STAT 730 Chapter 5: Hypothesis Testing

Timothy Hanson

Department of Statistics, University of South Carolina

Stat 730: Multivariate Analysis

def'n: Data $\mathbf{X}$ depend on $\boldsymbol{\theta}$. The likelihood ratio test statistic for $H_0 : \boldsymbol{\theta} \in \Omega_0$ vs. $H_1 : \boldsymbol{\theta} \in \Omega_1$is

$$\lambda(\mathbf{X}) = \frac{L_0^*}{L_1^*} = \frac{\max_{\boldsymbol{\theta} \in \Omega_0} L(\mathbf{X}; \boldsymbol{\theta})}{\max_{\boldsymbol{\theta} \in \Omega_1} L(\mathbf{X}; \boldsymbol{\theta})}.$$

def'n: The likelihood ratio test (LRT) of size α for testing H_0 vs. H_1 rejects H_0 when $\lambda(\mathbf{X}) < c$ where c solves $\sup_{\boldsymbol{\theta} \in \Omega_0} P_{\boldsymbol{\theta}}(\lambda(\mathbf{X}) < c) = \alpha$.

A very important large sample result for LRTs is

thm: For a LRT, if $\Omega_1 \subset \mathbb{R}^q$ and $\Omega_0 \subset \Omega_1$ be r-dimensional where $r < q$, then under some regularity conditions $-2 \log \lambda(\mathbf{X}) \xrightarrow{D} \chi^2_{q-r}$.

One-sample normal $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$

Let $\mathbf{x}_1, \ldots, \mathbf{x}_n \overset{iid}{\sim} N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ are unknown. We wish to test $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$.

Under H_0, $\hat{\boldsymbol{\mu}} = \boldsymbol{\mu}_0$ and $\hat{\boldsymbol{\Sigma}} = \mathbf{S} + (\bar{\mathbf{x}} - \boldsymbol{\mu}_0)(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)'$. Under H_1, $\hat{\boldsymbol{\mu}} = \bar{\mathbf{x}}$ and $\hat{\boldsymbol{\Sigma}} = \mathbf{S}$. We will show in class that

$$-2 \log \lambda(\mathbf{X}) = n \log\{1 + (\bar{\mathbf{x}} - \boldsymbol{\mu}_0)'\mathbf{S}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)\}.$$

Recall that $(n-1)(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)'\mathbf{S}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0) \sim T^2(p, n-1)$ when H_0 is true and has a scaled F-distribution.

Frets (1921) considers data on the head length and breadth of 1st and 2nd sons from $n = 25$ families. Let $\mathbf{x}_i = (\mathbf{x}_{i1}, \mathbf{x}_{i2})'$ be the head lengths (mm) of the first and second son in Frets' data. To test $H_0 : \boldsymbol{\mu} = (182, 182)'$ (p. 126) we can use Michail Tsagris' `hotel1T2` function in R. Note that your text incorrectly has $p = 3$ rather than $p = 2$ in the denominator of the test statistic.

```
install.packages("calibrate")
library("calibrate")
data(heads)
source("http://www.stat.sc.edu/~hansont/stat730/paketo.R")
colMeans(heads)
hotel1T2(heads[,c(1,3)],c(182,182),R=1) # normal theory
```

Bootstrapped p-value

The underlying assumption here is normality. If this assumption is suspect and the sample sizes are small, we can sometimes perform a nonparametric bootstrap.

The bootstrapped approximation to the sampling distribution under H_0 is simple to obtain. Simply resample the same number(s) of vectors *with replacement*, forcing H_0 upon the samples, and compute the sample statistic over and over again. This provides a Monte Carlo estimate of the sampling distribution.

Michail Tsagris' function automatically obtains a bootstrapped p-value via `hotel1T2(heads[,c(1,3)],c(182,182),R=2000)`. `R` is the number of bootstrapped samples used to compute the p-value. The larger `R` is the more accurate the estimate, but the longer it takes.

One-sample normal $H_0 : \mathbf{\Sigma} = \mathbf{\Sigma}_0$

Let $\mathbf{x}_1, \ldots, \mathbf{x}_n \stackrel{iid}{\sim} N_p(\boldsymbol{\mu}, \mathbf{\Sigma})$ where $(\boldsymbol{\mu}, \mathbf{\Sigma})$ are unknown. We wish to test $H_0 : \mathbf{\Sigma} = \mathbf{\Sigma}_0$.

