Inferences about Multivariate Means

Nathaniel E. Helwig

Assistant Professor of Psychology and Statistics
University of Minnesota (Twin Cities)

Updated 16-Jan-2017
Copyright © 2017 by Nathaniel E. Helwig
Outline of Notes

1) Single Mean Vector
   - Introduction
   - Hotelling’s $T^2$
   - Likelihood Ratio Tests
   - Confidence Regions
   - Simultaneous $T^2$ CIs
   - Bonferroni’s Method
   - Large Sample Inference
   - Prediction Regions

2) Multiple Mean Vectors
   - Introduction
   - Paired Comparisons
   - Repeated Measures
   - Two Populations
   - One-Way MANOVA
   - Simultaneous CIs
   - Equal Cov Matrix Test
   - Two-Way MANOVA
Inferences about a Single Mean Vector
Univariate Reminder: Student’s One-Sample $t$ Test

Let $(x_1, \ldots, x_n)$ denote a sample of iid observations sampled from a normal distribution with mean $\mu$ and variance $\sigma^2$, i.e., $x_i \sim \text{N}(\mu, \sigma^2)$.

Suppose $\sigma^2$ is unknown, and we want to test the hypotheses

$$H_0 : \mu = \mu_0 \quad \text{versus} \quad H_1 : \mu \neq \mu_0$$

where $\mu_0$ is some known value specified by the null hypothesis.

We use Student’s $t$ test, where the $t$ test statistic is given by

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

where $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$ and $s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$. 
Univariate Reminder: Student’s $t$ Test (continued)

Under $H_0$, the $t$ test statistic follows a Student’s $t$ distribution with degrees of freedom $\nu = n - 1$.

- We reject $H_0$ if $|t|$ is large relative to what we would expect
- Same as rejecting $H_0$ if $t^2$ is larger than we expect

We can rewrite the (squared) $t$ test statistic as

$$t^2 = n(\bar{x} - \mu_0)(s^2)^{-1}(\bar{x} - \mu_0)$$

which emphasizes the quadratic form of the $t$ test statistic.

- $t^2$ gets larger as $(\bar{x} - \mu_0)$ gets larger (for fixed $s^2$ and $n$)
- $t^2$ gets larger as $s^2$ gets smaller (for fixed $\bar{x} - \mu_0$ and $n$)
- $t^2$ get larger as $n$ gets larger (for fixed $\bar{x} - \mu_0$ and $s^2$)
Now suppose that $x_i \overset{iid}{\sim} N(\mu, \Sigma)$ where
- $x_i = (x_{i1}, \ldots, x_{ip})'$ is the $i$-th observation’s $p \times 1$ vector
- $\mu = (\mu_1, \ldots, \mu_p)'$ is the $p \times 1$ mean vector
- $\Sigma = \{\sigma_{jk}\}$ is the $p \times p$ covariance matrix

Suppose $\Sigma$ is unknown, and we want to test the hypotheses

$$H_0 : \mu = \mu_0 \quad \text{versus} \quad H_1 : \mu \neq \mu_0$$

where $\mu_0$ is some known vector specified by the null hypothesis.
Hotelling’s $T^2$ Test Statistic

Hotellings $T^2$ is multivariate extension of (squared) $t$ test statistic

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

where

- $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$ is the sample mean vector
- $S = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})'$ is the sample covariance matrix
- $\frac{1}{n} S$ is the sample covariance matrix of $\bar{x}$

Letting $X = \{x_{ij}\}$ denote the $n \times p$ data matrix, we could write

- $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i = n^{-1} X'1_n$
- $S = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})' = \frac{1}{n-1} X_c'X_c$
- $X_c = CX$ with $C = I_n - \frac{1}{n} 1_n 1_n'$ denoting a centering matrix
Inferences using Hotelling’s $T^2$

Under $H_0$, Hotelling’s $T^2$ follows a scaled $F$ distribution

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

where $F_{p,n-p}$ denotes an $F$ distribution with $p$ numerator degrees of freedom and $n - p$ denominator degrees of freedom.

- This implies that $\alpha = P(T^2 > \left[ \frac{p(n-1)}{(n-p)} \right] F_{p,n-p}(\alpha))$
- $F_{p,n-p}(\alpha)$ denotes upper $(100\alpha)$th percentile of $F_{p,n-p}$ distribution

We reject the null hypothesis if $T^2$ is too large, i.e., if

$$T^2 > \frac{(n-1)p}{(n-p)} F_{p,n-p}(\alpha)$$

where $\alpha$ is the significance level of the test.
Inferences about a Single Mean Vector

Comparing Student’s $t^2$ and Hotelling’s $T^2$

Student’s $t^2$ and Hotelling’s $T^2$ have a similar form

$$T^2_{p,n-1} = \sqrt{n}(\bar{x} - \mu_0)'[S]^{-1}\sqrt{n}(\bar{x} - \mu_0)$$

$$= (\text{MVN vector})' \left( \frac{\text{Wishart matrix}}{df} \right)^{-1} (\text{MVN vector})$$

$$t^2_{n-1} = \sqrt{n}(\bar{x} - \mu_0)[s^2]^{-1}\sqrt{n}(\bar{x} - \mu_0)$$

$$= (\text{UVN variable}) \left( \frac{\text{scaled } \chi^2 \text{ variable}}{df} \right)^{-1} (\text{UVN variable})$$

where MVN (UVN) = multivariate (univariate) normal and $df = n - 1$. 
Define Hotelling’s $T^2$ Test Function

```
T.test <- function(X, mu=0){
  X <- as.matrix(X)
  n <- nrow(X)
  p <- ncol(X)
  df2 <- n - p
  if(df2 < 1L) stop("Need nrow(X) > ncol(X).")
  if(length(mu) != p) mu <- rep(mu[1], p)
  xbar <- colMeans(X)
  S <- cov(X)
  T2 <- n * t(xbar - mu) %*% solve(S) %*% (xbar - mu)
  Fstat <- T2 / (p * (n-1) / df2)
  pval <- 1 - pf(Fstat, df1=p, df2=df2)
  data.frame(T2=as.numeric(T2), Fstat=as.numeric(Fstat),
              df1=p, df2=df2, p.value=as.numeric(pval), row.names="")
}
```
Inferences about a Single Mean Vector

Hotelling’s $T^2$ Test Example in R

```r
# get data matrix
> data(mtcars)
> X <- mtcars[,c("mpg","disp","hp","wt")]
> xbar <- colMeans(X)

# try Hotelling T^2 function
> xbar
   mpg     disp      hp      wt
  20.09062 230.72188 146.68750  3.21725
> T.test(X)
   T2   Fstat df1 df2    p.value
  7608.14 1717.967   4  28 0
> T.test(X, mu=c(20,200,150,3))
   T2   Fstat df1 df2    p.value
  10.78587 2.435519   4  28 0.07058328
> T.test(X, mu=xbar)
   T2   Fstat df1 df2    p.value
    0     0    4  28    1
```
Inferences about a Single Mean Vector

Hotelling’s $T^2$ Using `lm` Function in R

```r
> y <- as.matrix(X)
> anova(lm(y ~ 1))
Analysis of Variance Table

   Df  Pillai  approx F num Df den Df Pr(>F)
(Intercept) 1 0.99594 1718     4  28 < 2.2e-16 ***
Residuals  31
---
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> y <- as.matrix(X) - matrix(c(20,200,150,3),nrow(X),ncol(X),byrow=T)
> anova(lm(y ~ 1))
Analysis of Variance Table

   Df  Pillai  approx F num Df den Df Pr(>F)
(Intercept) 1 0.25812  2.4355     4  28  0.07058 .
Residuals  31
---
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> y <- as.matrix(X) - matrix(xbar,nrow(X),ncol(X),byrow=T)
> anova(lm(y ~ 1))
Analysis of Variance Table

   Df  Pillai  approx F num Df den Df Pr(>F)
(Intercept) 1 1.8645e-31 1.3052e-30     4  28       1
Residuals  31
```
Reminder: the log-likelihood function for $n$ independent samples from a $p$-variate normal distribution has the form

$$LL(\mu, \Sigma | X) = -\frac{np}{2} \log(2\pi) - \frac{n}{2} \log(|\Sigma|) - \frac{1}{2} \sum_{i=1}^{n} (x_i - \mu)' \Sigma^{-1} (x_i - \mu)$$

and the MLEs of the mean vector $\mu$ and covariance matrix $\Sigma$ are

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^{n} x_i = \bar{x}$$

$$\hat{\Sigma} = \frac{1}{n} \sum_{i=1}^{n} (x_i - \hat{\mu})(x_i - \hat{\mu})'$$
Plugging the MLEs of $\mu$ and $\Sigma$ into the likelihood function gives

$$\max_{\mu, \Sigma} L(\mu, \Sigma \mid X) = \frac{\exp(-np/2)}{(2\pi)^{np/2} |\hat{\Sigma}| n/2}$$

If we assume that $\mu = \mu_0$ under $H_0$, then we have that

$$\max_{\Sigma} L(\Sigma \mid \mu_0, X) = \frac{\exp(-np/2)}{(2\pi)^{np/2} |\hat{\Sigma}_0| n/2}$$

where $\hat{\Sigma}_0 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu_0)(x_i - \mu_0)'$ is the MLE of $\Sigma$ under $H_0$. 
Likelihood Ratio Test Statistic (and Wilks’ lambda)

Want to test \( H_0 : \mu = \mu_0 \) versus \( H_1 : \mu \neq \mu_0 \)

The likelihood ratio test statistic is

\[
\Lambda = \frac{\max_{\Sigma} L(\Sigma | \mu_0, X)}{\max_{\mu, \Sigma} L(\mu, \Sigma | X)} = \left( \frac{|\hat{\Sigma}|}{|\hat{\Sigma}_0|} \right)^{n/2}
\]

and we reject \( H_0 \) if the observed value of \( \Lambda \) is too small.

The equivalent test statistic \( \Lambda^{2/n} = \frac{|\hat{\Sigma}|}{|\hat{\Sigma}_0|} \) is known as Wilks’ lambda.
There is a simple relationship between $T^2$ and $\Lambda$

$$\Lambda^{2/n} = \left(1 + \frac{T^2}{n-1}\right)^{-1}$$

which derives from the definition of the matrix determinant.\(^1\)

This implies that we can just use the $T^2$ distribution for $\Lambda$ inference.

- Reject $H_0$ for small $\Lambda^{2/n} \iff$ large $T^2$

---

\(^1\)For a proof, see p 218 of Johnson & Wichern (2007).
General Likelihood Ratio Tests

Let \( \theta \in \Theta \) denote a \( p \times 1 \) vector of parameters, which takes values in the parameter set \( \Theta \), and let \( \theta_0 \in \Theta_0 \) where \( \Theta_0 \subset \Theta \).

A likelihood ratio test rejects \( H_0 : \theta \in \Theta_0 \) in favor of \( H_1 : \theta \notin \Theta_0 \) if

\[
\Lambda = \frac{\max_{\theta \in \Theta_0} L(\theta)}{\max_{\theta \in \Theta} L(\theta)} < c_\alpha
\]

where \( c_\alpha \) is some constant and \( L(\cdot) \) is the likelihood function.

