Contents

Statistical Analysis of Multivariate Data

- **Data-analytic techniques:-**
  - Techniques based just on the structure of data
  - Use statistical ideas of variances, means, ….
  - Provide data descriptions & simplification e.g.
    - PCA
    - MD Scaling
    - LDA
- **PCA is the vital first step in the multivariate analysis of continuous data**

- No assumptions of statistical distributions
  - c.f. non-parametric methods
    - use statistical ideas but not distributional models
- Now introduce a statistical distribution as a model for multivariate data & review basic results of related distributions and statistical tests

- **Quotation:–**
  - Much classical and formal theoretical work in Multivariate Analysis rests on assumptions of underlying *multivariate normality* — resulting in techniques of very limited value (Gnanadesikan, page 2)
  - Extremely difficult to test distributional assumptions in high dimensions since need a LOT of data
    - so results can be very dependent on distributional assumptions
    - same true of Bayesian approaches

- First consider generalizations of 1-d results
  - Normal N(μ, σ^2)
    χ^2, t & F-distributions
  - MLEs of μ & $\sigma \Rightarrow$ sample mean and [~] s.d.
    - divisor n not n –1
      - (but variance with divisor (n –1) gives unbiased of σ^2)
  - one sample test of $\mu = 0$
  - two sample tests of $\mu_1 = \mu_2$
  - analysis of variance to test $\mu_1 = \mu_2 = \ldots = \mu_k$
  - &c.,&c.,…..

- Next:-
  - Outline introduction to LRT tests
    - Likelihood Ratio Tests

    of more complex hypotheses

    Then
- Union-Intersection Tests
  - UITs
  - New general purpose method for constructing hypothesis tests
    - Greater interpretability, c.f. PCA/LDA loadings

©NRJF, University of Sheffield, 2011/12 Semester 1

- **Univariate (p=1)**
- $N(\mu, \sigma^2)$
  - μ scalar
  - σ^2 scalar
  - sample mean unbiased for μ
  - sample variance unbiased for σ^2
    - sample variance is a scalar

- **Multivariate p-dims**
- $N_p(\mu, \Sigma)$
  - μ p-vector
  - Σ p×p matrix
  - sample mean unbiased for μ
  - sample variance unbiased for Σ
    - sample variance is a p×p matrix

- **Univariate (p=1)**
- $N(\mu, \sigma^2)$
  - sample mean $\sim N(\mu, \sigma^2/n)$
  - Linear combinations ~ normal with appropriate mean & variance

- **Multivariate p-dims**
- $N_p(\mu, \Sigma)$
  - sample mean $\sim N_p(\mu, \Sigma/n)$
  - Linear combinations ~ normal with appropriate mean & variance **and dimension**

- **Univariate (p=1)**
- $N(\mu, \sigma^2)$
  - MLE of μ is sample mean
  - MLE of σ^2 is sample variance × $(n-1)/n$ i.e. divisor n not n –1

- **Multivariate p-dims**
- $N_p(\mu, \Sigma)$
  - MLE of μ is sample mean
  - MLE of Σ is sample variance × $(n-1)/n$ i.e. divisor n not n –1

- **N.B.** to shew this in multivariate case
  - Need to differentiate w.r.t. vector μ
    - standard result since only terms of form $\mu'\mathbf{A}\mu$ involved
      - derivative is $2\mathbf{A}\mu$
  - And to differentiate w.r.t. matrix Σ
    - Difficult and **non-examinable**
      - but see lecture notes, neater to use $T=\Sigma^{-1}$

- Key formulæ:
  - See p202

$$\hat{\mu} = \bar{x}, \quad \hat{\Sigma} = \tfrac{n-1}{n} S$$

- And if $d = \bar{x} - \hat{\mu}$ (e.g. under H_0: $\mu = \mu_0$ then $d = \bar{x} - \mu_0$)

$$\hat{\Sigma} = \tfrac{n-1}{n} S + dd'$$

- used for calculating L_{max} under H_0 and H_A

- Likelihood ratio tests constructed by finding the values of *maximized likelihoods*
  - i.e. plug in MLEs of parameters into likelihood
  - General theory says 'best' test statistic is $L_{max}(H_A)/L_{max}(H_0)$ and reject H_0 in favour of H_A if this is large
  - i.e. *if* the [maximized] likelihood when H_A is true is much larger than that when H_0 is true *then* evidence favours H_A
    - very sensible!