Under H_0, $\hat{\boldsymbol{\mu}} = \bar{\mathbf{x}}$ and $\hat{\mathbf{\Sigma}} = \mathbf{\Sigma}_0$. Under H_1, $\hat{\boldsymbol{\mu}} = \bar{\mathbf{x}}$ and $\hat{\mathbf{\Sigma}} = \mathbf{S}$. We will show in class that

$$-2 \log \lambda(\mathbf{X}) = n \text{ tr } \mathbf{\Sigma}_0^{-1}\mathbf{S} - n \log |\mathbf{\Sigma}_0^{-1}\mathbf{S}| - np.$$

Note that the statistic depends on the eigenvalues of $\mathbf{\Sigma}_0^{-1}\mathbf{S}$. Your book discusses small sample results. For large samples we can use the usual $\chi^2_{p(p+1)/2}$ approximation.

```
cov.equal(heads[,c(1,3)],diag(c(100,100)))     # 1st hypothesis, p. 127
cov.equal(heads[,c(1,3)],matrix(c(100,50,50,100),2,2)) # 2nd hypothesis
```

A parametric bootstrapped p-value can also be obtained here.

Parametric bootstrap

The parametric bootstrap conditions on sufficient statistics under H_0, typically nuisance parameters, to compute the p-value. Conditioning on a sufficient statistic is a common approach in hypothesis testing (e.g. Fisher's test of homogeneity for 2×2 tables). Since the sampling distribution relies on the underlying parametric model, a parametric bootstrap may be sensitive to parametric assumptions, unlike the nonparametric bootstrap.

Recall that MLEs are functions of sufficient statistics. The parametric boostrap proceeds by simulating R samples $\mathbf{X}_r^* = [\mathbf{x}_{r1}^* \cdots \mathbf{x}_{rn}^*]'$ of size n from the parametric model under the MLE from H_0 and computing $\lambda(\mathbf{X}_r^*)$. The p-value is $\hat{p} = \frac{1}{R} \sum_{r=1}^{R} I\{\lambda(\mathbf{X}_r^*) < \lambda(\mathbf{X})\}$.

Let's see how this works for the test of $H_0 : \mathbf{\Sigma} = \mathbf{\Sigma}_0$.

Parametric bootstrap

```
Sigma0=matrix(c(100,50,50,100),2,2)
# Sigma0=matrix(c(100,0,0,100),2,2)
lambda=function(x,Sigma0){
 n=dim(x)[1]; p=dim(x)[2]
 S0inv=solve(Sigma0); S=(n-1)*cov(x)/n
 n*sum(diag(S0inv%*%S))-n*det(S0inv%*%S)-n*p
}
ts=lambda(heads[,c(1,3)],Sigma0)
xd=heads[,c(1,3)]; n=dim(xd)[1]; p=dim(xd)[2]
xbar=colMeans(heads[,c(1,3)]); S0root=t(chol(Sigma0)) # MLEs under H0
R=2000; pval=0
for(i in 1:R){
 xd=t(S0root%*%matrix(rnorm(p*n),p,n)+xbar)
 if(lambda(xd,Sigma0)>ts){pval=pval+1}
}
cat("Parametric bootstrapped p-value = ",pval/R,"\n")
```

Note that parametric bootstraps can take a long time to complete depending on the model, hypothesis, and sample size.

Union-intersection test of $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$

Often a multivariate hypothesis can be written as the union of univariate hypotheses. $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ holds $\Leftrightarrow$ $H_{0\mathbf{a}} : \mathbf{a}'\boldsymbol{\mu} = \mathbf{a}'\boldsymbol{\mu}_0$ holds for every $\mathbf{a} \in \mathbb{R}^p$. In this sense, $H_0 = \cap_{\mathbf{a}\in\mathbb{R}^p} H_{0\mathbf{a}}$.