For a large sample size \( n \), we have that

\[
-2 \log(\Lambda) \approx \chi^2_{\nu - \nu_0}
\]

where \( \nu \) and \( \nu_0 \) are the dimensions of \( \Theta \) and \( \Theta_0 \).
Extending Confidence Intervals to Regions

A 100(1 − α)% confidence interval (CI) for \( \theta \in \Theta \) is defined such that

\[
P[L_\alpha(x) \leq \theta \leq U_\alpha(x)] = 1 - \alpha
\]

where the interval \([L_\alpha(x), U_\alpha(x)] \subset \Theta\) is a function of the data vector \(x\) and the significance level \(\alpha\).

A confidence region is a multivariate extension of a confidence interval.

A 100(1 − α)% confidence region (CR) for \( \theta \in \Theta \) is defined such that

\[
P[\theta \in R_\alpha(X)] = 1 - \alpha
\]

where the region \(R_\alpha(X) \subset \Theta\) is a function of the data matrix \(X\) and the significance level \(\alpha\).
Before we collect \( n \) samples from a \( p \)-variate normal distribution

\[
P[T^2 \leq \nu_{n,p}F_{p,n-p}(\alpha)] = 1 - \alpha
\]

where \( T^2 = n(\bar{x} - \mu)'S^{-1}(\bar{x} - \mu) \) and \( \nu_{n,p} = p(n - 1)/(n - p) \).

The 100\((1 - \alpha)\)% confidence region (CR) for a mean vector from a \( p \)-variate normal distribution is ellipsoid formed by all \( \mu \in \mathbb{R}^p \) such that

\[
n(\bar{x} - \mu)'S^{-1}(\bar{x} - \mu) \leq \nu_{n,p}F_{p,n-p}(\alpha)
\]

where \( \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \) and \( S = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})' \).
Inferences about a Single Mean Vector

Confidence Regions

Forming Confidence Regions in R

n <- nrow(X)
p <- ncol(X)
xbar <- colMeans(X)
S <- cov(X)
library(car)
tconst <- sqrt((p/n)*((n-1)/(n-p)) * qf(0.99,p,n-p))
id <- c(1,3)
plot(ellipse(center=xbar[id], shape=S[id,id], radius=tconst, draw=F), xlab="mpg", ylab="hp")

Nathaniel E. Helwig (U of Minnesota)
Linear Combinations of Normal Variables

Suppose that $\mathbf{X} = (X_1, \ldots, X_p) \sim \mathcal{N}(\mu, \Sigma)$ and let $\mathbf{a} = (a_1, \ldots, a_p) \in \mathbb{R}^p$ denote some linear transformation vector.

The random variable $Z = \sum_{j=1}^{p} a_j X_j = \mathbf{a}' \mathbf{X}$ has the properties

$$
\mu_Z = E(Z) = \sum_{j=1}^{p} a_j E(X_j) = \mathbf{a}' \mu
$$

$$
\sigma_Z^2 = \text{Var}(Z) = \mathbf{a}' \Sigma \mathbf{a}
$$

and because $Z$ is a linear transformation of normal variables, we know

$$
Z \sim \mathcal{N}(\mu_Z, \sigma_Z^2)
$$

i.e., $Z$ follows a univariate normal with mean $\mu_Z$ and variance $\sigma_Z^2$. 
Suppose that $\mathbf{x}_i \overset{iid}{\sim} \mathcal{N}(\mu, \Sigma)$ and $z_i = \mathbf{a}'\mathbf{x}_i$ for $i \in \{1, \ldots, n\}$.

The sample mean and variance of the $z_i$ terms are

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

$$
\hat{s}_z^2 = \mathbf{a}'\mathbf{S}\mathbf{a}
$$

where $\bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i$ and $\mathbf{S} = \frac{1}{n-1} \sum_{i=1}^{n} (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})'$.

Because $\bar{z}$ is a linear transformation of normal variables, we know

$$
\bar{z} \sim \mathcal{N}(\mu_Z, \sigma^2_Z/n)
$$

i.e., $\bar{z}$ follows a univariate normal with mean $\mu_Z$ and variance $\sigma^2_Z/n$. 

Nathaniel E. Helwig (U of Minnesota)
Inferences about a Single Mean Vector

Simultaneous $T^2$ Confidence Intervals

Confidence Intervals for Single Linear Combination

Fixing $a$ and assuming $\sigma_Z^2$ is unknown, the $t$ test statistic is

$$t = \frac{\bar{z} - \mu_Z}{s_z/\sqrt{n}} = \frac{\sqrt{n}(a'\bar{x} - a'\mu)}{\sqrt{a'Sa}}$$

and the corresponding confidence interval has the form

$$\bar{z} - \frac{s_z}{\sqrt{n}}t_{n-1}(\alpha/2) \leq \mu_Z \leq \bar{z} + \frac{s_z}{\sqrt{n}}t_{n-1}(\alpha/2)$$

$$a'\bar{x} - \frac{\sqrt{a'Sa}}{\sqrt{n}}t_{n-1}(\alpha/2) \leq \mu_Z \leq a'\bar{x} + \frac{\sqrt{a'Sa}}{\sqrt{n}}t_{n-1}(\alpha/2)$$

where $t_{n-1}(\alpha/2)$ is the critical value that cuts off the upper $(100\alpha/2)\%$ tail of the $t_{n-1}$ distribution.
The confidence interval has level $\alpha$ separately for each CI we form.

- Each interval separately satisfies $P[\mu_j \in R_j(X)] = 1 - \alpha$

For multiple CIs, the familywise significance level will exceed $\alpha$.

- Intervals do not satisfy $P[\mu_1 \in R_1(X) \cap \cdots \cap \mu_p \in R_p(X)] = 1 - \alpha$

If we want to form a CI for each $\mu_j$ term, we need to consider some approach that will control the familywise error rate.
Defining a Simultaneous Confidence Interval

For a fixed $a$, the $t$ test statistic CI is the set of $a'\mu$ values such that

$$ t^2 = \frac{(\bar{z} - \mu_Z)^2}{s_Z^2/n} = \frac{n(a'\bar{x} - a'\mu)^2}{a'Sa} \leq t_{n-1}(\alpha/2) $$

For a simultaneous CI, we want the above to hold for all choices of $a$.

Start by considering the maximum possible $t^2$ that we could see

$$ \max_a t^2 = \max_a \frac{n[a'(\bar{x} - \mu)]^2}{a'Sa} = n(\bar{x} - \mu)'S^{-1}(\bar{x} - \mu) = T^2 $$

which occurs when $a \propto S^{-1}(\bar{x} - \mu)$. 
The fact that \( \max_a t^2 = T^2 \) leads to the following simultaneous CI

\[
\bar{z} - \frac{S_z}{\sqrt{n}} \sqrt{\nu_{n,p}^F_{p,n-p}(\alpha)} \leq \mu \leq \bar{z} + \frac{S_z}{\sqrt{n}} \sqrt{\nu_{n,p}^F_{p,n-p}(\alpha)}
\]

\[
a' \bar{x} - \frac{\sqrt{a'Sa}}{\sqrt{n}} \sqrt{\nu_{n,p}^F_{p,n-p}(\alpha)} \leq \mu \leq a' \bar{x} + \frac{\sqrt{a'Sa}}{\sqrt{n}} \sqrt{\nu_{n,p}^F_{p,n-p}(\alpha)}
\]

which uses the fact that \( T^2 \sim \nu_{n,p}^F_{p,n-p} \) where \( \nu_{n,p} = p(n-1)/(n-p) \).

Simultaneously for all linear combination vectors \( a \in \mathbb{R}^p \), the above interval will contain \( a' \mu \) with probability \( 1 - \alpha \).
T.ci <- function(mu, Sigma, n, avec=rep(1,length(mu)), level=0.95) {
  p <- length(mu)
  if(nrow(Sigma) != p) stop("Need length(mu) == nrow(Sigma).")
  if(ncol(Sigma) != p) stop("Need length(mu) == ncol(Sigma).")
  if(length(avec) != p) stop("Need length(mu) == length(avec).")
  if(level <= 0 | level >= 1) stop("Need 0 < level < 1.")
  cval <- qf(level, p, n-p) * p * (n-1) / (n-p)
  zhat <- crossprod(avec, mu)
  zvar <- crossprod(avec, Sigma %*% avec) / n
  const <- sqrt(cval * zvar)
  c(lower = zhat - const, upper = zhat + const)
}
Example of Simultaneous $T^2$ Confidence Intervals

```r
> X <- mtcars[,c("mpg","disp","hp","wt")]
> n <- nrow(X)
> p <- ncol(X)
> xbar <- colMeans(X)
> S <- cov(X)
> xbar

  mpg   disp     hp    wt
20.09062 230.72188 146.68750 3.21725

> T.ci(mu=xbar, Sigma=S, n=n, avec=c(1,0,0,0))

  lower    upper
16.39689 23.78436

> T.ci(mu=xbar, Sigma=S, n=n, avec=c(0,1,0,0))

  lower    upper
154.7637  306.6801

> T.ci(mu=xbar, Sigma=S, n=n, avec=c(0,0,1,0))

  lower    upper
104.6674  188.7076

> T.ci(mu=xbar, Sigma=S, n=n, avec=c(0,0,0,1))

  lower    upper
 2.617584  3.816916
```
Inferences about a Single Mean Vector

Simultaneous $T^2$ Confidence Intervals

Compare Simultaneous $T^2$ CI to Classic $t$ CI

TCI <- tCI <- NULL
for(k in 1:4){
  avec <- rep(0, 4)
  avec[k] <- 1
  TCI <- c(TCI, T.ci(xbar, S, n, avec))
  tCI <- c(tCI,
          xbar[k] - sqrt(S[k,k]/n) * qt(0.975, df=n-1),
          xbar[k] + sqrt(S[k,k]/n) * qt(0.975, df=n-1))
}
rtab <- rbind(TCI, tCI)

> round(rtab, 2)

   mpg.lower  mpg.upper  disp.lower  disp.upper  hp.lower  hp.upper  wt.lower  wt.upper
  TCI     16.40    23.78     154.76     306.68  104.67    188.71     2.62     3.82
  tCI     17.92    22.26     186.04     275.41  121.97    171.41     2.86     3.57
If \( p \) and/or the number of linear combinations \( a_1, \ldots, a_q \) is small, we may be able to form better (i.e., narrower) simultaneous CIs.

Let \( C_k \) denote some confidence statement, and note that

\[
P[\text{all } C_k \text{ true}] = 1 - P[\text{at least one } C_k \text{ false}] \\
\geq 1 - \sum_{k=1}^{q} P(C_k \text{ false}) \\
= 1 - \sum_{k=1}^{q} \alpha_k
\]

where \( \alpha_k \) is the significance level for the \( k \)-th test.
Simultaneous CIs via Bonferroni’s Method

Result on the previous slide is a special case of Bonferroni’s inequality.

