{\cdot 2} \\ \vdots \\ \mathbf{y}_{\cdot p} \end{bmatrix}}_{np\times 1} = \underbrace{\begin{bmatrix} \underset{n\times r_1}{\mathbf{X}_1} & & & \mathbf{0} \\ & \underset{n\times r_2}{\mathbf{X}_2} & & \\ & & \ddots & \\ \mathbf{0} & & & \underset{n\times r_p}{\mathbf{X}_p} \end{bmatrix}}_{np\times(\sum_{j=1}^p r_j)} \underbrace{\begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \vdots \\ \boldsymbol{\beta}_p \end{bmatrix}}_{(\sum_{j=1}^p r_j)\times 1} + \underbrace{\begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \\ \vdots \\ \mathbf{e}_p \end{bmatrix}}_{np\times 1}$$

OR

$$\mathbf{y}^* = \mathbf{X}^*\boldsymbol{\beta}^* + \mathbf{e}^*$$

$$\text{cov}\{\mathbf{e}_i, \mathbf{e}_j\} = \sigma_{ij}\mathbf{I}_n \quad \leftarrow \text{ independent observations}$$

$$\Rightarrow \text{var}\{\mathbf{e}^*\} = \underset{p\times p}{\boldsymbol{\Sigma}} \otimes \mathbf{I}_n = \underset{np\times np}{\boldsymbol{\Sigma}^*}$$

- OLS estimator

$$\hat{\boldsymbol{\beta}}^*_{\mathrm{OLS}} = \left(\mathbf{X}^{*\prime}\mathbf{X}^*\right)^{-1}\mathbf{X}^{*\prime}\mathbf{y}^* = \begin{bmatrix} \left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1}\mathbf{X}_1'\mathbf{y}_{\cdot 1} \\ \vdots \\ \left(\mathbf{X}_p'\mathbf{X}_p\right)^{-1}\mathbf{X}_p'\mathbf{y}_{\cdot p} \end{bmatrix} = \begin{bmatrix} \hat{\boldsymbol{\beta}}_{1,\mathrm{OLS}} \\ \vdots \\ \hat{\boldsymbol{\beta}}_{p,\mathrm{OLS}} \end{bmatrix}$$

is <u>not</u> BLUE in general

– Zyskind condition states that the OLS estimator is BLUE if and only if there exists a matrix $\mathbf{Q}$ such that $\boldsymbol{\Sigma}^*\mathbf{X}^* = \mathbf{X}^*\mathbf{Q}$

  * In standard multivariate regression, $\boldsymbol{\Sigma}^* = \underset{p\times p}{\boldsymbol{\Sigma}} \otimes \mathbf{I}_n$ and $\mathbf{X}^* = \mathbf{I}_p \otimes \underset{n\times r}{\mathbf{X}}$ and

$$\begin{aligned}
\boldsymbol{\Sigma}^*\mathbf{X}^* &= \left(\underset{p\times p}{\boldsymbol{\Sigma}} \otimes \mathbf{I}_n\right)\left(\mathbf{I}_p \otimes \underset{n\times r}{\mathbf{X}}\right) \\
&= \boldsymbol{\Sigma} \otimes \mathbf{X} \\
&= \underbrace{\left(\mathbf{I}_p \otimes \underset{n\times r}{\mathbf{X}}\right)}_{\mathbf{X}^*}\underbrace{\left(\underset{p\times p}{\boldsymbol{\Sigma}} \otimes \mathbf{I}_n\right)}_{\mathbf{Q}}
\end{aligned}$$

    Therefore, OLS is BLUE under standard multivariate regression assumptions

  * In SUR case, there is no simple way of writing $\mathbf{X}^*$, and in general, there exists no $\mathbf{Q}$ satisfying $\boldsymbol{\Sigma}^*\mathbf{X}^* = \mathbf{X}^*\mathbf{Q}$

- Use $\hat{\boldsymbol{\beta}}^*_{\text{SUR}} = \left(\mathbf{X}^{*\prime}\boldsymbol{\Sigma}^{*-1}\mathbf{X}^*\right)^{-1}\mathbf{X}^{*\prime}\boldsymbol{\Sigma}^{*-1}\mathbf{y}^*$

III.C.iii Canonical Correlation Analysis

Objective: summarize the linear relationship between two groups of variables $\mathbf{y} = (y_1, \ldots, y_p)$ and $\mathbf{x} = (x_1, \ldots, x_q)$

- Neither $\mathbf{x}$ nor $\mathbf{y}$ considered "dependent"
- Multivariate extension of the squared multiple correlation coefficient (used to relate a single response y with $\mathbf{x}$).