- More general theory (***Wilks' Theorem)*** says instead of using $L_{max}(H_A)/L_{max}(H_0)$ easier to use

$$\lambda = 2\{\log_e [L_{max}(H_A)] - \log_e[L_{max}(H_0)]\}$$

- Since if H_0 is true then $\lambda \sim \chi^2_r$
  - where r is calculated from difference in number of parameters estimated under H_A and H_0
    - strictly need some technical regularity conditions here

- This provides a general 'handle turning' method of constructing a test of H_A against H_0
  - Assume H_0 is true and find $\log_e(L_{max})$
  - Assume H_A is true and find $\log_e(L_{max})$
  - Take $\lambda = 2 \times$ difference
  - Compare with χ^2_r
  - If improbably large (i.e. $\lambda > \chi^2_r$(5% crit value) then reject H_0 in favour of H_A

- Key formulæ (ct^d.):
  - See **p206**
  - In standard case the maximized likelihood involves the data only through |S | (i.e. det(S)
  - In more general case when $d = \bar{x} - \hat{\mu}$

    Then maximized likelihood has term

    $\log_e\{|S + dd'|\}$ instead of $\log_e\{|S|\}$
  - So $\lambda = 2[\log_e\{|S + dd'|\} - \log_e\{|S|\}]$

- **Univariate (p=1)**
- $N(\mu, \sigma^2)$
  - Sample mean & variance are independent
  - $(n-1) \times s^2 \sim \sigma^2 \chi^2_{n-1}$
  - **Chi-squared** dist^n on n–1 degrees of freedom scaled by σ^2

- **Multivariate p-dims**
- $N_p(\mu, \Sigma)$
  - Sample mean & variance are independent
  - $(n-1) \times S \sim W_p(\Sigma, n-1)$
  - p-dim **Wishart** dist^n on (n–1) degrees of freedom scaled by Σ

- Wishart distributions are MV generalisations of chi-squared
  - dist^n for a p×p symmetric matrix
    - i.e. a joint p.d.f for each [distinct] element
      - so it is a ½p(p+1)-dimensional distribution
        - upper triangle + leading diagonal elements
  - Properties analogous to χ^2
    - **Additivity on d.f. parameter for common scale**
      - Cochran's theorem in Design of Experiments &c., &c.,…
    - main use is as an intermediate step

- **Univariate (p=1)**
- N(0, 1)
  - $d \sim N(0,1)$, $m^2 \sim \chi^2_n$ (independently)
  - $t = x\sqrt{n/m} \Rightarrow t \sim t_n$
  - $\Rightarrow t^2 \sim F_{1,n}$

- **Multivariate p-dims**
- $N_p(0, I_p)$
  - $d \sim N_p(0,I_p)$, $M \sim W_p(I_p,n)$ (independently)
  - $T^2 = nd'M^{-1}d$
  - $\Rightarrow T^2 \sim T^2(p,n)$
  - **Hotelling's T^2-distr$^{n.}$**
  - p-dimensional version of the F-distribution

©NRJF, University of Sheffield, 2011/12 Semester 1

- NB: Although Hotelling's T^2-distribution is a 'p-dimensional version of the F-distribution' it is a ***univariate distribution***
  - i.e. the statistic T^2 is calculated from p-dimensions but it is a scalar
  - $T^2=nd'M^{-1}d$ which is $1\times 1 \times 1\times p \times p\times p \times p\times 1$ — $n \times d' \times M^{-1} \times d$

- **Univariate (p=1)**
- $N(\mu, \sigma^2)$
  - $x \sim N(\mu, \sigma^2)$
  - $d=(x-\mu)/\sigma$
  - $d \sim N(0,1)$
  - $(n-1)s^2/\sigma^2 \sim \chi^2_{n-1}$
  - $\Rightarrow (x-\mu)/s \sim t_{(n-1)}$
  - (N.B.:- the σ cancels & the n–1 cancels)
- **Multivariate p-dims**
- $N_p(\mu, \Sigma)$
  - $d=\Sigma^{-½}(x-\mu)$
  - $d \sim N_p(0,I_p)$
  - $M=(n-1)\Sigma^{-½}S\Sigma^{-½}$
  - $\Rightarrow M \sim W_p(I_p,(n-1))$
  - $\Rightarrow (x-\mu)'S^{-1}(x-\mu) \sim T^2(p,n-1)$
  - (N.B.:- the Σ 'cancels' & n–1 cancels)

- **Univariate (p=1)**
- $N(\mu, \sigma^2)$
  - $\bar{x} \sim N(\mu, \sigma^2/n)$
  - $n(\bar{x}-\mu)^2/s^2 \sim F_{1,n-1}$
- **Multivariate p-dims**
- $N_p(\mu, \Sigma)$
  - $\bar{x} \sim N(\mu, \Sigma/n)$
  - $n(\bar{x}-\mu)'S^{-1}(\bar{x}-\mu) \sim T^2(p,n-1)$
- Note that the n comes from the divisor in $\bar{x}$ and the n–1 from the d.f. in s^2