To control familywise error rate, we just need to adjust the error rates of the individual tests, i.e., the $\alpha_k$ terms.

If no prior knowledge is available,\(^2\) we simply set $\alpha_k = \alpha/q$ to control the familywise error rate at $\alpha$ when conducting $q$ significance tests.

\[100(1 - \alpha)\% \text{ CI for } q \text{ tests: } \bar{z} \pm (s_{z}/\sqrt{n})t_{n-1}(\alpha_k/2) \text{ with } \alpha_k = \alpha/q\]

\(^2\)If we have prior knowledge about the importance of the individual tests, we could adjust each $\alpha_k$ individually with the constraint that $\sum_{k=1}^{q} \alpha_k = \alpha$. 
Simultaneous CIs via Bonferroni’s Method in R

TCI <- tCI <- bon <- NULL
alpha <- 1 - 0.05/(2*4)
for(k in 1:4){
  avec <- rep(0, 4)
  avec[k] <- 1
  TCI <- c(TCI, T.ci(xbar, S, n, avec))
  tCI <- c(tCI,
           xbar[k] - sqrt(S[k,k]/n) * qt(0.975, df=n-1),
           xbar[k] + sqrt(S[k,k]/n) * qt(0.975, df=n-1))
  bon <- c(bon,
           xbar[k] - sqrt(S[k,k]/n) * qt(alpha, df=n-1),
           xbar[k] + sqrt(S[k,k]/n) * qt(alpha, df=n-1))
}
rtab <- rbind(TCI, tCI, bon)
> round(rtab, 2)

<table>
<thead>
<tr>
<th></th>
<th>mpg.lower</th>
<th>mpg.upper</th>
<th>disp.lower</th>
<th>disp.upper</th>
<th>hp.lower</th>
<th>hp.upper</th>
<th>wt.lower</th>
<th>wt.upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCI</td>
<td>16.40</td>
<td>23.78</td>
<td>154.76</td>
<td>306.68</td>
<td>104.67</td>
<td>188.71</td>
<td>2.62</td>
<td>3.82</td>
</tr>
<tr>
<td>tCI</td>
<td>17.92</td>
<td>22.26</td>
<td>186.04</td>
<td>275.41</td>
<td>121.97</td>
<td>171.41</td>
<td>2.86</td>
<td>3.57</td>
</tr>
<tr>
<td>bon</td>
<td>17.27</td>
<td>22.92</td>
<td>172.62</td>
<td>288.82</td>
<td>114.55</td>
<td>178.83</td>
<td>2.76</td>
<td>3.68</td>
</tr>
</tbody>
</table>
Suppose that \( \mathbf{x}_i = (x_{i1}, \ldots, x_{ip})' \) is a random sample from some distribution with finite mean \( \mu \) and finite covariance matrix \( \Sigma \).

As the sample size gets large, i.e., as \( n \to \infty \), we have that

\[
\sqrt{n}(\bar{\mathbf{x}} - \mu) \approx N(0, \Sigma)
\]

\[
n(\bar{\mathbf{x}} - \mu)'S^{-1}(\bar{\mathbf{x}} - \mu) \approx \chi^2_p
\]

where \( \approx \) denotes “is approximately distributed as”.

Result holds for non-normal data too, as long as \( n - p \) is large!
Update `T.test` Function: Add `asymp` Option

```r
T.test <- function(X, mu=0, asymp=FALSE){
  X <- as.matrix(X)
  n <- nrow(X)
  p <- ncol(X)
  df2 <- n - p
  if(df2 < 1L) stop("Need nrow(X) > ncol(X).")
  if(length(mu) != p) mu <- rep(mu[1], p)
  xbar <- colMeans(X)
  S <- cov(X)
  T2 <- n * t(xbar - mu) %*% solve(S) %*% (xbar - mu)
  Fstat <- T2 / (p * (n-1) / df2)
  if(asymp){
    pval <- 1 - pchisq(T2, df=p)
  } else {
    pval <- 1 - pf(Fstat, df1=p, df2=df2)
  }
  data.frame(T2=as.numeric(T2), Fstat=as.numeric(Fstat),
              df1=p, df2=df2, p.value=as.numeric(pval),
              asymp=asymp, row.names="")
}
```

Nathaniel E. Helwig (U of Minnesota)
Compared Finite Sample and Large Sample $p$-values

# compare finite sample and large sample p-values (n=10)
> set.seed(1)
> XX <- matrix(rnorm(10*4), 10, 4)
> T.test(XX)

```
  T2  Fstat df1 df2   p.value  asymp
1.963739 0.3272899  4   6 0.8503944 FALSE
```

> T.test(XX, asymp=TRUE)

```
  T2  Fstat df1 df2   p.value  asymp
1.963739 0.3272899  4   6 0.7424283 TRUE
```

# compare finite sample and large sample p-values (n=50)
> set.seed(1)
> XX <- matrix(rnorm(50*4), 50, 4)
> T.test(XX)

```
  T2  Fstat df1 df2   p.value  asymp
4.348226 1.020502  4  46 0.4067571 FALSE
```

> T.test(XX, asymp=TRUE)

```
  T2  Fstat df1 df2   p.value  asymp
4.348226 1.020502  4  46 0.3609251 TRUE
```

# compare finite sample and large sample p-values (n=100)
> set.seed(1)
> XX <- matrix(rnorm(100*4), 100, 4)
> T.test(XX)

```
  T2  Fstat df1 df2   p.value  asymp
1.972616 0.47821  4  96 0.7516411 FALSE
```

> T.test(XX, asymp=TRUE)

```
  T2  Fstat df1 df2   p.value  asymp
1.972616 0.47821  4  96 0.7407957 TRUE
```
For large $n$, we have that $T^2 \approx \chi_p^2$, which implies

$$P[T^2 \leq \chi_p^2(\alpha)] \approx 1 - \alpha$$

where $\chi_p^2(\alpha)$ is the upper $(100 \alpha)$th percentile of the $\chi_p^2$ distribution.

The implied large sample CI has the form

$$\bar{z} - \frac{s_z}{\sqrt{n}} \sqrt{\chi_p^2(\alpha)} \leq \mu_Z \leq \bar{z} + \frac{s_z}{\sqrt{n}} \sqrt{\chi_p^2(\alpha)}$$

$$a'\bar{x} - \frac{\sqrt{a'Sa}}{\sqrt{n}} \sqrt{\chi_p^2(\alpha)} \leq \mu_Z \leq a'\bar{x} + \frac{\sqrt{a'Sa}}{\sqrt{n}} \sqrt{\chi_p^2(\alpha)}$$
Forming Large Sample Confidence Intervals in R

```r
chi <- NULL
for(k in 1:4){
  chi <- c(chi,
          xbar[k] - sqrt(S[k,k]/n) * sqrt(qchisq(0.95, df=p)),
          xbar[k] + sqrt(S[k,k]/n) * sqrt(qchisq(0.95, df=p)))
}

> round(rtab, 2)
    mpg.lower mpg.upper disp.lower disp.upper
TCI  16.40  23.78  154.76   306.68
 tCI  17.92  22.26  186.04   275.41
bon  17.27  22.92  172.62   288.82
chi  16.81  23.37  163.24   298.21
    hp.lower hp.upper wt.lower wt.upper
TCI  104.67  188.71   2.62    3.82
 tCI  121.97  171.41   2.86    3.57
bon  114.55  178.83   2.76    3.68
chi  109.35  184.02   2.68    3.75
```
Prediction Regions for Future Observations

Suppose $x_i \overset{iid}{\sim} N(\mu, \Sigma)$ and $\bar{x}$ and $S$ have been calculated from a sample of $n$ independent observations.

If $x_*$ is some new observation sampled from $N(\mu, \Sigma)$, then

$$T^2_* = \frac{n}{n+1}(x_* - \bar{x})' S^{-1} (x_* - \bar{x}) \sim \frac{(n-1)p}{n-p} F_{p,n-p}$$

given that $\text{Cov}(x_* - \bar{x}) = \text{Cov}(x_*) + \text{Cov}(\bar{x}) = \left[(n+1)/n\right] \Sigma$.

The $100(1 - \alpha)\%$ prediction ellipsoid is given by all $x_*$ that satisfy

$$(x_* - \bar{x})' S^{-1} (x_* - \bar{x}) \leq \frac{(n^2 - 1)p}{n(n - p)} F_{p,n-p}(\alpha)$$
Forming Prediction Regions in R

```r
n <- nrow(X)
p <- ncol(X)
xbar <- colMeans(X)
S <- cov(X)
library(car)
pconst <- sqrt((p/n)*((n^2-1)/(n-p)) * qf(0.99,p,n-p))
id <- c(1,3)
plot(ellipse(center=xbar[id], shape=S[id,id], radius=pconst, draw=F), xlab="mpg", ylab="hp")
```
Inferences about Multiple Mean Vectors
Univariate Reminder: Student’s Two-Sample $t$ Tests

Remember: there are two types of two-sample $t$ tests:
- **Dependent samples:** two repeated measures from same subject
- **Independent samples:** measurements from two different groups

For the dependent samples $t$ test, we test

$$H_0 : \mu_d = \mu_0 \quad \text{versus} \quad H_1 : \mu_d \neq \mu_0$$

where $\mu_d = E(d_i)$ with $d_i = x_{i1} - x_{i2}$ denoting a difference score.

For the independent samples $t$ test, we test

$$H_0 : \mu_x - \mu_y = \mu_0 \quad \text{versus} \quad H_1 : \mu_x - \mu_y \neq \mu_0$$

where $\mu_x = E(x_i)$ and $\mu_y = E(y_i)$ are the two population means.
Multivariate Extensions of Two-Sample $t$ Tests

In this section, we will consider multivariate extensions of the (univariate) two-sample $t$ tests.

Similar to the univariate case, the multivariate dependent samples $T^2$ test performs the one-sample test on a difference score.

The independent samples case involves a modification of the $T^2$ statistic, and can be extended to $K > 2$ samples.

- $K > 2$ is a multivariate analysis of variance (MANOVA) model
To test the hypotheses $H_0 : \mu_d = \mu_0$ versus $H_1 : \mu_d \neq \mu_0$ we use

$$t = \frac{\bar{d} - \mu_0}{s_d / \sqrt{n}}$$

where $\bar{d} = \frac{1}{n} \sum_{i=1}^{n} d_i$ and $s_d^2 = \frac{1}{n-1} \sum_{i=1}^{n} (d_i - \bar{d})^2$ with $d_i = x_{i1} - x_{i2}$.