- Consider a single random variable y and a random vector $\mathbf{x}$

  Recall

$$\operatorname{var}\left\{\begin{bmatrix}\mathbf{y}\\ \mathbf{x}\end{bmatrix}\right\} = \mathbf{S} = \begin{bmatrix}\mathbf{S}_{yy} & \mathbf{S}_{yx}\\ \mathbf{S}_{xy} & \mathbf{S}_{xx}\end{bmatrix} \text{ if } \overset{p=1}{=} \begin{bmatrix}\underset{1\times 1}{s_{yy}} & \mathbf{s}'_{yx}\\ \mathbf{s}_{xy} & \underset{q\times q}{\mathbf{S}_{xx}}\end{bmatrix}$$

  and

$$\operatorname{corr}\left\{\begin{bmatrix}\mathbf{y}\\ \mathbf{x}\end{bmatrix}\right\} = \mathbf{R} = \begin{bmatrix}\mathbf{R}_{yy} & \mathbf{R}_{yx}\\ \mathbf{R}_{xy} & \underset{q\times q}{\mathbf{R}_{xx}}\end{bmatrix} \text{ if } \overset{p=1}{=} \begin{bmatrix}1 & \mathbf{r}'_{yx}\\ \mathbf{r}_{xy} & \mathbf{R}_{xx}\end{bmatrix}$$

Squared multiple correlation between y and $(x_1, \ldots, x_q)$ is

$$R^2_{y|\mathbf{x}} = \frac{\mathbf{s}_{yx}\mathbf{S}^{-1}_{xx}\mathbf{s}_{xy}}{s_{yy}} = \mathbf{r}'_{yx}\mathbf{R}^{-1}_{xx}\mathbf{r}_{xy}$$

where $R^2_{y|\mathbf{x}}$ is themaximum correlation between y and a linear combination of the x's

- Extending to the case with $\mathbf{y} = (y_1, \ldots, y_p)$ and $\mathbf{x} = (x_1, \ldots, x_q)$, a measure of association between $\mathbf{y}$ and $\mathbf{x}$ is

$$R^2_M = |\mathbf{S}^{-1}_{yy}\mathbf{S}_{yx}\mathbf{S}^{-1}_{xx}\mathbf{S}_{xy}| = \prod_{i=1}^{s} r_i^2 \quad (s = \min(p, q))$$

where $r_1^2, \ldots, r_s^2$ are the eigenvalues of $\mathbf{S}^{-1}_{yy}\mathbf{S}_{yx}\mathbf{S}^{-1}_{xx}\mathbf{S}_{xy}$.

  - R^2_M too small and too heavily dependent on smallest eigenvalues.
  - Instead, work directly with $r_1^2, \ldots, r_s^2$ called the "squared canonical correlations"

- Largest squared canonical correlation r_1^2 is the maximum squared correlation between a linear combination of $\mathbf{x}$ and a linear combination of $\mathbf{y}$.

$$\sqrt{r_1^2} = \text{corr}\{\underbrace{\mathbf{a}_1'\mathbf{y}}_{u_1}, \underbrace{\mathbf{b}_1'\mathbf{x}}_{v_1}\} = \max_{\mathbf{a},\mathbf{b}} \text{corr}\{\mathbf{a}'\mathbf{y}, \mathbf{b}'\mathbf{x}\}$$

  - $u_1 = \mathbf{a}_1'\mathbf{y}$ and $v_1 = \mathbf{b}_1'\mathbf{x}$ are the "first canonical variates"
  - First s eigenvalues of $\underbrace{\mathbf{S}_{xx}^{-1}\mathbf{S}_{xy}\mathbf{S}_{yy}^{-1}\mathbf{S}_{yx}}_{q\times q}$ are same as first s eigenvalues of $\underbrace{\mathbf{S}_{yy}^{-1}\mathbf{S}_{yx}\mathbf{S}_{xx}^{-1}\mathbf{S}_{xy}}_{p\times p}$, but eigenvectors are different.
  - $(\mathbf{S}_{yy}^{-1}\mathbf{S}_{yx}\mathbf{S}_{xx}^{-1}\mathbf{S}_{xy} - r^2\mathbf{I}_p)\mathbf{a} = \mathbf{0}$
    * If $q < p$, only q of the eigenvectors $\mathbf{a}$ are meaningful

    $(\mathbf{S}_{xx}^{-1}\mathbf{S}_{xy}\mathbf{S}_{yy}^{-1}\mathbf{S}_{yx} - r^2\mathbf{I}_q)\mathbf{b} = \mathbf{0}$
    * If $p < q$, only p of the eigenvectors $\mathbf{b}$ are meaningful