- Key formulæ:
  - see p209

$$n(\bar{x}-\mu)'S^{-1}(\bar{x}-\mu) \sim T^2(p,n-1)$$

- use this for performing 1- & 2-sample tests

$$T^2(p,n) \equiv \tfrac{np}{n-p+1}F_{p,n-p+1}$$

- use for finding p-value from F-tables

- Test H_0: $\mu = \mu_0$ vs H_A: $\mu \neq \mu_0$ by rejecting H_0 if

$$\tfrac{n(n-p)}{(n-1)p}(\bar{x}-\mu_0)'S^{-1}(\bar{x}-\mu_0) > F_{p,n-p}(5\%\text{ point})$$

- Calculate this 'by hand' for p=2 (?)
  - or use matrix facilities in package
  - Or `HotellingsT2(.)` in package `ICSNP`
    - see p212

- Test H_0: $\mu_1 = \mu_2$ vs H_A: $\mu_1 \neq \mu_2$ by rejecting H_0 if

$$\tfrac{n_1n_2(n-p-1)}{n(n-2)p}(\bar{x}_1-\bar{x}_2)'S^{-1}(\bar{x}_1-\bar{x}_2) > F_{p,n-p-1}(5\%\text{ point})$$

- Calculate this 'by hand' for p=2 (?)
  - or use matrix facilities in package
  - **OR use MANOVA in package**
    - **(if you have original data not summary statistics)**
    - see p212

- **Examples of calculations**
  - **1-sample T^2-test**
  - Data on head length of 1st and 2nd sons in 25 families (x_1 and x_2 respectively)
  - n=25
  - H_0: both mean head lengths are = 182
  - i.e. $\mu_0 = (182,182)'$ and $x = (x_1, x_2)'$

$$T^2 = n(\bar{x}-\mu_0)'S^{-1}(\bar{x}-\mu_0)$$

- **Data:**

$$\bar{x} = \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \end{pmatrix} = \begin{pmatrix} 185.72 \\ 183.84 \end{pmatrix}$$

$$S = \begin{pmatrix} 91.481 & 66.875 \\ 66.875 & 96.775 \end{pmatrix}$$

$$T^2 = n(\bar{x}-\mu_0)'S^{-1}(\bar{x}-\mu_0)$$

$$T^2 = n(\bar{x}-\mu_0)'S^{-1}(\bar{x}-\mu_0)$$

$$\bar{x}-\mu_0 = \begin{pmatrix} \bar{x}_1-\mu_{01} \\ \bar{x}_2-\mu_{02} \end{pmatrix} = \begin{pmatrix} 185.72-182 \\ 183.84-182 \end{pmatrix} = \begin{pmatrix} 3.72 \\ 1.84 \end{pmatrix}$$

$$n = 25, \quad p = 2$$

$$T^2 = 25(3.72,1.84)S^{-1}(3.72,1.84)'$$

$$S = \begin{pmatrix} 91.481 & 66.875 \\ 66.875 & 96.775 \end{pmatrix}$$

$$|S| = 91.481 \times 96.775 - 66.875^2 = 4380.81$$

$$S^{-1} = \tfrac{1}{4380.81}\begin{pmatrix} 96.775 & -66.875 \\ -66.875 & 91.481 \end{pmatrix}$$

$$T^2 = 25(3.72,1.84)S^{-1}(3.72,1.84)'$$

$$T^2 = \tfrac{25}{4380.81}(3.72,1.84)S^{-1}(3.72,1.84)'$$

$$= \tfrac{25}{4380.81}(3.72,1.84)\begin{pmatrix} 96.775 & -66.875 \\ -66.875 & 91.481 \end{pmatrix}\begin{pmatrix} 3.72 \\ 1.84 \end{pmatrix}$$

$$= \tfrac{25}{4380.81}(3.72,1.84)\begin{pmatrix} 96.775 \times 3.72 - 66.875 \times 1.84 \\ -66.875 \times 3.72 + 91.481 \times 1.84 \end{pmatrix}$$

$$= \tfrac{25}{4380.81}(3.72,1.84)\begin{pmatrix} 360.00 - 123.85 \\ -248.78 + 168.33 \end{pmatrix}$$

$$= \tfrac{25}{4380.81}(3.72,1.84)\begin{pmatrix} 236.15 \\ -80.45 \end{pmatrix}$$

$$= \tfrac{25}{4380.81}(3.72 \times 236.15 - 1.84 \times 80.45)$$

$$= \tfrac{25}{4380.81}730.45 = 4.17$$

$$T^2(p,n) \equiv \tfrac{np}{n-p+1}F_{p,n-p+1} \rightarrow T^2(p,n-1) \equiv \tfrac{(n-1)p}{n-p}F_{p,n-p}$$

$$F_{p,n-p} = \tfrac{n-p}{(n-1)p}T^2(p,n-1) = \tfrac{25-2}{48}4.17 = 2.00$$

$2.00 <$ $F_{2,23}(95\%) = 3.4$ so little evidence that the mean head lengths are not both equal to 182.0