For each $\mathbf{a}$ we have a simple univariate hypothesis, based on univariate data $\mathbf{a}'\mathbf{x}_1, \ldots, \mathbf{a}'\mathbf{x}_n \overset{iid}{\sim} N(\mathbf{a}'\boldsymbol{\mu}, \mathbf{a}'\boldsymbol{\Sigma}\mathbf{a})$, and testable using the usual t-test

$$t_{\mathbf{a}} = \frac{\mathbf{a}'\bar{\mathbf{x}} - \mathbf{a}'\boldsymbol{\mu}_0}{\sqrt{\mathbf{a}'\mathbf{S}_u\mathbf{a}/n}}.$$

We reject $H_{0\mathbf{a}}$ if $|t_{\mathbf{a}}| > t_{n-1}(1 - \frac{\alpha}{2})$, which means we reject H_0 *if any* $|t_{\mathbf{a}}| > t_{n-1}(1 - \frac{\alpha}{2})$, i.e. the union of all rejection regions, which happens when $\max_{\mathbf{a}\in\mathbb{R}^p} |t_{\mathbf{a}}| > t_{n-1}(1 - \frac{\alpha}{2})$.

UIT of $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$, continued

Squaring both sides of $\max_{\mathbf{a} \in \mathbb{R}^p} |t_{\mathbf{a}}| > t_{n-1}(1 - \frac{\alpha}{2})$, the multivariate test statistic can be rewritten

$$\max_{\mathbf{a} \in \mathbb{R}^p} t_{\mathbf{a}}^2 = \max_{\mathbf{a} \in \mathbb{R}^p} (n-1) \frac{\mathbf{a}'(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)'\mathbf{a}}{\mathbf{a}'\mathbf{S}\mathbf{a}} = (n-1)(\bar{\mathbf{x}} - \boldsymbol{\mu}_0)'\mathbf{S}^{-1}(\bar{\mathbf{x}} - \boldsymbol{\mu}_0),$$

the one-sample Hotelling's T^2. We used a maximization result from Chapter 4 here.

For $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$, the LRT and the UIT give the same test statistic.

$H_0 : \mathbf{R}\boldsymbol{\mu} = \mathbf{r}$

Here, $\mathbf{r} \in \mathbb{R}^q$. From Chapter 4 we have $\hat{\boldsymbol{\mu}} = \bar{\mathbf{x}} - \mathbf{S}\mathbf{R}'[\mathbf{R}\mathbf{S}\mathbf{R}']^{-1}(\mathbf{R}\bar{\mathbf{x}} - \mathbf{r})$. We can then show

$$(n-1)[\lambda(\mathbf{X})^{-2/n} - 1] = (n-1)(\mathbf{R}\bar{\mathbf{x}} - \mathbf{r})'(\mathbf{R}\mathbf{S}\mathbf{R}')^{-1}(\mathbf{R}\bar{\mathbf{x}} - \mathbf{r}).$$

Your book shows that $(n-1)[\lambda(\mathbf{X})^{-2/n} - 1] \sim T^2(q, n-1)$. This is a special case of the general test discussed in Chapter 6.

- If $\boldsymbol{\mu}' = (\boldsymbol{\mu}_1', \boldsymbol{\mu}_2')$, then $H_0 : \boldsymbol{\mu}_2 = \mathbf{0}$ is this type of test where $\mathbf{R} = [\mathcal{I}\ \mathbf{0}]$ and $\mathbf{r} = \mathbf{0}$.
- $H_0 : \boldsymbol{\mu} = \kappa\boldsymbol{\mu}_0$ where $\boldsymbol{\mu}_0$ is given is also a special case. Here, the $k = p - 1$ rows of $\mathbf{R}$ are orthogonal to $\boldsymbol{\mu}_0$ and $\mathbf{r} = \mathbf{0}$.

The UIT gives the same test statistic.

Test for sphericity $H_0 : \boldsymbol{\Sigma} = \kappa \mathcal{I}_p$

On p. 134, MKB derive

$$-2 \log \lambda(\mathbf{X}) = np \log \left\{ \frac{\frac{1}{p}\text{tr } \mathbf{S}}{|\mathbf{S}|^{1/p}} \right\}.$$

Asymptotically, this is $\chi^2_{(p-1)(p+2)/2}$.

```
ts=function(x){
 n=dim(x)[1]; p=dim(x)[2]
 S=cov(x)*(n-1)/n
 d=n*p*(log(sum(diag(S))/p)-log(det(S))/p)
 cat("Asymptotic p-value = ",1-pchisq(d,0.5*(p-1)*(p+2)),"\n")
}

ts(heads)
```

A parametric bootstrap can also be employed. R has `mauchly.test` built in for testing sphericity.