The $100(1 - \alpha)\%$ CI for the population mean of the difference score is

$$\bar{d} - t_{n-1}(\alpha/2) \frac{s_d}{\sqrt{n}} \leq \mu_d \leq \bar{d} + t_{n-1}(\alpha/2) \frac{s_d}{\sqrt{n}}$$
Let $x_{ijk}$ denote the $k$-th repeated measurement of the $j$-th variable collected from the $i$-th subject, and define $x_{ki} = (x_{i1k}, \ldots, x_{ipk})'$. The $i$-th subject’s vector of difference scores is defined as

$$d_i = x_{1i} - x_{2i} = \begin{pmatrix} x_{i11} - x_{i12} \\ x_{i21} - x_{i22} \\ \vdots \\ x_{ip1} - x_{ip2} \end{pmatrix}$$

and note that $d_i \overset{iid}{\sim} N(\mu_d, \Sigma_d)$ assuming the subjects are independent.
Inferences about Multiple Mean Vectors

Paired Comparisons

Hotelling’s $T^2$ for Difference Score Vectors

Given that $d_i \overset{iid}{\sim} N(\mu_d, \Sigma_d)$, the $T^2$ statistic has the form

$$T^2_d = n(\bar{d} - \mu_d)'S_d^{-1}(\bar{d} - \mu_d) \sim \frac{(n-1)p}{n-p} F_{p,n-p}$$

where $\bar{d} = \frac{1}{n} \sum_{i=1}^{n} d_i$ and $S_d = \frac{1}{n-1} \sum_{i=1}^{n} (d_i - \bar{d})(d_i - \bar{d})'$

We use the same inference procedures as before:

- Reject $H_0 : \mu_d = \mu_0$ if $(\bar{d} - \mu_0)'S_d^{-1}(\bar{d} - \mu_0) > \frac{(n-1)p}{n(n-p)} F_{p,n-p}(\alpha)$
- Use same procedures for forming confidence regions/intervals
In a repeated measures design units participate in $q > 2$ treatments.

- Assuming a single response variable at $q$ treatments
- $X = \{x_{ij}\}$ is the $n$ units $\times$ $q$ treatments data matrix

We can assume that $x_i \overset{iid}{\sim} N(\mu, \Sigma)$ where

- $\mu = (\mu_1, \ldots, \mu_q)$ is the mean vector
- $\Sigma$ is the $q \times q$ covariance matrix

Could use Hotelling’s $T^2$, but we will consider a new parameterization.
We could consider making contrasts of the component means such as

\[
\begin{pmatrix}
\mu_1 - \mu_2 \\
\mu_1 - \mu_3 \\
\vdots \\
\mu_1 - \mu_q
\end{pmatrix}
= \begin{pmatrix}
1 & -1 & 0 & \cdots & 0 \\
1 & 0 & -1 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
1 & 0 & 0 & \cdots & -1
\end{pmatrix}
\begin{pmatrix}
\mu_1 \\
\mu_2 \\
\vdots \\
\mu_q
\end{pmatrix}
= C_1\mu
\]

or

\[
\begin{pmatrix}
\mu_2 - \mu_1 \\
\mu_3 - \mu_2 \\
\vdots \\
\mu_q - \mu_{q-1}
\end{pmatrix}
= \begin{pmatrix}
-1 & 1 & 0 & \cdots & 0 & 0 \\
0 & -1 & 1 & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & -1 & 1
\end{pmatrix}
\begin{pmatrix}
\mu_1 \\
\mu_2 \\
\vdots \\
\mu_q
\end{pmatrix}
= C_2\mu
\]

which allows us to compare select mean differences.
Hotelling’s $T^2$ with Contrast Matrices

Note that a contrast matrix is any matrix $C_j$ that

1. Has linearly independent rows
2. Satisfies $C_j1_q = 0$ (i.e., rows sum to 0)

If $\mu \propto 1_q$ (i.e., $\mu_1 = \cdots = \mu_q$), then $C_j\mu = 0$ for any contrast matrix $C_j$.

We can use $T^2$ to test $H_0 : \mu \propto 1_q$ versus $H_1 : \mu \propto 1_q$

$$T^2 = n(C\bar{x})'(CSC')^{-1}C\bar{x} \sim \frac{(n-1)(q-1)}{n-q+1} F_{q-1,n-q+1}(\alpha)$$

where $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$ and $S = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})'$
Simultaneous $T^2$ CRs and CIs with Contrast Matrices

Given $C$, a $100(1 - \alpha)\%$ confidence region (CR) for $C\mu$ is defined as the set of all $C\mu$ that satisfy

$$n(C\bar{x} - C\mu)'(CSC')^{-1}(C\bar{x} - C\mu) \leq \frac{(n - 1)(q - 1)}{n - q + 1} F_{q-1,n-q+1}$$

This implies that a simultaneous $100(1 - \alpha)\%$ confidence interval (CI) for a single contrast $c'\mu$ has the form

$$c'\bar{x} \pm \sqrt{\frac{c'Sc}{n}} \sqrt{\frac{(n - 1)(q - 1)}{n - q + 1}} F_{q-1,n-q+1}(\alpha)$$
R Function for $T^2$ with Contrast Matrices

```r
RM.test <- function(X, mu=0, C=NULL){
  X <- as.matrix(X)
  n <- nrow(X)
  p <- ncol(X)
  df2 <- n - p + 1
  if(df2 < 1L) stop("Need nrow(X) > ncol(X).")
  if(length(mu) != p) mu <- rep(mu[1], p)
  xbar <- colMeans(X)
  S <- cov(X)
  if(is.null(C)){
    C <- matrix(0, p-1, p)
    for(k in 1:(p-1)) C[k, 1:2 + 1*(k-1)] <- c(1, -1)
  } else {
    if(nrow(C) != (p-1)) stop("Need [ncol(X)-1] == nrow(C).")
    if(ncol(C) != p) stop("Need ncol(X) == ncol(C).")
    if(any(rowSums(C)>0L)) stop("Need rowSums(C) == rep(0, nrow(C)).")
  }
  T2 <- n * t(C %*% (xbar - mu)) %*% solve(C %*% S %*% t(C)) %*% (C %*% (xbar - mu))
  Fstat <- T2 / ((p-1) * (n-1) / df2)
  pval <- 1 - pf(Fstat, df1=p-1, df2=df2)
  data.frame(T2=as.numeric(T2), Fstat=as.numeric(Fstat),
             df1=p-1, df2=df2, p.value=as.numeric(pval), row.names="")
}
```
Example of $T^2$ with Contrast Matrices in R: $H_0$ True

```r
# RM.test example w/ H0 true (10 data points)
> set.seed(1)
> XX <- matrix(rnorm(10*4), 10, 4)
> RM.test(XX)
   T2  Fstat df1 df2 p.value
1 1.83 0.48  3  7  0.7094449

# RM.test example w/ H0 true (100 data points)
> set.seed(1)
> XX <- matrix(rnorm(100*4), 100, 4)
> RM.test(XX)
   T2  Fstat df1 df2 p.value
1 1.29 0.42  3  97  0.7389465

# RM.test example w/ H0 true (500 data points)
> set.seed(1)
> XX <- matrix(rnorm(500*4), 500, 4)
> RM.test(XX)
   T2  Fstat df1 df2 p.value
1 1.23 0.41  3 497  0.7466049
```
Example of $T^2$ with Contrast Matrices in R: $H_0$ False

```r
# RM.test example w/ H0 false (10 data points)
> set.seed(1)
> XX <- matrix(rnorm(10*4), 10, 4)
> XX <- XX + matrix(c(0,0,0,0.25), 10, 4, byrow=TRUE)
> RM.test(XX)

   T2  Fstat df1 df2  p.value
 2.975373 0.7713929 3 7 0.5456821

# RM.test example w/ H0 false (100 data points)
> set.seed(1)
> XX <- matrix(rnorm(100*4), 100, 4)
> XX <- XX + matrix(c(0,0,0,0.25), 100, 4, byrow=TRUE)
> RM.test(XX)

   T2  Fstat df1 df2  p.value
 6.868081 2.243111 3 97 0.08812235

# RM.test example w/ H0 false (500 data points)
> set.seed(1)
> XX <- matrix(rnorm(500*4), 500, 4)
> XX <- XX + matrix(c(0,0,0,0.25), 500, 4, byrow=TRUE)
> RM.test(XX)

   T2  Fstat df1 df2  p.value
19.72918 6.550036 3 497 0.0002380274
```
Univariate Reminder: Independent Samples $t$ test

To test $H_0 : \mu_1 - \mu_2 = \mu_0$ versus $H_1 : \mu_1 - \mu_2 \neq \mu_0$ we use

$$t = \frac{\bar{x}_1 - \bar{x}_2 - \mu_0}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

where

- $\bar{x}_k = \frac{1}{n_k} \sum_{i=1}^{n_k} x_{ik}$ is the $k$-th group’s sample mean
- $s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2}$ is the pooled variance estimate
- $s_k^2 = \frac{1}{n_k-1} \sum_{i=1}^{n_k} (x_{ik} - \bar{x}_k)^2$ is the $k$-th group’s sample variance

The $100(1 - \alpha)$% CI for the difference in population means is

$$(\bar{x}_1 - \bar{x}_2) - t_{n_1 + n_2 - 2}(\alpha/2)s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \leq \mu_1 - \mu_2 \leq (\bar{x}_1 - \bar{x}_2) + t_{n_1 + n_2 - 2}(\alpha/2)s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$
Multivariate Independent Samples $T^2$ Test

Let $x_{ki} \overset{iid}{\sim} N(\mu_k, \Sigma)$ for $k \in \{1, 2\}$ and assume that the elements of $\{x_{1i}\}_{i=1}^{n_1}$ and $\{x_{2i}\}_{i=1}^{n_2}$ are independent of one another.

The pooled estimate of the covariance matrix has the form

$$S_p = \frac{n_1 - 1}{n_1 + n_2 - 2} S_1 + \frac{n_2 - 1}{n_1 + n_2 - 2} S_2$$

where $\bar{x}_k = \frac{1}{n_k} \sum_{i=1}^{n_k} x_{ki}$ and $S_k = \frac{1}{n_k - 1} \sum_{i=1}^{n_k} (x_{ki} - \bar{x}_k)(x_{ki} - \bar{x}_k)'$.

The $T^2$ test statistic for testing $H_0 : \mu_1 - \mu_2 = \mu_0$ has the form

$$T^2 = (\bar{x}_1 - \bar{x}_2 - \mu_0)' \left[ \left( \frac{1}{n_1} + \frac{1}{n_2} \right) S_p \right]^{-1} (\bar{x}_1 - \bar{x}_2 - \mu_0)$$
Assuming that $x_{ki} \overset{iid}{\sim} N(\mu_k, \Sigma)$ for $k \in \{1, 2\}$, we have that

$$T^2 = (\bar{x}_1 - \bar{x}_2 - [\mu_1 - \mu_2])' \left[ \left( \frac{1}{n_1} + \frac{1}{n_2} \right) S_p \right]^{-1} (\bar{x}_1 - \bar{x}_2 - [\mu_1 - \mu_2])$$

$$\sim \frac{(n_1 + n_2 - 2)p}{n_1 + n_2 - p - 1} F_{p, n_1 + n_2 - p - 1}$$

which is an analogue of the one sample $T^2$ statistic.