- **Examples of calculations**
  - **2-sample T^2-test**
  - British Museum mummy pots
  - 2 batches: n_1=15 $n_2$10
  - 2 measures: rim & base circumferences
  - To test whether the two batches have the same mean circumferences

$$\tfrac{n_1n_2(n-p-1)}{n(n-2)p}(\bar{x}_1-\bar{x}_2)'S^{-1}(\bar{x}_1-\bar{x}_2) > F_{p,n-p-1}(5\%\text{ point})$$

- **Data**

$$\bar{x}_1 = \begin{pmatrix} 476.33 \\ 225.67 \end{pmatrix} \quad \bar{x}_2 = \begin{pmatrix} 544.00 \\ 242.50 \end{pmatrix}$$

$$S_1 = \begin{pmatrix} 1483.81 & 113.33 \\ 113.33 & 420.95 \end{pmatrix} \quad S_2 = \begin{pmatrix} 1915.55 & 786.11 \\ 786.11 & 784.72 \end{pmatrix}$$

$$\tfrac{n_1n_2(n-p-1)}{n(n-2)p}(\bar{x}_1-\bar{x}_2)'S^{-1}(\bar{x}_1-\bar{x}_2) > F_{p,n-p-1}(5\%\text{ point})$$

$$(\bar{x}_1-\bar{x}_2) = \begin{pmatrix} -67.67 \\ -16.83 \end{pmatrix}$$

$$S = [(n_1-1)S_1 + (n_2-1)S_2]/(n_1+n_2-2)$$

$$= \begin{pmatrix} 1652.75 & 376.59 \\ 376.59 & 563.30 \end{pmatrix}$$

$$\tfrac{n_1n_2(n-p-1)}{n(n-2)p}(\bar{x}_1-\bar{x}_2)'S^{-1}(\bar{x}_1-\bar{x}_2) > F_{p,n-p-1}(5\%\text{ point})$$

$$S^{-1} = 10^{-4}\begin{pmatrix} 7.14 & -4.77 \\ -4.77 & 20.94 \end{pmatrix}$$

$$\frac{n_1n_2(n-p-1)}{n(n-2)p} = \frac{10.15.(25-2-1)}{25.23.2} = 2.87$$

Gives F-statistic of 7.96 > 3.44 = $F_{2,22}$

conclude that there is strong evidence of a difference in mean circumferences between the two batches

$$\tfrac{n_1n_2(n-p-1)}{n(n-2)p}(\bar{x}_1-\bar{x}_2)'S^{-1}(\bar{x}_1-\bar{x}_2) > F_{p,n-p-1}(5\%\text{ point})$$

- **Computational Note:**
  - Quadratic forms such as those above can be easily calculated at sight:

$$(x,y)\begin{pmatrix} a & c \\ c & b \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = ax^2 + 2cxy + by^2$$

$$(x,y,z)\begin{pmatrix} a & d & e \\ d & b & f \\ e & f & c \end{pmatrix}\begin{pmatrix} x \\ y \\ z \end{pmatrix} =$$

$$ax^2 + by^2 + cz^2 + 2dxy + 2exz + 2fyz$$

- **Likelihood Ratio Tests**
  - General procedure for constructing hypothesis tests
  - Test statistic based on difference in maximized log-likelihoods under null & alternative hypotheses
  - General theory gives [asymptotic] distribution of test statistic
    - (typically a χ^2 distribution)
  - General theory says test is most powerful
  - Can implement procedure computationally

- **Examples**
  - Generalizations of univariate tests: e.g. standard 1 & 2 –sample tests on means
    - H_0: $\mu=\mu_0$ *vs* H_A: $\mu \neq \mu_0$
    - H_0: $\mu_1 = \mu_2$ *vs* H_A: $\mu_1 \neq \mu_2$
  - Hypotheses that can only arise in multi-dimensions
    - H_0: $\mu'\mu = r^2$ *vs* H_A: $\mu'\mu \neq r^2$
      - i.e. the mean lies on a sphere of radius r
    - H_0: $\Sigma = \lambda\Sigma_0$ *vs* H_A: $\Sigma \neq \lambda\Sigma_0$ (some scalar λ)
      - the variance matrix has a specified form

- **However**
  - rejection of a multivariate hypothesis only reveals that at least one aspect of it is untenable in the light of the data
  - e.g. rejecting H_0: $\mu = \mu_0$ means that there is evidence that $(\mu_1, \mu_2,\ldots,\mu_p) \neq (\mu_{01}, \mu_{02},\ldots,\mu_{0p})$
    - for at least one μ_i
  - H_0 could be false because of only one μ_i

