

Stat 8053: Manova (rev November 4, 2013)

Wishart Distribution

1. Suppose $\mathbf{z}_i \sim N_p(\mathbf{0}, \mathbf{I})$, $i = 1, \dots, n$. Then we say that

$$\mathbf{z}_i \mathbf{z}_i' \sim W_p(\mathbf{I}, 1)$$

has a Wishart distribution in p dimensions with covariance matrix \mathbf{I} and with 1 df. The diagonal elements of $\mathbf{z}_i \mathbf{z}_i'$ are all distributed as $\chi^2(1)$.

2. Write \mathbf{Z} as an $n \times p$ matrix with rows \mathbf{x}_i' . If the \mathbf{z}_i are independent, then

$$\mathbf{Z}'\mathbf{Z} = \sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i' \sim \sum_{i=1}^n W_p(\mathbf{I}, 1) = W_p(\mathbf{I}, n)$$

has a Wishart distribution with covariance matrix \mathbf{I} and n degrees of freedom. If $p = 1$, $\sum z_i^2 \sim \chi^2(n)$. Also, $E(W_p(\mathbf{I}), n) = n\mathbf{I}$.

3. Let $\mathbf{Y} = \mathbf{Z}\Sigma^{1/2} \sim N_p(\mathbf{0}, \Sigma)$. Then

$$\begin{aligned}\mathbf{Y}'\mathbf{Y} &\sim W_p(\Sigma, n) \\ E(W_p(\Sigma, n)) &= n\Sigma\end{aligned}$$

If $\Sigma^{-1/2}$ is any square root of Σ , then $\Sigma^{-1/2}\mathbf{Y} \sim N_p(\mathbf{0}, \mathbf{I})$, and

$$\begin{aligned}\Sigma^{-1/2}\mathbf{Y}'\mathbf{Y}\Sigma^{-1/2'} &\sim W_p(\mathbf{I}, n) \\ \mathbf{Y}'\mathbf{Y} &\sim \Sigma^{1/2}W_p(\mathbf{I}, n)\Sigma^{1/2'}\end{aligned}$$

If $p = 1$, then $\mathbf{Y}'\mathbf{Y} \sim \sigma^2\chi^2(n)$. Like the χ^2 distributions, the Wishart distribution with non-identity covariance matrix differs from the standard Wishart by multiplication by constants.

4. If \mathbf{A} is $q \times p$, and \mathbf{a} is a $p \times 1$ vector, then

$$\begin{aligned}\mathbf{A}(\mathbf{Y}'\mathbf{Y})\mathbf{A}' &\sim W_q(\mathbf{A}\Sigma\mathbf{A}', n) = \Sigma^{-1/2}\mathbf{A}'W_p(\mathbf{I}, n)\mathbf{A}\Sigma^{-1/2'} \\ \mathbf{a}'(\mathbf{Y}'\mathbf{Y})\mathbf{a} &\sim (\mathbf{a}'\Sigma\mathbf{a})\chi^2(n)\end{aligned}$$

5. Suppose that Σ is a $p \times p$ symmetric matrix. By the *Spectral Theorem*, we can find a $p \times p$ orthogonal matrix $\Gamma = (\gamma_1, \dots, \gamma_p)$, $\Gamma\Gamma' = \Gamma'\Gamma = \mathbf{I}$, and a $p \times p$ diagonal matrix Λ with non-negative diagonal elements $\lambda_1, \dots, \lambda_p$, such that

$$\begin{aligned}\Sigma &= \Gamma\Lambda\Gamma' = \sum \lambda_i \gamma_i \gamma_i' \\ \Sigma^{-1} &= \Gamma\Lambda^{-1}\Gamma' = \sum \lambda_i^{-1} \gamma_i \gamma_i'\end{aligned}$$

where the inverse requires all the $\lambda_i > 0$.