Testing $H_0 : \mathbf{\Sigma}_{12} = \mathbf{0}$, i.e. $\mathbf{x}_{1i}$ indep. $\mathbf{x}_{2i}$

Let $\mathbf{x}_i' = (\mathbf{x}_{1i}', \mathbf{x}_{2i}')$, where $\mathbf{x}_{1i} \in \mathbb{R}^{p_1}$, $\mathbf{x}_{2i} \in \mathbb{R}^{p_2}$ and $p_1 < p_2$. We can always rearrange elements via a permutation matrix to test any subset is indpendent of another subset. Partition $\mathbf{\Sigma}$ and $\mathbf{S}$ accordingly. MKB (p. 135) find

$$-2 \log \lambda(\mathbf{X}) = -n \log \prod_{i=1}^{p_1} (1 - \lambda_i) = -n \ \log |\mathcal{I} - \mathbf{S}_{11}^{-1} \mathbf{S}_{12} \mathbf{S}_{22}^{-1} \mathbf{S}_{21}|,$$

where λ_i are the e-values of $\mathbf{S}_{11}^{-1} \mathbf{S}_{12} \mathbf{S}_{22}^{-1} \mathbf{S}_{21}$.

Exactly, $\lambda^{2/n} \sim \Lambda(p_1, n - 1 - p_2, p_2)$ assuming $n - 1 \geq p$. When both $p_i > 2$ we can use
$-(n - \frac{1}{2}(p + 3)) \log |\mathcal{I} - \mathbf{S}_{11}^{-1} \mathbf{S}_{12} \mathbf{S}_{22}^{-1} \mathbf{S}_{21}| \sim \chi^2_{p_1 p_2}$.

If $p_1 = 1$ then $-2 \log \lambda = -n \log(1 - R^2)$ where R^2 is the sample multiple correlation coefficient between x_{1i} and $\mathbf{x}_{2i}$, discussed in the next Chapter.

The UIT gives the largest e-value λ_1 of $\mathbf{S}_{11}^{-1} \mathbf{S}_{12} \mathbf{S}_{22}^{-1} \mathbf{S}_{21}$ as the test statistic, different from the LRT.

Head length/breadth data

Say we want to test that the head/breadths are independent between sons. Let $\mathbf{x}_i = (x_{i1}, x_{i2}, x_{i3}, x_{i4})'$ be the 1st son's length & breadth followed by the 2nd son's. We want to test

$H_0 : \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{22} \end{bmatrix}$ where all submatrices are 2×2.

```
S=24*cov(heads)/25
S2=solve(S[1:2,1:2])%*%S[1:2,3:4]%*%solve(S[3:4,3:4])%*%S[3:4,1:2]
ts=det(diag(2)-S2) # ~WL(p1,n-1-p2,p2)=WL(2,22,2)
# p. 83 => (21/2)*(1-sqrt(ts))/sqrt(ts)~F(4,42)
1-pf((21/2)*(1-sqrt(ts))/sqrt(ts),4,42)
ats=-(25-0.5*7)*log(ts) # ~chisq(4) asymptotically
1-pchisq(ats,4) # asymptotics work great here!
```

What do we conclude about how head shape is related between 1st and 2nd sons?

$H_0 : \boldsymbol{\Sigma}$ is diagonal

We may want to test that all variables are independent. Under H_0 $\hat{\boldsymbol{\mu}} = \bar{\mathbf{x}}$ and $\hat{\boldsymbol{\Sigma}} = \text{diag}(s_{11}, \ldots, s_{pp})$. Then

$$-2 \log \lambda(\mathbf{X}) = -n \log |\mathbf{R}|.$$

This is approximately $\chi^2_{p(p-1)/2}$ in large samples. The parametric bootstrap can be used here as well.

```
1-pchisq(-25*log(det(cov2cor(S))),6)
```

One-way MANOVA, LRT

Want to test $H_0 : \boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_k$ assuming $\boldsymbol{\Sigma}_1 = \cdots = \boldsymbol{\Sigma}_k = \boldsymbol{\Sigma}$.