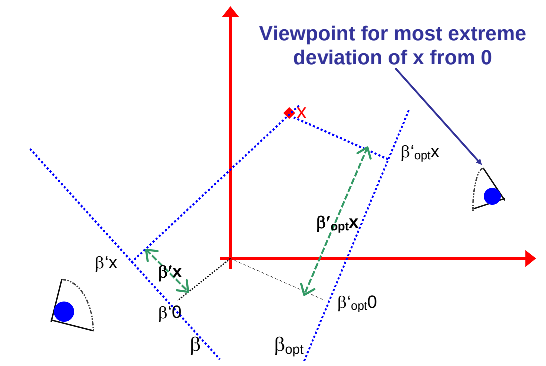
- In univariate case of $x\sim N(\mu, \sigma^2)$ if we reject H_0: $\mu = \mu_0$ we can see whether $\bar{x} > \mu_0$ or $\bar{x} < \mu_0$ and thus whether $\mu > \mu_0$ or $\mu < \mu_0$
- For multivariate μ we cannot say $\mu>\mu_0$
- i.e. do not know ***direction*** of departure from H_0.

- **Union-Intersection Tests**
  - Validity rests on the Cramér–Wold Theorem
    - gives connection between one-dimensional projections and the multivariate distribution
  - The distribution of a p-vector x is completely determined by the set of **all** 1-dimensional distributions of 1-dimensional projections of x, t′x, where t∈{all fixed p-vectors}
  - Proof is a result of uniqueness of characteristic functions

- C-W Theorem provides a way of making inferences about p-dimensions from one-dimensional projections
  - c.f. dimensionality reduction methods
- Need to know about **ALL** 1-dim projections
  - not just the p marginal variates
- e.g. if $X\sim N_p(\mu,\Sigma)$ then **ALL** 1-dim projections are normally distributed **and vice versa**
  - 1-dim projection ≡ linear combination
- But possible all p marginals Normal but **NOT** multivariate normal

- Idea is to project p-dim hypothesis into 1-dim and test all such projected 1-dim hypotheses
  - If any 1-dim hypothesis is rejected we know **which** projection 'failed'
  - i.e. we know ***direction*** of departure

- Typically we maximize a 1-dim test statistic wrt to the projection vector
  - e.g. $X \sim N_p(\mu,\Sigma)$ and hypothesis about μ &/or Σ
    - e.g. H_0: $\mu = \mu_0$
  - Then project into 1-dim by vector β & consider $Y = \beta'X \sim N_1(\beta'\mu, \beta'\Sigma\beta)$
    - e.g. $H_{0\beta}$: $\beta'\mu = \beta'\mu_0$
  - Find test statistic for $H_{0\beta}$ and maximise wrt β
  - c.f. finding direction β that gives most extreme evidence against H_0



- **Example**
  - $X \sim N_p(\mu,\Sigma)$; H_0: $\mu=0$ (Σ known)
  - consider $Y = \beta'X \sim N_1(\beta'\mu, \beta'\Sigma\beta)$
    - $H_{0\beta}$: $\beta'\mu = 0$
  - test statistic for $H_{0\beta}$ is

$$z_\beta = \left| \frac{\bar{y}}{\sqrt{\beta'\Sigma\beta / n}} \right|$$

And reject $H_{0\beta}$ if $z_\beta > c$ [suitable critical value]

- Reject H_0 if ***any*** $H_{0\beta}$ is rejected
- H_0 not rejected if ***no*** $H_{0\beta}$ is rejected i.e. if no $z_\beta >$ critical value i.e. if $\max\{z_\beta\} < c$
- Now

$$z_\beta^2 = \frac{n\bar{y}^2}{\beta'\Sigma\beta} = \frac{n\beta'\bar{x}\bar{x}'\beta}{\beta'\Sigma\beta}$$

- and this is maximized [using The Procedure] when β = [only] eigenvector of

$$\Sigma^{-1}\bar{x}\bar{x}'$$

- The [only] eigenvector of $\Sigma^{-1}\bar{x}\bar{x}'$ is $\Sigma^{-1}\bar{x}$
- This ***direction*** exhibits the greatest deviation from $\mu=0$
- Examining coefficients of variables in $\Sigma^{-1}\bar{x}$ will shew which variables in which combinations 'cause' rejection of H_0

♦ Substituting $\beta = \Sigma^{-1}\bar{x}$ in expression for z_β^2 gives

$$z_{\beta_{opt}}^2 = \frac{n\bar{x}'\Sigma^{-1}\bar{x}\,\bar{x}'\Sigma^{-1}\bar{x}}{\bar{x}'\Sigma^{-1}\bar{x}} = n\bar{x}'\Sigma^{-1}\bar{x}$$

and under H_0 this has an [exact] χ_p^2 distribution (not just an asymptotic as given by general MLE theory)

