$$F_{p,n-p} \stackrel{D}{=} \frac{\chi^2_p/p}{\underset{\text{ind}}{\perp} \chi^2_{n-p}/(n-p)} \xrightarrow{D} \frac{\chi^2_p}{p}$$

$\chi^2_{n-p} \stackrel{D}{=} \sum_{i=1}^{n-p} X_i$, $E X_i = 1$, X_i iid χ^2_1

Also $T_C^2 = n(\bar{\underline{x}} - \underline{\mu})^T S^{-1} (\bar{\underline{x}} - \underline{\mu}) \xrightarrow{D} \chi^2_p$

and $T_C^2 = n(T - \underline{\mu})^T \hat{D}^{-1} (T - \underline{\mu}) \xrightarrow{D} \chi^2_p$ if $\sqrt{n}(T - \underline{\mu}) \xrightarrow{D} N_p(\underline{0}, D)$.

So $\frac{n-p}{n-1} \frac{1}{p} T_C^2 \xrightarrow{D} \frac{\chi^2_p}{p}$.

So Hotelling's T_C^2 test with $(\bar{\underline{x}}, S)$ is a large sample ~~test~~ if $\underline{x}_1, \ldots, \underline{x}_n$ are iid from a distribution with nonsingular covariance matrix $\Sigma_{\underline{x}}$.

So the test is robust to nonnormality.

ex) (p. 173) X_1 = sweat rate, X_2 = sodium content, X_3 = potassium content, $\underline{\mu}_0 = (4, 50, 10)^T$

$T_C^2 = 9.74$, $n = 20$, $p = 3$

i) H_0: $\underline{\mu} = \underline{\mu}_0$ $\quad H_A$: $\underline{\mu} \neq \underline{\mu}_0$

ii) $T_C^2 = 9.74$, $\frac{n-p}{n-1} \frac{1}{p} T_C^2 = \frac{20-3}{20-1} \frac{1}{3} 9.74 = 2.905$

iii) $2.905 < 3.20$ so pval > 0.05

Den	num 3 = p
n − p = 17	3.20

Den		num 3
15	.9	2.49
	.95	3.29
20	.9	2.38
	.95	3.10

$.05 <$ pval $< .1$

iv) fail to reject H_0, not enough evidence to conclude that $\underline{\mu} \neq \begin{pmatrix} 4 \\ 50 \\ 10 \end{pmatrix}$.

33.5

3) The one sample Hotelling's T^2 test is the analog of the univariate t test for μ. Now there are $p>1$ variables and their correlations are important. Many multivariate procedures are the multivariate analog of univariate techniques.

4) The two F tables give left tail area. want right tail area.

For 15.5 F table with $\alpha = 0.05$, reject Ho if

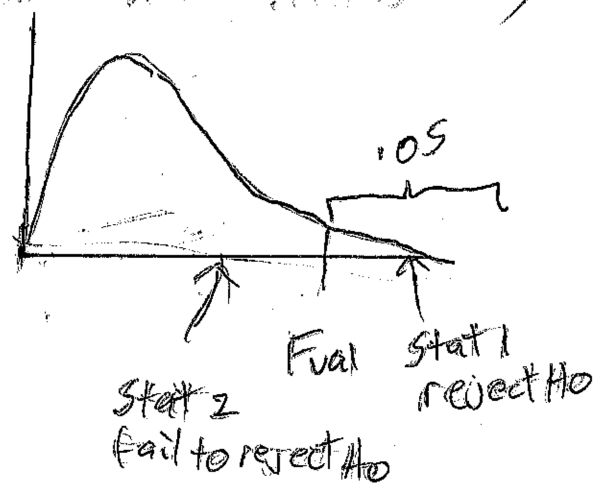
$$\frac{n-p}{(n-1)p} T_c^2 > F\text{value}, \text{ since pval} < 0.05.$$

If $\frac{n-p}{(n-1)p} T_c^2 < F\text{value}$, fail to reject Ho

since pval > 0.05.

(2.905 < 3.2 in tables)

num df = p = 2
den df = n-p = 10



	$\frac{n-p}{(n-1)p}T_c^2$	$p=2$ $n=12$ pval
	.5	.5 < pval
.5	.743	.5
	2	.1 < pval < .5
.9	2.92	.1
	3	.05 < pval < .1
.95	4.10	.05
	5	.025 < pval < .05
.975	5.46	.025
	6	.01 < pval < .025
.99	7.56	.01
	8	.005 < pval < .01
.995	9.43	.005
	12	.001 < pval < .005
.999	14.9	.001
	20	< .001

(§6.2)

5) large sample multivariate matched pairs test

$k=2$ treatments applied to the "same" n cases

p measurements taken for both treatments $(\underline{y}_i, \underline{z}_i)$.

eg $p=2$ systolic and diastolic blood pressure

want these to be highly correlated pairs

before medication (treatment 1) and after medication (treatment 2)

	sys	dys	$\swarrow$ p in general	sys	dys	$\swarrow$ p in general
person 1	y_{11}	y_{12}	$=\underline{y}_1^T$	z_{11}	z_{12}	$=\underline{z}_1^T$
⋮	⋮	⋮		⋮	⋮	
n	y_{n1}	y_{n2}	$=\underline{y}_n^T$	z_{n1}	z_{n2}	$=\underline{z}_n^T$