This implies that a $100(1 - \alpha)\%$ confidence region can be formed from

$$P \left[ T^2 \leq \frac{(n_1 + n_2 - 2)p}{n_1 + n_2 - p - 1} F_{p, n_1 + n_2 - p - 1}(\alpha) \right] = 1 - \alpha$$
Multivariate Independent Samples $T^2$ Test R Function

```r
T.test <- function(X, Y=NULL, mu=0, paired=FALSE, asymp=FALSE){
  if(is.null(Y)){
    # one-sample T^2 test: same code as before (omitted here)
  } else {
    if(paired){
      # dependent two-sample T^2 test
      X <- as.matrix(X)
      Y <- as.matrix(Y)
      if(!identical(dim(X),dim(Y))) stop("Need dim(X) == dim(Y).")
      xx <- T.test(X-Y, mu=mu, asymp=asymp)
      xx$type <- "dep-sample"
      return(xx)
    } else {
      # independent two-sample T^2 test
      X <- as.matrix(X)
      Y <- as.matrix(Y)
      nx <- nrow(X)
      ny <- nrow(Y)
      p <- ncol(X)
      df2 <- nx + ny - p - 1
      if(p != ncol(Y)) stop("Need ncol(X) == ncol(Y).")
      if(min(nx,ny) <= p) stop("Need min(nrow(X),nrow(Y)) > ncol(X).")
      Sp <- ((nx-1)*cov(X) + (ny-1)*cov(Y)) / (nx + ny - 2)
      dbar <- colMeans(X) - colMeans(Y)
      T2 <- (1/((1/nx) + (1/ny))) * t(dbar - mu) %*% solve(Sp) %*% (dbar - mu)
      Fstat <- T2 / ((nx + ny - 2) * p / df2)
      if(asymp){
        pval <- 1 - pchisq(T2, df=p)
      } else {
        pval <- 1 - pf(Fstat, df1=p, df2=df2)
      }
      return(data.frame(T2=as.numeric(T2), Fstat=as.numeric(Fstat),
                        df1=p, df2=df2, p.value=as.numeric(pval),
                        type="ind-sample", asymp=asymp, row.names=""))
    } # end if(paired)
  } # end if(is.null(Y))
} # end T.test function
```
Multivariate Independent Samples $T^2$ Test in R

```r
> X4 <- subset(mtcars, cyl==4)[,c("mpg","disp","hp","wt")]
> X6 <- subset(mtcars, cyl==6)[,c("mpg","disp","hp","wt")]
> X8 <- subset(mtcars, cyl==8)[,c("mpg","disp","hp","wt")]
> T.test(X4, X6)

    T2  Fstat df1 df2 p.value  type       asymp
 39.61 8.05  4  13 0.0017 ind-sample FALSE

> T.test(X4, X8)

    T2  Fstat df1 df2 p.value  type       asymp
185.61 40.35  4  20 2.63e-09 ind-sample FALSE

> T.test(X6, X8)

    T2  Fstat df1 df2 p.value  type       asymp
 61.97 13.05  4  16 6.55e-05 ind-sample FALSE
```
Comparing Finite Sample and Large Sample p-values

<table>
<thead>
<tr>
<th></th>
<th>T2</th>
<th>Fstat</th>
<th>df1</th>
<th>df2</th>
<th>p.value</th>
<th>type</th>
<th>asymp</th>
</tr>
</thead>
<tbody>
<tr>
<td>Finite Sample</td>
<td>3.582189</td>
<td>0.7462893</td>
<td>4</td>
<td>15</td>
<td>0.5754577</td>
<td>ind-sample</td>
<td>FALSE</td>
</tr>
<tr>
<td>Large Sample</td>
<td>3.582189</td>
<td>0.7462893</td>
<td>4</td>
<td>95</td>
<td>0.5110603</td>
<td>ind-sample</td>
<td>TRUE</td>
</tr>
</tbody>
</table>

# compare finite sample and large sample p-values (n=100)

<table>
<thead>
<tr>
<th></th>
<th>T2</th>
<th>Fstat</th>
<th>df1</th>
<th>df2</th>
<th>p.value</th>
<th>type</th>
<th>asymp</th>
</tr>
</thead>
<tbody>
<tr>
<td>Finite Sample</td>
<td>3.270955</td>
<td>0.8053489</td>
<td>4</td>
<td>195</td>
<td>0.5230921</td>
<td>ind-sample</td>
<td>FALSE</td>
</tr>
<tr>
<td>Large Sample</td>
<td>3.270955</td>
<td>0.8053489</td>
<td>4</td>
<td>195</td>
<td>0.5135473</td>
<td>ind-sample</td>
<td>TRUE</td>
</tr>
</tbody>
</table>

Nathaniel E. Helwig (U of Minnesota)
A $100(1 - \alpha)\%$ confidence interval of the form

$$a'(\bar{x}_1 - \bar{x}_2) \pm \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} a'S_p a \sqrt{\frac{(n_1 + n_2 - 2)p}{n_1 + n_2 - p - 1} F_{p, n_1+n_2-p-1}(\alpha)}$$

covers $a'(\mu_1 - \mu_2)$ with probability $1 - \alpha$ simultaneously for all $a$.

With $a = e_j$ (the $j$-th standard basis vector), we have

$$(\bar{x}_{1j} - \bar{x}_{2j}) \pm \sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)} s_{jj}(p) \sqrt{\frac{(n_1 + n_2 - 2)p}{n_1 + n_2 - p - 1} F_{p, n_1+n_2-p-1}(\alpha)}$$

where $s_{jj}(p)$ is the $j$-th diagonal element of $S_p$. 
Simultaneous CIs for Independent Samples $T^2$ in R

```r
T.ci <- function(mu, Sigma, n, avec=rep(1,length(mu)), level=0.95){
  p <- length(mu)
  if(nrow(Sigma)!=p) stop("Need length(mu) == nrow(Sigma).")
  if(ncol(Sigma)!=p) stop("Need length(mu) == ncol(Sigma).")
  if(length(avec)!=p) stop("Need length(mu) == length(avec).")
  if(level <=0 | level >= 1) stop("Need 0 < level < 1.")
  zhat <- crossprod(avec, mu)
  if(length(n)==1L){
    cval <- qf(level, p, n-p) * p * (n-1) / (n-p)
    zvar <- crossprod(avec, Sigma %*% avec) / n
  } else {
    cval <- qf(level, p, df2) * p * (n[1]+n[2]-2) / df2
    zvar <- crossprod(avec, Sigma %*% avec) * ( (1/n[1]) + (1/n[2]) )
  }
  const <- sqrt(cval * zvar)
  c(lower = zhat - const, upper = zhat + const)
}
```

Nathaniel E. Helwig (U of Minnesota)
Example of Simultaneous CIs for Indep. Samples $T^2$

```r
X4 <- subset(mtcars, cyl==4)[,c("mpg","disp","hp","wt")]
X6 <- subset(mtcars, cyl==6)[,c("mpg","disp","hp","wt")]
n4 <- nrow(X4)
n6 <- nrow(X6)
dbar <- colMeans(X4) - colMeans(X6)
Sp <- ((n4-1)*cov(X4) + (n6-1)*cov(X6)) / (n4 + n6 - 2)
dbar

   mpg    disp     hp      wt
6.9207792 -78.1779221 -39.6493506 -0.8314156

> T.ci(dbar, Sp, c(n4,n6), c(1,0,0,0))
lower    upper
-0.1082001 13.9497585

> T.ci(dbar, Sp, c(n4,n6), c(0,1,0,0))
lower    upper
-141.59106 -14.76478

> T.ci(dbar, Sp, c(n4,n6), c(0,0,1,0))
lower    upper
-82.189575 2.890874

> T.ci(dbar, Sp, c(n4,n6), c(0,0,0,1))
lower    upper
-1.7885075 0.1256764
```
Let $x_{ki} \overset{iid}{\sim} N(\mu_k, \Sigma_k)$ for $k \in \{1, 2\}$ and assume that the elements of \{\(x_{1i}\)\}_{i=1}^{n_1}$ and \(\{x_{2i}\}_{i=1}^{n_2}$ are independent of one another.

We cannot define a "distance" measure like $T^2$, whose distribution does not depend on the unknown population parameters $\Sigma_1$ and $\Sigma_2$.

Use the modified $T^2$ statistic with non-pooled covariance matrices

$$T^2 = (\bar{x}_1 - \bar{x}_2 - [\mu_1 - \mu_2])' \left[n_1^{-1}S_1 + n_2^{-1}S_2\right]^{-1} (\bar{x}_1 - \bar{x}_2 - [\mu_1 - \mu_2])$$

where $\bar{x}_k = \frac{1}{n_k} \sum_{i=1}^{n_k} x_{ki}$ and $S_k = \frac{1}{n_k - 1} \sum_{i=1}^{n_k} (x_{ki} - \bar{x}_k)(x_{ki} - \bar{x}_k)'$. 
Inferences about Multiple Mean Vectors

Large and Small Sample $T^2$ Inferences when $\Sigma_1 \neq \Sigma_2$

If $\min(n_1, n_2) - p$ is large, we can use the large sample approximation:

$$P[T^2 \leq \chi^2_p(\alpha)] \approx 1 - \alpha$$

which (asymptotically) works for non-normal multivariate data too!

If $\min(n_1, n_2) - p$ is small and we assume normality, we can use

$$T^2 \approx \frac{\nu p}{\nu - p + 1} F_{p, \nu - p + 1}$$

where the degrees of freedom parameter $\nu$ is estimated as

$$\nu = \frac{p + p^2}{\sum_{k=1}^{2} \frac{1}{n_k} \left\{ \text{tr} \left[ \left( \frac{1}{n_k} S_k S_0^{-1} \right)^2 \right] + \left( \text{tr} \left[ \frac{1}{n_k} S_k S_0^{-1} \right] \right)^2 \right\}}$$

with $S_0 = \frac{1}{n_1} S_1 + \frac{1}{n_2} S_2$. Note that $\min(n_1, n_2) \leq \nu \leq n_1 + n_2$. 
Add `var.equal` Option to $T^2$ Test R Function

```r
T.test <- function(X, Y=NULL, mu=0, paired=FALSE, asymp=FALSE, var.equal=TRUE){
  if(is.null(Y)){# one-sample $T^2$ test: same code as before (omitted here)
  } else {
    if(paired){# dependent two-sample $T^2$ test: same code as before (omitted here)
    } else {
      # independent two-sample $T^2$ test
      X <- as.matrix(X)
      Y <- as.matrix(Y)
      nx <- nrow(X)
      ny <- nrow(Y)
      p <- ncol(X)
      if(p != ncol(Y)) stop("Need ncol(X) == ncol(Y).")
      if(min(nx,ny) <= p) stop("Need min(nrow(X),nrow(Y)) > ncol(X).")
      dbar <- colMeans(X) - colMeans(Y)
      if(var.equal){
        df2 <- nx + ny - p - 1
        Sp <- ((nx-1)*cov(X) + (ny-1)*cov(Y)) / (nx + ny - 2)
        T2 <- (1/((1/nx) + (1/ny))) * t(dbar - mu) %*% solve(Sp) %*% (dbar - mu)
        Fstat <- T2 / ((nx + ny - 2) * p / df2)
      } else {
        Sx <- cov(X)
        Sy <- cov(Y)
        Sp <- (Sx/nx) + (Sy/ny)
        T2 <- t(dbar - mu) %*% solve(Sp) %*% (dbar - mu)
        SpInv <- solve(Sp)
        SxSpInv <- (1/nx) * Sx %*% SpInv
        SySpInv <- (1/ny) * Sy %*% SpInv
        nudx <- sum(diag(SxSpInv))^2 / nx
        nudy <- sum(diag(SySpInv))^2 / ny
        nu <- (p + p^2) / (nudx + nudy)
        df2 <- nu - p + 1
        Fstat <- T2 / (nu + p / df2)
      }
      if(asymp){
        pval <- 1 - pchisq(T2, df=p)
      } else {
        pval <- 1 - pf(Fstat, df1=p, df2=df2)
      }
      return(data.frame(T2=as.numeric(T2), Fstat=as.numeric(Fstat),
          df1=p, df2=df2, p.value=as.numeric(pval),
          type="ind-sample", asymp=asymp, var.equal=var.equal, row.names=""))
    } # end if(paired)
  } # end if(is.null(Y))
} # end T.test function
```