- NB $\bar{x}'\Sigma^{-1}\bar{x}$ is $1\times p \times p\times p\times p\times 1 = 1\times 1$

  i.e. a scalar, so cancels from top and bottom

Other Examples

♦ 2-sample test (c.f. task sheet 10)
- Test on data projected into 1-dimension gives a standard two-sample t-test.
- Maximizing wrt projection gives two-sample T²-test
- Direction of greatest difference is

$$S^{-1}(\bar{x}_1 - \bar{x}_2)$$

♦ Example: Iris Versicolor *vs* Virginica

Example: Iris Versicolor *vs* Virginica

- (cf task sheet 9)

We have

$$\bar{x}_1 - \bar{x}_2 = \begin{pmatrix} -0.652 \\ -0.210 \\ -1.290 \\ -0.700 \end{pmatrix} \quad S = \begin{pmatrix} .336 & .092 & .243 & .052 \\ * & .105 & .080 & .045 \\ * & * & .263 & .061 \\ * & * & * & .057 \end{pmatrix}$$

and then

$$S^{-1}(\bar{x}_1 - \bar{x}_2) = \begin{pmatrix} 3.524 \\ 5.610 \\ -6.988 \\ -12.460 \end{pmatrix}$$

♦ c.f. eigenvector from MANOVA

- differs only by a factor ~37.5

$$\begin{pmatrix} 0.095 \\ 0.150 \\ -0.187 \\ -0.333 \end{pmatrix}$$

Interpretation:

♦ Variable	coefficient in $S^{-1}(\bar{x}_1 - \bar{x}_2)$
Sepal-l	3.524
Sepal-w	5.610
Petal-l	-6.988
Petal-w	-12.460

♦ Greatest difference between varieties is exhibited in a direction contrasting size of sepals with size of petals

& particularly widths of sepals and petals

Other Examples

♦ Test of H_0: $\Sigma = \Sigma_0$ *vs* $\Sigma \neq \Sigma_0$ with μ unknown
- c.f. § 8.6.5

♦ UIT gives directions of deviations from H_0 as the smallest and largest eigenvectors of $\Sigma_0^{-1}S$ (i.e. corres. to λ_1 & λ_p) with test statistics λ_1 & λ_p rejecting H_0 if λ_1 is improbably big **or** if λ_p is improbably small
- Assess by simulating from $N_p(0, \Sigma_0)$ and calculating λ_1 & λ_p many times

- **Example: Iris Versicolor**
  - To test whether $\Sigma = \Sigma_0 = 0.158\, I_4$
    - (artificial example: 0.158 chosen as average trace of S, 'tests whether lengths and widths of sepals & petals are independent with equal variances)
  - We have (see Q3, Tasks 9) $S = \begin{pmatrix} 0.266 & 0.091 & 0.183 & 0.056 \\ \star & 0.106 & 0.089 & 0.043 \\ \star & \star & 0.221 & 0.073 \\ \star & \star & \star & 0.039 \end{pmatrix}$

- Eigenvalues of $\Sigma_0^{-1}S$ are
  - $0.158^{-1} \times (0.494, 0.073, 0.055, 0.010)$ $= (3.127, 0.462, 0.348, 0.063)$
    - So LRT statistic is ~172.61 on 10df (see §8.5.2)
- First eigenvector is
  - (0.681,0.326, 0.620, 0.214)′

  and last is
  - (0.100, -0.216, -0.314, 0.919)′

- To assess whether the UIT rejects H_0: $\Sigma = 0.158\, I_4$
  - need to simulate 50 samples from $N_4(0, 0.158\, I_4)$,
  - calculate the sample covariance matrix S
  - find the largest and smallest eigenvalues of 0.158^{-1} S
  - repeat this 1000 times (say)
  - see how many of these are > 3.13 or < 0.063 (the observed values)
- Then, reject H_0 at the 5% level if fewer than 5% of the simulated values are outside these bounds

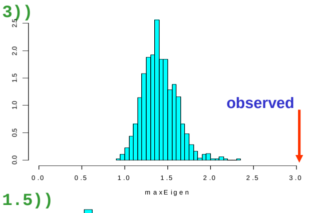
- **Interpretation (very difficult)**
  - First eigenvector is
    - (**0.681**,0.326, 0.**620**, 0.214)′

    and last is
    - (**0.100**, -0.216, -0.314, **0.919**)′
  - Direction of deviation is 'weighted' towards 1st & 3rd comps (i.e.lengths) $S = \begin{pmatrix} 0.266 & 0.091 & 0.183 & 0.056 \\ \star & 0.106 & 0.089 & 0.043 \\ \star & \star & 0.221 & 0.073 \\ \star & \star & \star & 0.039 \end{pmatrix}$
    - (which have largest covariance)
  - & 1st & 4th [with max & min variances] different from 2nd & 3rd