Let $\underline{d}_i = \underline{x}_i = \underline{y}_i - \underline{z}_i$. Assume $\underline{x}_i$ are iid with mean $\underline{\mu}$ and covariance matrix $\Sigma_{\underline{x}}$

Compute $(\overline{\underline{x}}, S_{\underline{x}})$ for the $\underline{x}_i$ $i=1,\ldots,n$.

i) H_0 $\underline{\mu} = \underline{0}$ $\quad H_1$ $\underline{\mu} \neq \underline{0}$

one sample test applied to the $\underline{x}_i$ with $\underline{\mu}_0 = \underline{0}$

ii) $T_m^2 = n\, \overline{\underline{x}}^T S_{\underline{x}}^{-1} \overline{\underline{x}}$

iii) $pval = P\left(\frac{(n-p)}{(n-1)p} T_m^2 < F_{p,n-p}\right) = P\left(F_{p,n-p} > \frac{n-p}{(n-1)p} T_m^2\right)$

iv) reject H_0 if $pval < \alpha$ and conclude $\underline{\mu} \neq \underline{0}$

fail to reject H_0 if $pval \geq \alpha$ and conclude either $\underline{\mu} = \underline{0}$ or that there is not enough evidence to conclude that $\underline{\mu} \neq \underline{0}$.

Give a nontechnical sentence if possible.

ex) (p213-4) waste water from a sewage treatment plant was sent to two labs. measurements of biochemical oxygen demand (bod) and suspended solids (ss) were obtained. (345)

	lab1		lab 2	
	bod	ss	bod	ss
1	6	27	25	15
⋮				
11	20	14	39	21

$n=11, \ p=2 \qquad T_M^2 = 13.6$. Do test;

i) $H_0 \ \underline{\mu} = \underline{0} \qquad H_1 \ \underline{\mu} \neq \underline{0}$

ii) $T_M^2 = 13.6$

iii) $\frac{n-p}{(n-1)p} T_M^2 = \frac{11-2}{11-1} \frac{1}{2} \ 13.6 = 6.12$

$pval = P(F_{2,9} > 6.12)$ ($2 = p$, $9 = n-p$)

table 15.5

	2
9	4.26

other table

		2
9	.975	5.71
	.99	8.02

$6.12 > 4.26$

$pval < 0.05$

$.01 < pval < .025$

iv) reject H_0, $\underline{\mu} \neq \underline{0}$. The two labs are giving different mean measurements for bod and or ss.

6) Repeated measurements $\approx$ longitudinal data analysis; take p measurements on the same unit (case, individual), often the same type of measurement at several time periods so blood pressure at end of week 1, 2, ..., p.

Note: cross sectional study: p variables are measured at the same time for an individual

longitudinal study (or repeated measurements): individual is measured repeatedly through time

Morrison P 135 – 141

measurement at ith time = ith variable

Unit	X_1	X_2	...	X_p	
1	x_{11}	x_{12}		x_{1p}	$= \underset{\sim}{x}_1^T$
2	x_{21}	x_{22}		x_{2p}	$= \underset{\sim}{x}_2^T$
⋮					
n	x_{n1}	x_{n2}		x_{np}	$= \underset{\sim}{x}_p^T$

$$\underset{\sim}{\mu} = (\mu+\tau_1, \ldots, \mu+\tau_p)^T = (\mu_1, \ldots, \mu_p)^T$$

$$H_0 \ \underbrace{\mu_1 = \cdots = \mu_p}_{\text{equal response effects}} \quad H_A \text{ not } H_0$$

$$H_0 \begin{pmatrix} \mu_1 - \mu_2 \\ \vdots \\ \mu_{p-1} - \mu_p \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}, \quad H_A \begin{pmatrix} \mu_1 - \mu_2 \\ \vdots \\ \mu_{p-1} - \mu_p \end{pmatrix} \neq \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$$

Let $y_{ij} = x_{ij} - x_{i,j+1}$ $\quad i = 1, \ldots, n; \quad j = 1, \ldots, p-1.$

$$\underline{\bar{y}} = (\bar{y}_1, \ldots, \bar{y}_{p-1})^T = (\bar{x}_1 - \bar{x}_2, \ldots, \bar{x}_{p-1} - \bar{x}_p)^T.$$

Let $S_{\underline{y}}$ be the sample covariance matrix of the $\underset{\sim}{y}_i$.