Nathaniel E. Helwig (U of Minnesota)
Example with `var.equal` Option

```r
> X4 <- subset(mtcars, cyl==4)[,c("mpg","disp","hp","wt")]
> X6 <- subset(mtcars, cyl==6)[,c("mpg","disp","hp","wt")]
> T.test(X4, X6)
    T2  Fstat  df1  df2     p.value  type     asymp var.equal
39.60993  8.045767  4  13 0.001713109 ind-sample FALSE      TRUE
> T.test(X4, X6, var.equal=FALSE)
    T2  Fstat  df1  df2     p.value  type     asymp var.equal
46.04706  9.266334  4 12.38026 0.001067989 ind-sample FALSE     FALSE

> set.seed(1)
> n <- 100
> XX <- matrix(rnorm(n*4), n, 4)
> YY <- matrix(rnorm(n*4), n, 4)
> T.test(XX,YY)
    T2  Fstat  df1  df2     p.value  type     asymp var.equal
 3.270955  0.8053489  4 195 0.5230921 ind-sample FALSE      TRUE
> T.test(XX,YY,var.equal=F)
    T2  Fstat  df1  df2     p.value  type     asymp var.equal
 3.270955  0.8053198  4 194.537 0.5231144 ind-sample FALSE     FALSE
```
Suppose that $x_{ki} \overset{\text{ind}}{\sim} N(\mu_k, \sigma^2)$ for $k \in \{1, \ldots, g\}$ and $i \in \{1, \ldots, n_k\}$.

The one-way analysis of variance (ANOVA) model has the form

$$x_{ki} = \mu + \alpha_k + \epsilon_{ki}$$

where $\mu$ is the overall mean, $\alpha_k$ is the $k$-th group’s treatment effect with the constraint that $\sum_{k=1}^{g} n_k \alpha_k = 0$, and $\epsilon_{ki} \overset{\text{iid}}{\sim} N(0, \sigma^2)$ are error terms.

The sample estimates of the model parameters are

$$\hat{\mu} = \bar{x} \quad \text{and} \quad \hat{\alpha}_k = \bar{x}_k - \bar{x} \quad \text{and} \quad \hat{\epsilon}_{ki} = x_{ki} - \bar{x}_k$$

where $\bar{x} = \frac{1}{\sum_{k=1}^{g} n_k} \sum_{k=1}^{g} n_k \bar{x}_k$ and $\bar{x}_k = \frac{1}{n_k} \sum_{i=1}^{n_k} x_{ki}$.
Univariate Reminder: One-Way ANOVA (continued)

We want to test the hypotheses $H_0 : \alpha_k = 0$ for all $k \in \{1, \ldots, g\}$ versus $H_1 : \alpha_k \neq 0$ for some $k \in \{1, \ldots, g\}$.

The decomposition of the sums-of-squares has the form

$$\sum_{k=1}^{g} \sum_{i=1}^{n_k} (x_{ki} - \bar{x})^2 = \sum_{k=1}^{g} n_k (\bar{x}_k - \bar{x})^2 + \sum_{k=1}^{g} \sum_{i=1}^{n_k} (x_{ki} - \bar{x}_k)^2$$

The ANOVA $F$ test rejects $H_0$ at level $\alpha$ if

$$F = \frac{SSB/(g-1)}{SSW/(n-g)} > F_{g-1,n-g}(\alpha)$$

where $n = \sum_{k=1}^{g} n_k$. 
Multivariate Extension of One-Way ANOVA

Let $\mathbf{x}_{ki} \overset{iid}{\sim} \mathcal{N}(\mu_k, \Sigma)$ for $k \in \{1, \ldots, g\}$ and assume that the elements of $\{\mathbf{x}_{ki}\}_{i=1}^{n_k}$ and $\{\mathbf{x}_{li}\}_{i=1}^{n_l}$ are independent of one another.

The one-way multivariate analysis of variance (MANOVA) has the form

$$\mathbf{x}_{ki} = \mu + \alpha_k + \epsilon_{ki}$$

where $\mu_k = \mu + \alpha_k$, $\mu$ is the overall mean vector, $\alpha_k$ is the $k$-th group’s treatment effect vector (with $\sum_{k=1}^{g} n_k \alpha_k = \mathbf{0}_p$), and $\epsilon_{ki} \overset{iid}{\sim} \mathcal{N}(\mathbf{0}_p, \Sigma)$.

The sample estimates of the model parameters are

$$\hat{\mu} = \bar{x} \quad \text{and} \quad \hat{\alpha}_k = \bar{x}_k - \bar{x} \quad \text{and} \quad \hat{\epsilon}_{ki} = \mathbf{x}_{ki} - \bar{x}_k$$

where $\bar{x} = \frac{1}{\sum_{k=1}^{g} n_k} \sum_{k=1}^{g} n_k \bar{x}_k$ and $\bar{x}_k = \frac{1}{n_k} \sum_{i=1}^{n_k} \mathbf{x}_{ki}$. 
The MANOVA sums-of-squares and crossproducts decomposition is

\[
\sum_{k=1}^{g} \sum_{i=1}^{n_k} (x_{ki} - \bar{x})(x_{ki} - \bar{x})' = \sum_{k=1}^{g} n_k (\bar{x}_k - \bar{x})(\bar{x}_k - \bar{x})' + \sum_{k=1}^{g} \sum_{i=1}^{n_k} (x_{ki} - \bar{x}_k)(x_{ki} - \bar{x}_k)'
\]

SSCP Total

SSCP Between

SSCP Within

and note that the within SSCP matrix has the form

\[
W = \sum_{k=1}^{g} \sum_{i=1}^{n_k} (x_{ki} - \bar{x}_k)(x_{ki} - \bar{x}_k)'
\]

\[
= \sum_{k=1}^{g} (n_k - 1) S_k
\]

where \( S_k = \frac{1}{n_k - 1} \sum_{i=1}^{n_k} (x_{ki} - \bar{x}_k)(x_{ki} - \bar{x}_k)' \) is the \( k \)-th group’s sample covariance matrix.
Similar to the one-way ANOVA, we can summarize the SSCP information in a table

<table>
<thead>
<tr>
<th>Source</th>
<th>SSCP Matrix</th>
<th>D.F.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between</td>
<td>$B = \sum_{k=1}^{g} n_k (\bar{x}_k - \bar{x})(\bar{x}_k - \bar{x})'$</td>
<td>$g - 1$</td>
</tr>
<tr>
<td>Within</td>
<td>$W = \sum_{k=1}^{g} \sum_{i=1}^{n_k} (x_{ki} - \bar{x}<em>k)(x</em>{ki} - \bar{x}_k)'$</td>
<td>$\sum_{k=1}^{g} n_k - g$</td>
</tr>
<tr>
<td>Total</td>
<td>$B + W = \sum_{k=1}^{g} \sum_{i=1}^{n_k} (x_{ki} - \bar{x})(x_{ki} - \bar{x})'$</td>
<td>$\sum_{k=1}^{g} n_k - 1$</td>
</tr>
</tbody>
</table>

Between = treatment sum-of-squares and crossproducts
Within = residual (error) sum-of-squares and crossproducts
We want to test the hypotheses $H_0 : \alpha_k = 0_p$ for all $k \in \{1, \ldots, g\}$ versus $H_1 : \alpha_k \neq 0_p$ for some $k \in \{1, \ldots, g\}$.

The MANOVA test statistic has the form

$$\Lambda^* = \frac{|W|}{|B + W|}$$

which is known as Wilks’ lambda.

Reject $H_0$ if $\Lambda^*$ is smaller than expected under $H_0$. 
Distribution for Wilks’ Lambda

For certain special cases, the exact distribution of $\Lambda^*$ is known

<table>
<thead>
<tr>
<th>$p$</th>
<th>$g$</th>
<th>Sampling Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>$p = 1$</td>
<td>$g \geq 2$</td>
<td>$\left( \frac{n-g}{g-1} \right) \left( \frac{1-\Lambda^<em>}{\Lambda^</em>} \right) \sim F_{g-1,n-g}$</td>
</tr>
<tr>
<td>$p = 2$</td>
<td>$g \geq 2$</td>
<td>$\left( \frac{n-g-1}{g-1} \right) \left( \frac{1-\sqrt{\Lambda^<em>}}{\sqrt{\Lambda^</em>}} \right) \sim F_{2(g-1),2(n-g-1)}$</td>
</tr>
<tr>
<td>$p \geq 1$</td>
<td>$g = 2$</td>
<td>$\left( \frac{n-p-1}{p} \right) \left( \frac{1-\Lambda^<em>}{\Lambda^</em>} \right) \sim F_{p,n-p-1}$</td>
</tr>
<tr>
<td>$p \geq 1$</td>
<td>$g = 3$</td>
<td>$\left( \frac{n-p-2}{p} \right) \left( \frac{1-\sqrt{\Lambda^<em>}}{\sqrt{\Lambda^</em>}} \right) \sim F_{2p,2(n-p-2)}$</td>
</tr>
</tbody>
</table>

where $n = \sum_{k=1}^{g} n_k$.

If $n$ is large and $H_0$ is true, then

$$ - \left( n - 1 - \frac{p + g}{2} \right) \log(\Lambda^*) \approx \chi^2_{p(g-1)} $$
Other One-Way MANOVA Test Statistics

There are other popular MANOVA test statistics

- Lawley-Hotelling trace: $\text{tr}(BW^{-1})$
- Pillai trace: $\text{tr}(B[B + W]^{-1})$
- Roy’s largest root: maximum eigenvalue of $W(B + W)^{-1}$

Each of these test statistics has a corresponding (approximate) distribution, and all should produce similar inference for large $n$.