- The canonical correlations $r_1, \ldots, r_s$ respond to the s pairs of canonical variates:

$$\left.\begin{array}{ccc} u_1 = \mathbf{a}_1'\mathbf{y} & \text{and} & v_1 = \mathbf{b}_1'\mathbf{x} \\ u_2 = \mathbf{a}_2'\mathbf{y} & \text{and} & v_2 = \mathbf{b}_2'\mathbf{x} \\ \vdots & \vdots & \vdots \\ u_s = \mathbf{a}_s'\mathbf{y} & \text{and} & v_s = \mathbf{b}_s'\mathbf{x} \end{array}\right\} \begin{array}{c}\text{the } s \text{ nonredundant} \\ \text{dimensions of the} \\ \text{relationship} \\ (s = \min(p, q))\end{array}$$

  - u_i's are uncorrelated (so are v_i's)
  - u_i uncorrelated with v_j for $i \neq j$

- If software requires a symmetric matrix to obtain eigenvalues and eigenvectors, use

$$\mathbf{S}_{xx}^{-1/2}\mathbf{S}_{xy}\mathbf{S}_{yy}^{-1}\mathbf{S}_{yx}\mathbf{S}_{xx}^{-1/2}$$

  which has eigenvalues $r_1^2, \ldots, r_s^2$ and eigenvectors $\mathbf{S}_{xx}^{1/2}\mathbf{b}_i$ and

$$\mathbf{S}_{yy}^{-1/2}\mathbf{S}_{yx}\mathbf{S}_{xx}^{-1}\mathbf{S}_{xy}\mathbf{S}_{yy}^{-1/2}$$

  which has eigenvalues $r_1^2, \ldots, r_s^2$ and eigenvectors $\mathbf{S}_{yy}^{1/2}\mathbf{a}_i$

- Importance of the relationship between u_i and v_i (that is, importance of r_i^2) can be judged by the relative size of λ_i (the eigenvalues of $\mathbf{E}^{-1}\mathbf{H}$):

$$\frac{\lambda_i}{\sum_{j=1}^{s} \lambda_j}$$

- $\underset{p\times 1}{\mathbf{a}_i} = \frac{1}{r_i} \underset{p\times p}{\mathbf{S}_{yy}^{-1}} \underset{p\times q}{\mathbf{S}_{yx}} \underset{q\times 1}{\mathbf{b}_i}$

  $\underset{q\times 1}{\mathbf{b}_i} = \frac{1}{r_i} \underset{q\times q}{\mathbf{S}_{xx}^{-1}} \underset{q\times p}{\mathbf{S}_{xy}} \mathbf{a}_i$

- When interpreting the canonical variates, we prefer to use standardized coefficient vectors

$$\mathbf{c}_i = \begin{bmatrix} s_{y_1} & 0 & \cdots & 0 \\ 0 & s_{y_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & s_{y_p} \end{bmatrix} \mathbf{a}_i,$$

and

$$\mathbf{d}_i = \begin{bmatrix} s_{x_1} & 0 & \cdots & 0 \\ 0 & s_{x_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & s_{x_q} \end{bmatrix} \mathbf{b}_i$$

where $s_{y_i} = \sqrt{\text{var}\{y_i\}}$ and $s_{x_i} = \sqrt{\text{var}\{x_i\}}$.

  - More simply, conduct analysis using $\mathbf{R}_{yy}^{-1}\mathbf{R}_{yx}\mathbf{R}_{xx}^{-1}\mathbf{R}_{xy}$ and $\mathbf{R}_{xx}^{-1}\mathbf{R}_{xy}\mathbf{R}_{yy}^{-1}\mathbf{R}_{yx}$ which have eigenvectors $\mathbf{c}_i$ and $\mathbf{d}_i$, respectively, and have eigenvalues $r_1^2, \ldots, r_s^2$.

- Properties of canonical correlations:
  - r_i^2 invariant to change of scale on y's or x's
  - r_1 exceeds the absolute value of the correlation between any y and any or all of the x's.
  - $r_i^2 = R^2_{u_i|\mathbf{x}} = R^2_{v_i|\mathbf{y}}$ (where $R^2_{u_i|\mathbf{x}}$ is the squared multiple correlation between u_i and $(x_1, \ldots, x_q)$)

Statistical Inference:

- H_0: no linear relationship between y's and x's

  or

  $H_0 : \underset{q\times p}{\mathbf{B}_1} = \mathbf{0}$

  or

  H_0: independence of $\mathbf{y}$ and $\mathbf{x}$

- Test statistic:

$$\frac{|\mathbf{S}|}{|\mathbf{S}_{yy}||\mathbf{S}_{xx}|} = \frac{|\mathbf{R}|}{|\mathbf{R}_{yy}||\mathbf{R}_{xx}|} \sim \Lambda_{p,q,n-1-q}$$

$$\overset{q}{\equiv} \Lambda_{q,p,n-1-p}$$

$$\text{since } \Lambda_{p,q,n-1-q} \overset{q}{=} \Lambda_{q,p,n-1-p}$$

[An exact F test exists for $s = \min(p, q) \leq 2$, see $\mathbf{\Lambda}$-to-F conversions in the MANOVA section]