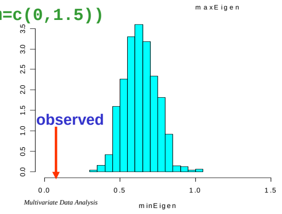

```
library(MASS)
V<-matrix(c(.266,.091,.183,.056,
+ .091,.106,.089,.043,
+ .183,.089,.221,.073,
+ .056,.043,.073,.039), nrow=4,ncol=4,byrow=T)
EigenV<-eigen(V)
sigma=diag(c(1,1,1,1))*0.158
sigma
mu<-matrix(c(0,0,0,0),nrow=4,ncol=1,byrow=T)
maxEigen<-numeric(1000)
minEigen<-numeric(1000)
for (i in 1:1000) {
X<-mvrnorm(n=50,mu,sigma)
S<-cov(X)/.158
eigenS<-eigen(S)
maxEigen[i]<-eigenS$values[1]
minEigen[i]<-eigenS$values[4]
}
```



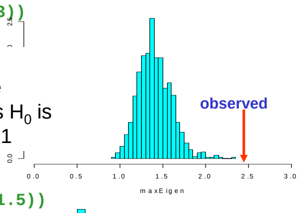
```
truehist(maxEigen,xlim=c(0,3))
```




```
truehist(minEigen,xlim=c(0,1.5))
```

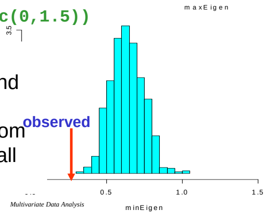


```
truehist(maxEigen,xlim=c(0,3))
```



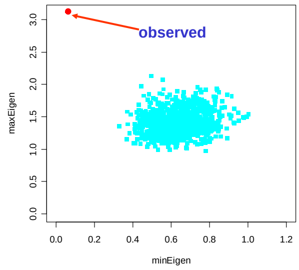
So none of the 1000 simulated values outside observed range and thus H_0 is rejected at a level < 0.001

```
truehist(minEigen,xlim=c(0,1.5))
```



Evidence both that the variances are unequal and that the covariances are non-zero so departure from scaled identity matrix in all respects


```
maxobs=EigenV$values[1]/0.158
minobs=EigenV$values[4]/0.158
plot(minEigen,maxEigen,xlim=c(0,1.2),ylim=c(0,3.1
),pch=15,col=5)
points(minobs,maxobs,pch=16,col=10,cex=1.5)
```



- **Key point:–**
  - UIT gives directions of deviations from H_0 as the first & last eigenvectors of $\Sigma_0^{-1}S$
    - gives directions of deviations
    - but need to use simulation test for p-value
  - **c.f. § 8.5.2:–** LRT for this problem gives [minimal] sufficient statistics in this case as sum and product of [all] eigenvalues of $\Sigma_0^{-1}S$
    - no specific direction of departure
    - but can refer LRT statistic to χ^2 on ½p(p+1) df
  - LRT more ***powerful*** but UIT more ***informative***

- **Multisample Tests — MANOVA**
- **Setup:–**
  - k ***known*** groups
  - n_i observations from each group, $\Sigma n_i = n$
  - Each observation is p-dimensional
    - **S** is the ***total variance***
    - **W** is the ***within groups variance***
    - **B** is the ***between groups variance***
  - Then it can be shewn that $(n-1)S = (k-1)B + (n-k)W$
    - – the one-way multivariate analysis of variance

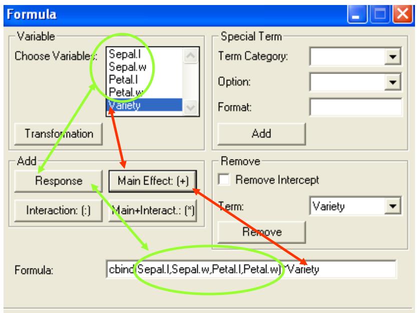
- Model is that data from i^{th} group $\sim N_p(\mu_i, \Sigma)$
  - – (common Σ note)
  - Standard hypothesis is H_0: $\mu_1 = \mu_2 = \ldots. = \mu_k$ *vs* at least one μ_i different
  - Various test statistics all based on eigenvalues of W and B