6. A (not the) square root of Σ and its inverse are $\Sigma^{1/2} = \Gamma\Lambda^{1/2}$ and $\Sigma^{-1/2} = \Gamma\Lambda^{-1/2}$. The point here is to show that a square root exists and is computable. The γ are eigenvectors and the λ are corresponding eigenvalues. A symmetric square root is given by $\Gamma\Lambda^{1/2}\Gamma'$.

Hotelling's T^2

1. If $\mathbf{y} \sim N_p(\boldsymbol{\mu}, \Sigma)$, then if Σ is of full rank, $\Sigma^{-1/2}(\mathbf{y} - \boldsymbol{\mu}) \sim N_p(\mathbf{0}, \mathbf{I})$, and $(\mathbf{y} - \boldsymbol{\mu})'\Sigma^{-1}(\mathbf{y} - \boldsymbol{\mu}) \sim W_p(\mathbf{I}, 1)$.
2. In most problems Σ is unknown. Suppose that $\mathbf{M} \sim W_p(\Sigma, n)$ is an estimate of Σ that is independent of \mathbf{y} . Then

$$T^2 = n(\mathbf{y} - \boldsymbol{\mu})'\mathbf{M}^{-1}(\mathbf{y} - \boldsymbol{\mu})$$

Is distributed as Hotelling's $T^2(p, n)$ in p dimensions with n degrees of freedom. If $p = 1$ then T^2 is the square of Student's $t(n)$.

- 3.

$$T^2(p, n) = \frac{np}{n - p + 1} F(p, n - p + 1)$$

so T^2 has a distribution proportional to a central F distribution, so the F can be used to get significance levels if T^2 is used for testing.

Cochran's Theorem

Cochran's theorem provides one of many ways to get to Anova and to Manova. Suppose \mathbf{Y} is an $n \times p$ matrix with iid $N_p(\mathbf{0}, \Sigma)$ rows. Also suppose that \mathbf{C} is any $n \times n$ symmetric matrix. Define $\mathbf{Y}'\mathbf{C}\mathbf{Y}$ to be a *sum of squares and cross products matrix*. With another application of the spectral theorem, we can always write $\mathbf{C} = \sum \delta_i \gamma_i \gamma_i'$ where the δ_i are nonnegative eigenvalues and the γ_i are eigenvectors.

1. For any \mathbf{C} ,

$$\begin{aligned}\mathbf{Y}'\mathbf{C}\mathbf{Y} &= \mathbf{Y}' \left(\sum \delta_i \boldsymbol{\gamma}_i \boldsymbol{\gamma}_i' \right) \mathbf{Y} = \sum \delta_i (\mathbf{Y}' \boldsymbol{\gamma}_i \boldsymbol{\gamma}_i' \mathbf{Y}) \\ &\sim \sum_{i=1}^n \delta_i W_p(\boldsymbol{\Sigma}, 1)\end{aligned}$$

which is a weighted linear combination of Wishart random variables with nonnegative weights. For $p = 1$, we get a weighted sum of central $\chi^2(1)$ s.

2. We will have

$$\mathbf{Y}'\mathbf{C}\mathbf{Y} \sim W_p(\boldsymbol{\Sigma}, r)$$

if and only if one of the following three equivalent conditions holds: (a) all δ_i equal either 0 or 1; (b) $\mathbf{C} = \mathbf{C}' = \mathbf{C}^2$; (c) \mathbf{C} is a projection matrix. Here $r = \sum \delta_i$ is the *rank* of the matrix \mathbf{C} .

3. The two sums of squares and cross products matrices $\mathbf{Y}'\mathbf{C}\mathbf{Y}$ and $\mathbf{Y}'\mathbf{B}\mathbf{Y}$ have independent Wishart distributions if and only if \mathbf{B} and \mathbf{C} are projections and $\mathbf{C}\mathbf{B} = \mathbf{0}$. This is the basis for both Anova and for Manova.