- Some evidence that Pillai’s trace is more robust to non-normality
One-Way MANOVA Example in R

```r
> X <- as.matrix(mtcars[,c("mpg","disp","hp","wt")])
> cylinder <- factor(mtcars$cyl)
> mod <- lm(X ~ cylinder)
> Manova(mod, test.statistic="Pillai")

Type II MANOVA Tests: Pillai test statistic

<table>
<thead>
<tr>
<th>Df</th>
<th>test stat</th>
<th>approx F</th>
<th>num Df</th>
<th>den Df</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>cylinder</td>
<td>2</td>
<td>1.0838</td>
<td>7.9845</td>
<td>8</td>
<td>54</td>
</tr>
</tbody>
</table>

---

Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> Manova(mod, test.statistic="Wilks")

Type II MANOVA Tests: Wilks test statistic

<table>
<thead>
<tr>
<th>Df</th>
<th>test stat</th>
<th>approx F</th>
<th>num Df</th>
<th>den Df</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>cylinder</td>
<td>2</td>
<td>0.091316</td>
<td>15.01</td>
<td>8</td>
<td>52</td>
</tr>
</tbody>
</table>

---

Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> Manova(mod, test.statistic="Roy")

Type II MANOVA Tests: Roy test statistic

<table>
<thead>
<tr>
<th>Df</th>
<th>test stat</th>
<th>approx F</th>
<th>num Df</th>
<th>den Df</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>cylinder</td>
<td>2</td>
<td>7.7873</td>
<td>52.564</td>
<td>4</td>
<td>27</td>
</tr>
</tbody>
</table>

---

Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

> Manova(mod, test.statistic="Hotelling-Lawley")

Type II MANOVA Tests: Hotelling-Lawley test statistic

<table>
<thead>
<tr>
<th>Df</th>
<th>test stat</th>
<th>approx F</th>
<th>num Df</th>
<th>den Df</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>cylinder</td>
<td>2</td>
<td>8.0335</td>
<td>25.105</td>
<td>8</td>
<td>50</td>
</tr>
</tbody>
</table>

---

Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Nathaniel E. Helwig (U of Minnesota)
Let $\alpha_{kj}$ denote the $j$-th element of $\alpha_k$ and note that

$$\hat{\alpha}_{kj} = \bar{x}_{kj} - \bar{x}_j$$

where $\bar{x}_j = \frac{1}{\sum_{k=1}^{g} n_k} \sum_{k=1}^{g} n_k \bar{x}_{kj}$ and $\bar{x}_{kj} = \frac{1}{n_k} \sum_{i=1}^{n_k} x_{kij}$

- $\bar{x}_{kj}$ is the $k$-th group’s mean for the $j$-th variable
- $\bar{x}_j$ is the overall mean of the $j$-th variable

The variance for the difference in the estimated treatment effects is

$$\text{Var}(\hat{\alpha}_{kj} - \hat{\alpha}_{\ell j}) = \text{Var}(\bar{x}_{kj} - \bar{x}_{\ell j}) = \left( \frac{1}{n_k} + \frac{1}{n_\ell} \right) \sigma_{jj}$$

where $\sigma_{jj}$ is the $j$-th diagonal element of $\Sigma$. 
To estimate the variance for the treatment effect difference, we use

\[ \hat{\text{Var}}(\hat{\alpha}_{kj} - \hat{\alpha}_{\ell j}) = \left( \frac{1}{n_k} + \frac{1}{n_\ell} \right) \frac{w_{jj}}{n - g} \]

where \( w_{jj} \) denotes the \( j \)-th diagonal of \( W \).

We can use Bonferroni’s method to control the familywise error rate

- Have \( p \) variables and \( g(g - 1)/2 \) pairwise comparisons
- Total of \( q = pg(g - 1)/2 \) tests to control for
- Use critical values \( t_{n-g}(\alpha/[2q]) = t_{n-g}(\alpha/[pg(g - 1)]) \)
# get least-squares means for each variable
> library(lsmeans)
> p <- ncol(X)
> lsm <- vector("list", p)
> names(lsm) <- colnames(X)
> for(j in 1:p){
+   wts <- rep(0, p)
+   wts[j] <- 1
+   lsm[[j]] <- lsmeans(mod, "cylinder", weights=wts)
+ }

> lsm[[1]]
cylinder  lsmean      SE  df lower.CL upper.CL
4       26.66364  0.9718008  29 24.67608  28.65119
6       19.74286  1.2182168  29 17.25132  22.23439
8       15.10000  0.8614094  29 13.33822  16.86178

Results are averaged over the levels of: rep.meas
Confidence level used: 0.95

> lsm[[3]]
cylinder  lsmean      SE  df lower.CL upper.CL
4       82.63636 11.43283  29  59.25361 106.0191
6      122.28571 14.33181  29  92.97388 151.5975
8      209.21429 10.13412  29 188.48769 229.9409

Results are averaged over the levels of: rep.meas
Confidence level used: 0.95
Form Simultaneous Confidence Intervals in R

```r
# get alpha level for Bonferroni correction
q <- p * 3 * (3-1) / 2
alpha <- 0.05 / (2*q)

# Bonferroni pairwise CIs for "mpg"
confint(contrast(lsm[[1]], "pairwise"), level=1-alpha, adj="none")

<table>
<thead>
<tr>
<th>contrast estimate</th>
<th>SE</th>
<th>df</th>
<th>lower.CL</th>
<th>upper.CL</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 - 6</td>
<td>6.920779</td>
<td>1.558348</td>
<td>29</td>
<td>1.6526941</td>
</tr>
<tr>
<td>4 - 8</td>
<td>11.563636</td>
<td>1.298623</td>
<td>29</td>
<td>7.1735655</td>
</tr>
<tr>
<td>6 - 8</td>
<td>4.642857</td>
<td>1.492005</td>
<td>29</td>
<td>-0.4009503</td>
</tr>
</tbody>
</table>

Results are averaged over the levels of: rep.meas
Confidence level used: 0.997916666666667

# Bonferroni pairwise CIs for "hp"
confint(contrast(lsm[[3]], "pairwise"), level=1-alpha, adj="none")

<table>
<thead>
<tr>
<th>contrast estimate</th>
<th>SE</th>
<th>df</th>
<th>lower.CL</th>
<th>upper.CL</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 - 6</td>
<td>-39.64935</td>
<td>18.33331</td>
<td>29</td>
<td>-101.6261</td>
</tr>
<tr>
<td>4 - 8</td>
<td>-126.57792</td>
<td>15.27776</td>
<td>29</td>
<td>-178.2252</td>
</tr>
<tr>
<td>6 - 8</td>
<td>-86.92857</td>
<td>17.55281</td>
<td>29</td>
<td>-146.2668</td>
</tr>
</tbody>
</table>