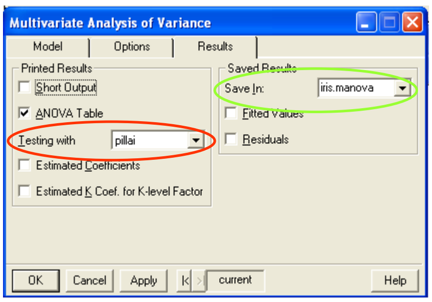
- **Likelihood Ratio Test**
  - Wilks Λ-test
    - Test statistic is $|W^{-1}B|$
      - – = product of all eigenvalues of $W^{-1}B$
- **UIT Test**
  - Roy's test
    - Test statistic is largest eigenvalue of $W^{-1}B$
    - Direction exhibiting greatest deviation from H_0 is corresponding eigenvector

- **Others**
  - Pillai Trace
    - Trace of $B(B+W)^{-1}$
      - = sum of eigenvlaues of $B(B+W)^{-1}$
  - Lawley-Hotelling Trace
    - Trace of $W^{-1}B$
      - = sum of eigenvlaues of $W^{-1}B$
- In case k = 2 all tests are the same
  - For k > 2 they are different
- **For k = 2 Hotelling's T²**
  = (n – 2) × Lawley-Hotelling trace

- **Key Advantage of MANOVA**:–
  - Partially overcomes problem of multiple testing of all variables separately
    - False positives, non-independent tests &c.
  - Use of UIT principle allows interpretation of which [combination] of variables exhibits deviation
- Generalizations to 2-way MANOVA & general multivariate linear model

- **R implementation:**–
  - Statistics>Multivariate>MANOVA
  - Create Formula
    - Variables are 'responses'
      - Sepal & Petal lengths & widths
    - Group indictor is 'main effect'
      - Variety
  - Results allow choice of Pillai (default), Roy, &c.
  - (or save object in 'Save In' & interrogate later)
  - or use command line `manova(.)`
    - `summary.manova(.)`






```
*** Multivariate Analysis of Variance Model ***

          Df Pillai Trace approx. F num df    den df    P-value
 Variety 2   .194    53.692         8       290         0

Residuals 147

> summary(iris.manova, test="hotelling-lawley")
          Df Hotelling-Lawley approx. F num df    den df
   Variety 2     32.759          585.562     8       286
Residuals 147

          P-value
  Variety   0
```

```
> attach(irisnf)
> iris.manova<-
   manova(cbind(Sepal.l,Sepal.w,Petal.l,Petal.w)~Variety)
> summary(iris.manova)
        Df Pillai Trace approx. F num df    den df
 Variety 2  .194          53.692      8       290
  0
Residuals 147

summary(iris.manova, test="r")
           Df Roy Largest approx. F num df    den df
  Variety 2      32.47      1177.05         4       145
Residuals 147

          P-value
  Variety       0
```


British Museum data, two circumferences

```
> attach(brmuseum)
> br.manova<-manova(cbind(rim.cir,base.circ)~batch)
> summary(br.manova)
     Df Pillai Trace approx. F    num df    den df  P-value
batch 1  0.4199       7.9619        2        22   0.0025
Residuals 23
```

For Hotelling's T^2 need Hotelling-Lawley trace:

```
> summary(br.manova,test="h")
     Df Hotelling-Lawley approx. F num df  den df  P-value
batch 1  0.7238            7.9619     2        22   0.0025
```

$T^2 = (n–2) \times \text{trace} = 23 \times 0.724 = 16.65$

So $F_{2,22} = (n – p+1)/np \times T^2$
$= [(25 – 2 –1)/25 \times 2] \times 16.65 = 7.962$

As calculated directly in earlier example (slide 380)

$$S^{-1} = 10^{-4}\begin{pmatrix} 7.14 & -4.77 \\ -4.77 & 20.94 \end{pmatrix}$$

$$\frac{n_1 n_2 (n-p-1)}{n(n-2)p} = \frac{10.15.(25-2-1)}{25.23.2} = 2.87$$

Gives F-statistic of 7.96 > 3.44 = $F_{2,22}$

conclude that there is strong evidence of a difference in mean circumferences between the two batches

$$\tfrac{n_1 n_2 (n-p-1)}{n(n-2)p}(\bar{x}_1 - \bar{x}_2)' S^{-1}(\bar{x}_1 - \bar{x}_2) > F_{p,n-p-1}(5\%\ \text{point})$$

- **Summary & Conclusions**
  - Univariate results extended to multivariate data
  - Introduction of Multivariate Distributions:
    - Normal
    - Wishart
    - Hotelling's T^2
  - sample mean and variance unbiased for population mean and variance
  - Normal ⇒
    - sample mean Normal
    - variance is Wishart
    - and they are **independent**

  - One and two-sample t-tests → T^2-tests
  - Generalized likelihood ratio tests (LRTs) for constructing hypotheses which cannot arise in one dimension
  - Union-Intersection Tests (UITs) provide an alternative strategy
    - similarities with multivariate EDA techniques such as PCA and LDA in construction and interpretation of 'directions'
- **standard packages have all standard tests**