- $\mathbf{\Lambda} = \prod_{i=1}^{s}(1 - r_i^2)$ is a function of $r_1^2, \ldots, r_s^2$, as are Pillai's (V), Lawley-Hotelling (U), and Roy's Largest root (θ)
  - As strength of relationship between $\mathbf{x}$ and $\mathbf{y}$ increases, r_i^2's increase and Λ decreases
  - Testing if the s canonical correlations (combined) are significant
  - Note: If $p = 1, \Lambda = 1 - R^2_{y|\mathbf{x}}$
  - If test rejects, next consider how many r_i^2's are significant
- H_0: The canonical correlations $r_m, \ldots, r_s$ are non-significant

$$\Lambda_m = \prod_{i=m}^{s}(1 - r_i^2) \sim \Lambda_{p-m+1,q-m+1,n-m-q}$$

$$\text{or } \Lambda_{q-m+1,p-m+1,n-m-p}$$

$$[\text{since } \Lambda_{p,\nu_H,\nu_E} \overset{q}{=} \Lambda_{\nu_H,p,\nu_H+\nu_E-p}]$$

  - An approach: Check $\Lambda_2, \ldots, \Lambda_s$ to determine number of significant r_i^2 values.

Interpretation of canonical variates (e.g., $u_1 = \mathbf{a}_1'\mathbf{y}$ and $v_1 = \mathbf{b}_1'\mathbf{x}$)

Wish to assess the contribution of each variable to the canonical correlation r_i^2.

- standardized coefficients
- correlations between y_1 and $u_j = \mathbf{a}_j'\mathbf{y}$

- Standardized coefficients

  Use $\mathbf{c}_i$ and $\mathbf{d}_i$ to account for differences in scaling among the variables

  - Absolute values of coefficients $\mathbf{c}_i$ show contribution of each y_i in the presence of the other y_i's.
  - Add or remove y_i's $\Rightarrow$ $\mathbf{c}_i$ changes

    We want this property in multivariate analysis!!

- Correlations between y_i and $u_j = \mathbf{a}_j'\mathbf{y}$ (and between x_i and $v_j = \mathbf{b}_j'\mathbf{x}$)
  - More frequently used and widely claimed to yield more valid interpretation of canonical variates (a.k.a. "structure coefficients")
  - corr$\{y_i, u_j\}$ is "stable" (not dramatically different) if we add or remove y_i's ... sounds nice, but it's not!

    In fact, these correlations provide <u>no</u> information about the <u>multivariate</u> contribution of the variable y_i to the correlation structure. (Analogous to T_p^2-test vs. p univariate t-tests.)

    Rencher (1988, 1992) showed that

$$\prod_{j=1}^{s} (\text{corr}\{y_i, u_j\})^2 r_j^2 = R^2_{y_i|\mathbf{x}}$$

    where $R^2_{y_i|\mathbf{x}}$ is the multiple correlation between y_i and $(x_1, \ldots, x_q)$

- Although $\text{corr}\{y_i, u_j\}$ might seem to quantify the importance of y_i in a multivariate relationship with $\mathbf{x}$ in the presence of the other y variables, it summarizes only a univariate relationship.

$\boxed{\text{ex}}$ chemical reaction data

? How many r_i^2's are necessary?

? Interpret u_1 and v_1.

- Recall $\sqrt{R^2_{x_i|\mathbf{y}}} = \sqrt{\sum_{j=1}^{3}(\text{corr}\{x_i, v_1\})^2 r_j^2}$.
- First canonical correlation is mostly due to relationship of: temperature and concentration (suppressed by x_1x_2, and to a lesser degree x_1x_3 and x_1^2) with: % changed.

INPUTS		d_{1i}	$\text{corr}\{x_i, v_1\}$	$\sqrt{R^2_{x_i\|\mathbf{y}}}$
x_1 = temperature	**	5.01	.69	.69
x_2 = concentration	**	5.86	.23	.24
x_3 = time		1.65	.45	.51
x_1x_2	**	-3.92	.41	.43
x_1x_3	*	-2.30	.54	.58
x_2x_3		0.53	.45	.48
x_1^2	*	-2.67	.69	.69
x_2^2		-1.23	.23	.23
x_3^2		0.57	.42	.47

YIELDS		c_{1i}	$\text{corr}\{y_i, u_1\}$	$\sqrt{R^2_{y_i\|\mathbf{x}}}$
y_1 = % unchanged	**	-1.54	-.996	.987
y_2 = % converted		-0.21	.64	.92
y_3 = % by-product		-0.47	.85	.91