Under H_0, $\hat{\boldsymbol{\mu}} = \bar{\mathbf{x}}$ and $\hat{\boldsymbol{\Sigma}}_0 = \mathbf{S}$ from the entire sample. Under H_a, $\hat{\boldsymbol{\mu}}_i = \bar{\mathbf{x}}_i$ and $\hat{\boldsymbol{\Sigma}}_a = \frac{1}{n}\sum_{i=1}^k n_i \mathbf{S}_i$. Here, $n = n_1 + \cdots + n_k$. Then

$$\lambda(\mathbf{X})^{2/n} = \frac{|\hat{\boldsymbol{\Sigma}}_a|}{|\hat{\boldsymbol{\Sigma}}_0|} = \frac{|\mathbf{W}|}{|\mathbf{T}|} = |\mathbf{W}\mathbf{T}^{-1}|,$$

where $\mathbf{T} = n\mathbf{S}$ is the total sums of squares and cross products (SSCP) and $\mathbf{W} = \sum_{i=1}^k n_i \mathbf{S}_i$ is the SSCP for error, or within groups SSCP. The SSCP for regression is $\mathbf{B} = \mathbf{T} - \mathbf{W}$, or between groups. Then

$$\lambda(\mathbf{X})^{2/n} = \frac{|\mathbf{W}|}{|\mathbf{B} + \mathbf{W}|} = \frac{1}{|\mathcal{I}_p + \mathbf{W}^{-1}\mathbf{B}|}.$$

A bit different than usual F-test

When $p = 1$ we have

$$\lambda(\mathbf{X})^{2/n} = \frac{1}{|\mathcal{I}_p + \mathbf{W}^{-1}\mathbf{B}|} = \frac{1}{1 + \frac{SSR}{SSE}} = \frac{1}{1 + \frac{k-1}{n-p}F^*},$$

where $F^* = \frac{MSR}{MSE}$.

The usual F-statistic is a monotone function of $\lambda(\mathbf{X})$ and vice-versa. For $p = 1$, $\lambda(\mathbf{X})$ is distributed beta.

ANOVA decomposition

Let $\mathbf{X}(n \times p)' = [\mathbf{X}_1' \cdots \mathbf{X}_k']$ be the k d.m. stacked on top of each other.

Recall that for $p = 1$

$$SSE = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2 = \mathbf{X}'[\mathcal{I}_n - \mathbf{P_Z}]\mathbf{X} \stackrel{def}{=} \mathbf{X}'\mathbf{C}_1\mathbf{X},$$

where $\mathbf{P_Z} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$ and $\mathbf{Z} = \text{block-diag}(\mathbf{1}_{n_1}, \ldots, \mathbf{1}_{n_k})$.

For $p > 1$ it is the same!

$$\mathbf{E} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)(\mathbf{x}_{ij} - \bar{\mathbf{x}}_i)' = \mathbf{X}'[\mathcal{I}_n - \mathbf{P_Z}]\mathbf{X} = \mathbf{X}'\mathbf{C}_1\mathbf{X}.$$

Note that $\mathbf{P_Z} = \text{block-diag}(\frac{1}{n_1}\mathbf{1}_{n_1}\mathbf{1}_{n_1}', \ldots, \frac{1}{n_k}\mathbf{1}_{n_k}\mathbf{1}_{n_k}')$.

Similarly, for $p = 1$

$$SSR = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (\bar{x}_i - \bar{x})^2 = \mathbf{X}'[\mathbf{P}_{\mathbf{Z}} - \mathbf{P}_{\mathbf{1}_n}]\mathbf{X} \stackrel{def}{=} \mathbf{X}'\mathbf{C}_2\mathbf{X},$$

where $\mathbf{P}_{\mathbf{1}_n} = \frac{1}{n}\mathbf{1}_n\mathbf{1}_n'$. This generalizes for $p > 1$ to

$$\mathbf{B} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (\bar{\mathbf{x}}_i - \bar{\mathbf{x}})(\bar{\mathbf{x}}_i - \bar{\mathbf{x}})' = \mathbf{X}'[\mathbf{P}_{\mathbf{Z}} - \mathbf{P}_{\mathbf{1}_n}]\mathbf{X} = \mathbf{X}'\mathbf{C}_2\mathbf{X}.$$