Applying Cochran's Theorem: One-sample Test

Suppose \mathbf{y}_i are iid, $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

1. Define $\mathbf{J} = \mathbf{1}\mathbf{1}'/n$ be the projection on the column of 1's, $\mathbf{J} = \mathbf{J}' = \mathbf{J}^2$. Then $\mathbf{J}\mathbf{Y}$ returns a matrix with all rows given by $\bar{\mathbf{y}}'$, the sample mean vector. The matrix $\mathbf{Y}'\mathbf{J}\mathbf{Y} = n\bar{\mathbf{y}}\bar{\mathbf{y}}'$. In the $p = 1$ case, this is $n\bar{y}^2$. $\mathbf{Y}'\mathbf{J}\mathbf{Y}$ does not have a Wishart distribution unless $\boldsymbol{\mu} = \mathbf{0}$.
2. $(\mathbf{I} - \mathbf{J}) = (\mathbf{I} - \mathbf{J})' = (\mathbf{I} - \mathbf{J})^2$ is also a projection, and $(\mathbf{I} - \mathbf{J})\mathbf{Y}$ returns a matrix like \mathbf{Y} but centered by subtracting the mean $\bar{\mathbf{y}}$ from each row.
3. $(\mathbf{I} - \mathbf{J})\mathbf{Y}$ has mean $\mathbf{0}$, and the corrected sum of squares and cross products matrix.

$$(n-1)\mathbf{S}_u = \mathbf{Y}'(\mathbf{I} - \mathbf{J})\mathbf{Y} \sim W_p(\boldsymbol{\Sigma}, n-1)$$

and so \mathbf{S}_u is an unbiased estimate of $\boldsymbol{\Sigma}$. Since $\mathbf{J}(\mathbf{I} - \mathbf{J}) = \mathbf{0}$, \mathbf{S}_u and $\bar{\mathbf{y}}$ are independent. The textbook defines $\mathbf{S} = ((n-1)/n)\mathbf{S}_u$.

4. Consider a test of $\boldsymbol{\mu} = \boldsymbol{\mu}_0$ versus $\boldsymbol{\mu} \neq \boldsymbol{\mu}_0$. Compute

$$T^2 = n(\bar{\mathbf{y}} - \boldsymbol{\mu}_0)' \mathbf{S}_u^{-1} (\bar{\mathbf{y}} - \boldsymbol{\mu}_0)$$

Then $T^2 \sim T^2(p, n-1) = [(n-1)p/(n-p)]F(p, n-p)$.

More generally...

Suppose we have the general multivariate linear model.

$$\begin{aligned}\mathbf{Y}|\mathbf{X} &= \mathbf{X}\boldsymbol{\beta} + \mathbf{E} \\ &= \mathbf{1}\beta_0 + \mathbf{X}_1\boldsymbol{\beta}'_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{E}\end{aligned}$$

Here \mathbf{Y} is an $n \times m$ data matrix of responses for the m response variables, $(\mathbf{1}, \mathbf{X}_1)$ is $n \times q$ for the first q columns of \mathbf{X} , and \mathbf{X}_2 is the remaining columns. The $p \times m$ matrix $\boldsymbol{\beta}$ has first row β'_0 , next q rows $\boldsymbol{\beta}'_1$ and remaining rows $\boldsymbol{\beta}_2$. The error matrix \mathbf{E} has iid $N_p(0, \boldsymbol{\Sigma})$ elements.

Let \mathbf{P}_1 be the projection on $(\mathbf{1}, \mathbf{X}_1)$, \mathbf{P}_{12} the projection on \mathbf{X} , and so $\mathbf{P}_{2|1} = \mathbf{P}_{12} - \mathbf{P}_1$ is the projection on the part of \mathbf{X}_2 orthogonal to $(\mathbf{1}, \mathbf{X}_1)$, and $\mathbf{P}_1\mathbf{P}_{2|1} = \mathbf{0}$. In this over-brief presentation, we skip general methods for computing the \mathbf{P} . Then