Results are averaged over the levels of: rep.meas
Confidence level used: 0.997916666666667
```
Testing the Homogeneity of Covariances Assumption

The one-way MANOVA model assumes that $\Sigma_1 = \cdots = \Sigma_g$.

To test $H_0 : \Sigma_1 = \cdots = \Sigma_g$ versus $H_1 : \Sigma_k \neq \Sigma_\ell$ for some $k, \ell \in \{1, \ldots, g\}$, we use the likelihood ratio test (LRT) statistic

$$\Lambda = \prod_{k=1}^{g} \left( \frac{|S_k|}{|S_P|} \right)^{(n_k-1)/2}$$

where $S_P = \frac{1}{n-g} W$ is the pooled covariance matrix estimate.

It was shown (by George Box) that $-2 \log(\Lambda) \approx \frac{1}{1-u} \chi^2_\nu$ where

$$u = \left( \sum_{k=1}^{g} \frac{1}{n_k} - \frac{1}{n} \right) \left( \frac{2p^2 + 3p - 1}{6(p+1)(g-1)} \right)$$

$$\nu = p(p+1)(g-1)/2$$
Box’s M-test for Homogeneity of Covariance Matrices

data: X
Chi-Sq (approx.) = 40.851, df = 20, p-value = 0.003892

We reject the null hypothesis that $\Sigma_1 = \cdots = \Sigma_g$, so our previous MANOVA results may be invalid.

$M$-test is sensitive to non-normality, so it may be ok to proceed with the MANOVA in some cases when the $M$-test rejects the null hypothesis.
Univariate Reminder: Two-Way ANOVA

\[ x_{\ell ki} \overset{\text{ind}}{\sim} N(\mu_{k\ell}, \sigma^2) \text{ for } k \in \{1, \ldots, a\}, \ell \in \{1, \ldots, b\}, \text{ and } i \in \{1, \ldots, n\}. \]

The two-way analysis of variance (ANOVA) model has the form

\[ x_{\ell ki} = \mu + \alpha_k + \beta_\ell + \gamma_{k\ell} + \epsilon_{\ell ki} \]

where

- \( \mu \) is the overall mean
- \( \alpha_k \) is the main effect for factor 1 \( (\sum_{k=1}^{a} \alpha_k = 0) \)
- \( \beta_\ell \) is the main effect for factor 2 \( (\sum_{\ell=1}^{b} \beta_\ell = 0) \)
- \( \gamma_{k\ell} \) is the interaction effect between factors 1 and 2 \( (\sum_{k=1}^{a} \gamma_{k\ell} = \sum_{\ell=1}^{b} \gamma_{k\ell} = 0) \)
- \( \epsilon_{\ell ki} \overset{iid}{\sim} N(0, \sigma^2) \) are error terms
The two-way ANOVA model implies the decomposition

\[ x_{\ell ki} = \bar{x} + (\bar{x}_{k.} - \bar{x}) + (\bar{x}_{.\ell} - \bar{x}) + (\bar{x}_{k\ell} - \bar{x}_{k.} - \bar{x}_{.\ell} + \bar{x}) + (x_{\ell ki} - \bar{x}_{k\ell}) \]

where

- \( \bar{x} = \frac{1}{abn} \sum_{k=1}^{a} \sum_{\ell=1}^{b} \sum_{i=1}^{n} x_{\ell ki} \) is the overall mean
- \( \bar{x}_{k.} = \frac{1}{bn} \sum_{\ell=1}^{b} \sum_{i=1}^{n} x_{\ell ki} \) is the mean of the \( k \)-th level of factor 1
- \( \bar{x}_{.\ell} = \frac{1}{an} \sum_{k=1}^{a} \sum_{i=1}^{n} x_{\ell ki} \) is the mean of the \( \ell \)-th level of factor 2
- \( \bar{x}_{k\ell} = \frac{1}{n} \sum_{i=1}^{n} x_{\ell ki} \) is the mean of the \( k \)-th level of factor 1 and the \( \ell \)-th level of factor 2
Inferences about Multiple Mean Vectors

Two-Way Multivariate Analysis of Variance (MANOVA)

Univariate Reminder: Two-Way ANOVA Sum-of-Sq

The two-way ANOVA sum-of-squares decomposition is

\[
\sum_{k=1}^{a} \sum_{\ell=1}^{b} \sum_{i=1}^{n} (x_{\ell ki} - \bar{x})^2 = \sum_{k=1}^{a} bn(\bar{x}_k. - \bar{x})^2 + \sum_{\ell=1}^{b} an(\bar{x}_.\ell - \bar{x})^2
\]

\[
+ \sum_{k=1}^{a} \sum_{\ell=1}^{b} n(\bar{x}_{k\ell} - \bar{x}_k. - \bar{x}_.\ell + \bar{x})^2
\]

\[
+ \sum_{k=1}^{a} \sum_{\ell=1}^{b} \sum_{i=1}^{n} (x_{\ell ki} - \bar{x}_{k\ell})^2
\]

\[
SSTotal \quad SSA \quad SSB \quad SSAB \quad SSE\text{rror}
\]
### Univariate Reminder: Two-Way ANOVA Inference

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum-of-Squares</th>
<th>D.F.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor 1</td>
<td>$\sum_{k=1}^{a} bn(\bar{x}_k - \bar{x})^2$</td>
<td>$a - 1$</td>
</tr>
<tr>
<td>Factor 2</td>
<td>$\sum_{\ell=1}^{b} an(\bar{x}_{.\ell} - \bar{x})^2$</td>
<td>$b - 1$</td>
</tr>
<tr>
<td>Interaction</td>
<td>$\sum_{k=1}^{a} \sum_{\ell=1}^{b} n(\bar{x}_{k\ell} - \bar{x}<em>k - \bar{x}</em>{.\ell} + \bar{x})^2$</td>
<td>$(a - 1)(b - 1)$</td>
</tr>
<tr>
<td>Error</td>
<td>$\sum_{k=1}^{a} \sum_{\ell=1}^{b} \sum_{i=1}^{n} (x_{\ell ki} - \bar{x}_{k\ell})^2$</td>
<td>$ab(n - 1)$</td>
</tr>
<tr>
<td>Total</td>
<td>$\sum_{k=1}^{a} \sum_{\ell=1}^{b} \sum_{i=1}^{n} (x_{\ell ki} - \bar{x})^2$</td>
<td>$abn - 1$</td>
</tr>
</tbody>
</table>

Reject $H_0: \alpha_1 = \cdots = \alpha_a = 0$ if $\frac{SSA}{a-1} > F_{a-1,ab(n-1)}(\alpha)$

Reject $H_0: \beta_1 = \cdots = \beta_b = 0$ if $\frac{SSB}{b-1} > F_{b-1,ab(n-1)}(\alpha)$

Reject $H_0: \gamma_{11} = \cdots = \gamma_{k\ell} = 0$ if $\frac{SSAB}{(a-1)(b-1)} > F_{(a-1)(b-1),ab(n-1)}(\alpha)$
Multivariate Extension of Two-Way ANOVA

\( x_{\ell ki} \overset{\text{ind}}{\sim} N(\mu_{k\ell}, \Sigma) \) for \( k \in \{1, \ldots, a\}, \ell \in \{1, \ldots, b\}, \) and \( i \in \{1, \ldots, n\}. \)

The two-way multivariate analysis of variance (MANOVA) has the form

\[
    x_{\ell ki} = \mu + \alpha_k + \beta_\ell + \tau_{k\ell} + \epsilon_{\ell ki}
\]

where the terms are analogues of those in the two-way ANOVA model.

The sample estimates of the model parameters are

\[
    x_{\ell ki} = \bar{x} + (\bar{x}_k - \bar{x}) + (\bar{x}_\ell - \bar{x}) + (\bar{x}_{k\ell} - \bar{x}_k - \bar{x}_\ell + \bar{x}) + (x_{\ell ki} - \bar{x}_{k\ell})
\]

where the terms are analogues of those in the two-way ANOVA model.
## Two-Way MANOVA Table

<table>
<thead>
<tr>
<th>Source</th>
<th>SSCP</th>
<th>D.F.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor 1</td>
<td>$\sum_{k=1}^{a} bn(\bar{x}_k. - \bar{x})(\bar{x}_k. - \bar{x})'$</td>
<td>$a - 1$</td>
</tr>
<tr>
<td>Factor 2</td>
<td>$\sum_{\ell=1}^{b} an(\bar{x}<em>.\ell - \bar{x})(\bar{x}</em>.\ell - \bar{x})'$</td>
<td>$b - 1$</td>
</tr>
<tr>
<td>Interaction</td>
<td>$\sum_{k=1}^{a} \sum_{\ell=1}^{b} n\tilde{x}<em>{k\ell} \tilde{x}</em>{k\ell}'$</td>
<td>$(a - 1)(b - 1)$</td>
</tr>
<tr>
<td>Error</td>
<td>$\sum_{k=1}^{a} \sum_{\ell=1}^{b} \sum_{i=1}^{n} (x_{\ell ki} - \tilde{x}<em>{k\ell})(x</em>{\ell ki} - \tilde{x}_{k\ell})'$</td>
<td>$ab(n - 1)$</td>
</tr>
<tr>
<td>Total</td>
<td>$\sum_{k=1}^{a} \sum_{\ell=1}^{b} \sum_{i=1}^{n} (x_{\ell ki} - \bar{x})(x_{\ell ki} - \bar{x})'$</td>
<td>$abn - 1$</td>
</tr>
</tbody>
</table>

where $\tilde{x}_{k\ell} = (\bar{x}_{k\ell} - \bar{x}_k. - \bar{x}_.\ell + \bar{x})$.

$$SSCP_T = SSCP_A + SSCP_B + SSCP_{AB} + SSCP_E$$
Two-Way MANOVA Inference

Reject $H_0 : \alpha_1 = \cdots = \alpha_a = 0_p$ if $\nu_A \log(\Lambda_A^*) > \chi^2_{(a-1)p}(\alpha)$

\[ \Lambda_A^* = \frac{|SSCP_E|}{|SSCP_A + SSCP_E|} \text{ and } \nu_A = - \left[ ab(n - 1) - \frac{p+1-(a-1)}{2} \right] \]

Reject $H_0 : \beta_1 = \cdots = \beta_b = 0_p$ if $\nu_B \log(\Lambda_B^*) > \chi^2_{(b-1)p}(\alpha)$

\[ \Lambda_B^* = \frac{|SSCP_E|}{|SSCP_B + SSCP_E|} \text{ and } \nu_B = - \left[ ab(n - 1) - \frac{p+1-(b-1)}{2} \right] \]

Reject $H_0 : \gamma_{11} = \cdots = \gamma_{k\ell} = 0_p$ if $\nu_{AB} \log(\Lambda_{AB}^*) > \chi^2_{(a-1)(b-1)p}(\alpha)$

\[ \Lambda_{AB}^* = \frac{|SSCP_E|}{|SSCP_{AB} + SSCP_E|} \text{ and } \nu_{AB} = - \left[ ab(n - 1) - \frac{p+1-(a-1)(b-1)}{2} \right] \]

Use $t_{ab(n-1)}$ distribution with Bonferroni correction for CIs.
Inferences about Multiple Mean Vectors
Two-Way Multivariate Analysis of Variance (MANOVA)

Two-Way MANOVA Example in R (fit model)

```r
# two-way manova with interaction
> data(mtcars)
> X <- as.matrix(mtcars[,c("mpg","disp","hp","wt")])
> cylinder <- factor(mtcars$cyl)
> transmission <- factor(mtcars$am)
> mod <- lm(X ~ cylinder * transmission)
> Manova(mod, test.statistic="Wilks")

Type II MANOVA Tests: Wilks test statistic

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>test stat</th>
<th>approx F</th>
<th>num Df</th>
<th>den Df</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>cylinder</td>
<td>2</td>
<td>0.09550</td>
<td>12.8570</td>
<td>8</td>
<td>46</td>
<td>1.689e-09 ***</td>
</tr>
<tr>
<td>transmission</td>
<td>1</td>
<td>0.43720</td>
<td>7.4019</td>
<td>4</td>
<td>23</td>
<td>0.0005512 ***</td>
</tr>
<tr>
<td>cylinder:transmission</td>
<td>2</td>
<td>0.58187</td>
<td>1.7880</td>
<td>8</td>
<td>46</td>
<td>0.1040156</td>
</tr>
</tbody>
</table>

---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

# refit additive model
> mod <- lm(X ~ cylinder + transmission)
> Manova(mod, test.statistic="Wilks")

Type II MANOVA Tests: Wilks test statistic

<table>
<thead>
<tr>
<th></th>
<th>Df</th>
<th>test stat</th>
<th>approx F</th>
<th>num Df</th>
<th>den Df</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>cylinder</td>
<td>2</td>
<td>0.11779</td>
<td>11.9610</td>
<td>8</td>
<td>50</td>
<td>2.414e-09 ***</td>
</tr>
<tr>
<td>transmission</td>
<td>1</td>
<td>0.49878</td>
<td>6.2805</td>
<td>4</td>
<td>25</td>
<td>0.001217 **</td>
</tr>
</tbody>
</table>

---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```
Two-Way MANOVA Example in R (get LS means)

```r
> p <- ncol(X)
> lsm.cyl <- lsm.trn <- vector("list", p)
> names(lsm) <- colnames(X)
> for(j in 1:p){
+   wts <- rep(0, p*2)
+   wts[1:2 + (j-1)*2] <- 1
+   lsm.cyl[[j]] <- lsmeans(mod, "cylinder", weights=wts)
+   wts <- rep(0, p*3)
+   wts[1:3 + (j-1)*3] <- 1
+   lsm.trn[[j]] <- lsmeans(mod, "transmission", weights=wts)
+ }

# print mpg LS mean for cylinder effect
> lsm.cyl[[1]]
cylinder   lsmean      SE  df lower.CL upper.CL
4   26.08183  0.9724817 28  24.08979  28.07387
6   19.92571  1.1653575 28  17.53858  22.31284
8   16.01427  0.9431293 28  14.08236  17.94618

Results are averaged over the levels of: transmission, rep.meas
Confidence level used: 0.95

# print mpg LS mean for transmission effect
> lsm.trn[[1]]
transmission  lsmean      SE  df lower.CL upper.CL
0    19.39396  0.7974085 28  17.76054  21.02738
1    21.95391  0.9283388 28  20.05230  23.85553

Results are averaged over the levels of: cylinder, rep.meas
Confidence level used: 0.95
```
Two-Way MANOVA Example in R (simultaneous CIs)

```r
# get alpha level for Bonferroni correction
> q <- p * (3 * (3-1) / 2 + 2 * (2-1) / 2)
> alpha <- 0.05 / (2*q)

# Bonferroni pairwise CIs for "mpg" (cylinder effect)
> confint(contrast(lsm.cyl[[1]], "pairwise"), level=1-alpha, adj="none")
  contrast estimate    SE df lower.CL upper.CL
4 - 6    6.156118 1.535723 28  0.7758311 11.536404
4 - 8    10.067560 1.452082 28  4.9803008 15.154818
6 - 8     3.911442 1.470254 28 -1.2394803  9.062364

Results are averaged over the levels of: transmission, rep.meas
Confidence level used: 0.9984375

# Bonferroni pairwise CIs for "mpg" (transmission effect)
> confint(contrast(lsm.trn[[1]], "pairwise"), level=1-alpha, adj="none")
  contrast estimate    SE df lower.CL upper.CL
0 - 1    -2.559954 1.297579 28 -7.105921  1.986014

Results are averaged over the levels of: cylinder, rep.meas
Confidence level used: 0.9984375
```