ANOVA decomposition

Finally, for $p = 1$ we have

$$SST = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2 = \mathbf{X}'[\mathcal{I}_n - \mathbf{P}_{\mathbf{1}_n}]\mathbf{X},$$

which generalizes for $p > 1$ to

$$\mathbf{T} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (\mathbf{x}_{ij} - \bar{\mathbf{x}})(\mathbf{x}_{ij} - \bar{\mathbf{x}})' = \mathbf{X}'[\mathcal{I}_n - \mathbf{P}_{\mathbf{1}_n}]\mathbf{X}.$$

Now note that

$$\begin{aligned} \mathbf{T} &= \mathbf{X}'[\mathcal{I}_n - \mathbf{P}_{\mathbf{1}_n}]\mathbf{X} = \mathbf{X}'[\mathcal{I}_n - \mathbf{P}_{\mathbf{Z}} + \mathbf{P}_{\mathbf{Z}} - \mathbf{P}_{\mathbf{1}_n}]\mathbf{X} \\ &= \mathbf{X}'[\mathcal{I}_n - \mathbf{P}_{\mathbf{Z}}]\mathbf{X} + \mathbf{X}'[\mathbf{P}_{\mathbf{Z}} - \mathbf{P}_{\mathbf{1}_n}]\mathbf{X} \\ &= \mathbf{E} + \mathbf{B}. \end{aligned}$$

One-way MANOVA, LRT

Note that $\mathbf{X}$ d.m. $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ under $H_0 : \boldsymbol{\mu}_1 = \cdots \boldsymbol{\mu}_k$. Then $\mathbf{W} = \mathbf{X}'\mathbf{C}_1\mathbf{X}$, $\mathbf{B} = \mathbf{X}'\mathbf{C}_2\mathbf{X}$, where $\mathbf{C}_1$ and $\mathbf{C}_2$ are projection matrices of ranks $n - k$ and $k - 1$, and $\mathbf{C}_1\mathbf{C}_2 = \mathbf{0}$. Using Cochran's theorem and Craig's theorem,

$$\mathbf{W} \sim W_p(\boldsymbol{\Sigma}, n-k) \text{ indep. } \mathbf{B} \sim W_p(\boldsymbol{\Sigma}, k-1),$$

so then

$$\lambda^{2/n} \sim \Lambda(p, n-k, k-1),$$

under H_0 provided $n \geq p + k$.

On p. 139 your book argues that the test statistic for the UIT is the largest e-value of $\mathbf{W}^{-1}\mathbf{B}$.

The LRT and UIT lead to different test statistics, but they are based on the same matrix. There are actually four statistics in common use in multivariate regression settings.

Let $\lambda_1 > \cdots > \lambda_p$ be e-values from $\mathbf{W}^{-1}\mathbf{B}$. Then Roy's greatest root is λ_1, Wilk's lambda is $\prod_{i=1}^{p} \frac{1}{1+\lambda_i}$, Pillai-Bartlett trace is $\sum_{i=1}^{p} \frac{\lambda_i}{1-\lambda_i}$, and Hotelling-Lawley trace is $\sum_{i=1}^{p} \lambda_i$. Note that the one-way model is a special case of the general regression model in Chapter 6. MANOVA is further explored in Chapter 12.

The Hotelling's two-sample test of $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ is a special case of MANOVA where $k = 2$. In this case, Wilk's lambda boils down to

$$\lambda^{2/n} = 1 + n_1 n_2 (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)' \mathbf{S}_u^{-1} (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2).$$

The last term differs from the Hotelling's two-sample test statistic by a factor of n.

Behrens-Fisher problem

How to compare two means $H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2$ when $\boldsymbol{\Sigma}_1 \neq \boldsymbol{\Sigma}_2$? A standard LRT approach works but requires numerical optimization to obtain the MLEs; your book describes an iterative procedure.

The UIT approach fares better here. On pp. 143–144 a UIT test procedure is described that yields a test statistic that is approximately distributed Hotelling's T^2 with *df* computed using Welch's (1947) approximation. Tsagris rather implements an approach due to James (1954).