i) $H_0 \ \underline{\mu}_{\underline{y}} = \underline{0} \qquad H_A \ \underline{\mu}_{\underline{y}} \neq \underline{0}$ $\qquad$ one sample test on the $\underline{y}_i$

ii) $T_R^2 = n \ \underline{\bar{y}}^T S_{\underline{y}}^{-1} \underline{\bar{y}}$ $\qquad$ $(p-1) \times 1$

iii) $\text{pval} = P\left(F_{p-1,\, n-p+1} > \frac{n-p+1}{(n-1)(p-1)} T_R^2\right)$

iv) If pval $< \alpha$, reject H_0 and conclude not all p of the μ_i are equal. If pval $\geq \alpha$ fail to reject H_0 and conclude all p of the μ_i are equal or that there is not enough evidence to conclude that not all p of the μ_i are equal. If possible, give a nontechnical conclusion.

ex) Reaction times to visual stimuli were obtained for $n = 20$ normal young men under three conditions A, B, C of stimulus display. Let

$$\mu_A = \mu + \tau_1, \quad \mu_B = \mu + \tau_2, \quad \mu_C = \mu + \tau_3.$$

$$\bar{x}_A = 21.05 \quad \bar{x}_B = 21.65 \quad \bar{x}_C = 28.95$$

Test whether $\mu_A = \mu_B = \mu_C$ if $T_R^2 = 882.8$.

i) $H_0\ \underset{\sim}{\mu}_y = \underset{\sim}{0}$ $\quad H_1\ \underset{\sim}{\mu}_y \neq \underset{\sim}{0}$

ii) $T_R^2 = 882.8$

iii) $\frac{n-p+1}{(n-1)(p-1)} T_R^2 = \frac{20-3+1}{(20-1)(3-1)} 882.8 = 418.168$

$\text{pval} = P(418.168 < F_{2,18})$ $\quad$ $\begin{array}{c|c} & 2 \\ \hline 18 & 3.55 \end{array}$ $\quad = P(F_{2,18} > 418.168)$

$F_{\underset{p-1}{2},\, \underset{n-p+1}{18}}$

pval ≈ 0

iv) reject H_0, the three mean reaction times are different

7) For the repeated measurements test,

$\underline{y} = A\underline{x}$ where

$$\underset{(p-1)\times p}{A} = \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & -1 & 0 \cdots & 0 & 0 \\ \vdots & & & & \vdots & \vdots \\ 0 & \cdots & & & 0 \; 1 & -1 \end{bmatrix}.$$

$$T_R^2 = n\,\bar{\underline{x}}^T A^T (A S_{\underline{x}} A^T)^{-1} A \bar{\underline{x}} = T_R^2(A).$$

If G is any nonsingular $(p-1)\times(p-1)$ matrix,

then $\underline{y} = GA\underline{x}_i$ can be used to

test H_0, and $T_R^2(A) = T_R^2(GA)$.

other choices are

$$G_1 A = A_1 = \begin{bmatrix} 1 & -1 & \cdots & & 0 \\ 1 & 0 & -1 & \cdots & 0 \\ \vdots & & & & \vdots \\ 1 & 0 & \cdots & 0 & -1 \end{bmatrix}$$

or $$G_2 A = A_2 = \begin{bmatrix} -1 & \cdots & & 0 & 1 \\ 0 & -1 & \cdots & 0 & 1 \\ 0 & \cdots & & -1 & 1 \end{bmatrix},$$

36.5

8) p160 Suppose test rejects H_0: $\underline{\mu} = \underline{\mu}_0$.

want to know which values of $\mu_i \neq \mu_{0i}$.

Large sample simultaneous $100(1-\alpha)\%$ confidence intervals for $\underline{a}_1^T\underline{\mu}, \ldots, \underline{a}_J^T\underline{\mu}$

satisfy $P($each of the J CIs contains $\underline{a}_i^T\underline{\mu})$ $\rightarrow 1-\alpha$ as $n \rightarrow \infty$.

ex) The $J = p + \frac{p(p-1)}{2}$ CIs for μ_i and $\mu_i - \mu_k$ can be made so they all hold simultaneously with confidence $\rightarrow 1-\alpha$. Hence if $\alpha = 0.05$, in 100 samples, expect all $p + \frac{p(p-1)}{2}$ CIs to contain μ_i and $\mu_i - \mu_k$ about 95 times while about 5 times at least one CI will fail to contain its parameter.

9) Let $\underline{x}_1, \ldots, \underline{x}_n$ be iid with mean $\underline{\mu}$ and covariance matrix $\Sigma_{\underline{x}}$. The large sample simultaneous $100(1-\alpha)\%$ CIs for $\underline{a}^T\underline{\mu}$ have the form

$$(L_a, U_a) = \underline{a}^T\overline{\underline{x}} \pm \sqrt{\frac{p(n-1)}{n(n-p)} F_{p,n-p,1-\alpha}\, \underline{a}^T S \underline{a}}$$

go to top of 38 last ex

CI for μ_i: $\overline{x}_i \pm \sqrt{\frac{p(n-1)}{n-p} F_{p,n-p,1-\alpha}} \sqrt{\frac{S_{ii}}{n}}$

CI for $\mu_i - \mu_k$: $\overline{x}_i - \overline{x}_k \pm \sqrt{\frac{p(n-1)}{n-p} F_{p,n-p,1-\alpha}} \sqrt{\frac{S_{ii} - 2S_{ik} + S_{kk}}{n}}$