An alternative, for $k \geq 2$, is to test $H_0 : \boldsymbol{\mu}_1 = \cdots = \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_1 = \cdots = \boldsymbol{\Sigma}_k$ (complete homogeneity) vs. the alternative that the means and covariances are all different. This is the hypothesis that data all come from one population vs. k separate populations. MKB pp. 141–142 show $-2 \log \lambda = n \log |\frac{1}{n} \sum_{i=1}^{k} n_i \mathbf{S}_i| - \sum_{i=1}^{k} n_i \log |\mathbf{S}_i|$. This is asymptotically $\chi^2_{p(k-1)(p+3)/2}$.

The James (1954) approach can be extended to $k \geq 2$; Tsagris implements this in `maovjames`.

Dental data

```
library(reshape) # need to turn $\by$ into $\bY$.
library(car)     # allows for multivariate linear hypotheses
library(heavy)   # has dental data
data(dental)
d2=cast(melt(dental,id=c("Subject","age","Sex")),Subject+Sex~age)
names(d2)[3:6]=c("d8","d10","d12","d14")
# Hotelling's T-test, exact p-value and bootstrapped
hotel2T2(d2[d2$Sex=='Male',3:6],d2[d2$Sex=='Female',3:6],R=1)
hotel2T2(d2[d2$Sex=='Male',3:6],d2[d2$Sex=='Female',3:6],R=2000)
# MANOVA gives same p-value as Hotelling's parametric test
f=lm(cbind(d8,d10,d12,d14)~Sex,data=d2)
summary(Anova(f))
# James' T-tests do not assume the same covariance matrices
james(d2[d2$Sex=='Male',3:6],d2[d2$Sex=='Female',3:6],R=1)
james(d2[d2$Sex=='Male',3:6],d2[d2$Sex=='Female',3:6],R=2000)
```

Here's a one-way MANOVA on the iris data.

```
library(car)
scatterplotMatrix(~Sepal.Length+Sepal.Width+Petal.Length+Petal.Width|Species,
 data=iris,smooth=FALSE,reg.line=F,ellipse=T,by.groups=T,diagonal="none")
f=lm(cbind(Sepal.Length,Sepal.Width,Petal.Length,Petal.Width)~Species,
 data=iris)
summary(Anova(f))

f=manova(cbind(Sepal.Length,Sepal.Width,Petal.Length,Petal.Width)~Species,
 data=iris)
summary(f) # can ask for other tests besides Pillai

maovjames(iris[,1:4],as.numeric(iris$Species),R=1000) # bootstrapped
```

Homogeneity of covariance

Now we test $H_0 : \mathbf{\Sigma}_1 = \cdots = \mathbf{\Sigma}_k$ in the general model with differing means and covariances. Your book argues

$$-2 \log \lambda(\mathbf{X}) = n \log |\mathbf{S}| - \sum_{i=1}^{k} n_i \log |\mathbf{S}_i| = \sum_{i=1}^{k} n_i \log |\mathbf{S}_i^{-1}\mathbf{S}|.$$

Asymptotically, this has a $\chi^2_{p(p+1)(k-1)/2}$ distribution. Using Tsagris' functions, try

```
cov.likel(iris[1:4],as.numeric(iris$Species))
cov.Mtest(iris[1:4],as.numeric(iris$Species)) # Box's M-test
```

Your book discusses that the test can be improved in smaller samples (Box, 1949).

Non-normal data

MKB suggest the use of multivariate skew and kurtosis as statistics for assessing multivariate normality. (p. 149).

They further point out that, broadly, for non-normal data the normal-theory tests on means are sensitive to $\beta_{1,p}$ whereas tests on covariance are sensitive to $\beta_{2,p}$.
Both tests, along with some others, are performed in the the `MVN` package. They are also in the `psych` package.

```
library(MVN)
cork=read.table("http://www.stat.sc.edu/~hansont/stat730/cork.txt",header=T)
mardiaTest(cork,qqplot=T)
```

Another option is to examine the sample Mahalanobis distances $D_i^2 = (\mathbf{x}_i - \bar{\mathbf{x}})'\mathbf{S}^{-1}(\mathbf{x}_i - \bar{\mathbf{x}})$, or stratified versions of these for multi-sample situations. These are approximately χ_p^2. The `mardiaTest` function provides a Q-Q plot of the M-distances compared to what is expected under multivariate normality.