1. Under the hypothesis that $\boldsymbol{\beta}_2 = \mathbf{0}$, $\mathbf{B} = \mathbf{Y}'\mathbf{P}'_{2|1}\mathbf{Y} \sim W_p(\boldsymbol{\Sigma}, p - q)$. The matrix \mathbf{B} is the *between* sum of squares and cross-products.
2. For any value of $\boldsymbol{\beta}$, $\mathbf{W} = \mathbf{Y}'(\mathbf{I} - \mathbf{P}_{12})'\mathbf{Y} \sim W_p(\boldsymbol{\Sigma}, n - p)$. The matrix \mathbf{W} is the *within* sum of squares and cross-products.
3. A test of $\boldsymbol{\beta} = \mathbf{0}$ is then based of “comparing” these two Wishart matrices, via $\mathbf{W}^{-1}\mathbf{B}$.

One-way Multivariate Analysis of Variance

For the one-way Manova model with p groups, $\mathbf{X} = (\mathbf{1}, \mathbf{X}_2)$, where as in Anova \mathbf{X}_2 is some basis for the $p - 1$ dimensional space. For this model, \mathbf{X}_1 is empty. Using the R parametrization, $\mathbf{X}_2(i, j) = 1$ if the i th observation is in group $j + 1$ for $j = 2, \dots, p$, but as in Anova *the results here do not depend on parameterization, in general as long as the marginality property is followed*. We want to test the coefficients for $\mathbf{X}_2 = \mathbf{0}$.

The projection on the whole space is $\mathbf{P}_{12} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$, while the projection on the space spanned by $\mathbf{1}$ is \mathbf{J} . Then $\mathbf{P}_{2|1} = \mathbf{P}_X - \mathbf{J}$ is the projection on the columns of \mathbf{X}_2 orthogonal to \mathbf{J} .

Write $\bar{\mathbf{y}}_i$ for the mean in the n_i observations in the i -th group; $\bar{\mathbf{y}} = \sum n_i \bar{\mathbf{y}}_i / \sum n_i$, then the multivariate analysis of variance, Manova, is

Source	df	SS
Mean	1	$\mathbf{Y}'\mathbf{J}\mathbf{Y} = n\bar{\mathbf{y}}\bar{\mathbf{y}}'$
Groups	$d_{2 1} = g - 1$	$\mathbf{B} = \mathbf{Y}'\mathbf{P}_{2 1}\mathbf{Y} = \sum n_i(\bar{\mathbf{y}}_i - \bar{\mathbf{y}})(\bar{\mathbf{y}}_i - \bar{\mathbf{y}})'$
Error	$d = \sum(n_i - 1)$	$\mathbf{W} = \mathbf{Y}'(\mathbf{I} - \mathbf{P}_{12})\mathbf{Y} = \sum \sum (x_{ij} - \bar{\mathbf{y}}_i)(x_{ij} - \bar{\mathbf{y}}_i)'$

$\mathbf{S}_u = \mathbf{W} / (\sum(n_i - 1))$ is the unbiased estimate of the common $\boldsymbol{\Sigma}$.

In the $p = 1$ case the likelihood ratio test of all group means equal is an F test, but the multivariate test is *not* an exact F -test, even under normality. The likelihood ratio test can be shown to be

$$\Lambda = \frac{|\mathbf{W}|}{|\mathbf{W} + \mathbf{B}|} = \frac{1}{|\mathbf{I} + \mathbf{W}^{-1}\mathbf{B}|}$$

where $|\mathbf{A}|$ is the determinant of \mathbf{A} , and we reject for small values of Λ . Λ depends on the data only through the eigenvalues of the matrix $\mathbf{W}^{-1}\mathbf{B}$. If $\lambda_1 \geq \dots \lambda_p$ are the eigenvalues of $\mathbf{W}^{-1}\mathbf{B}$, then

$$\Lambda = \prod_{i=1}^p \frac{1}{1 + \lambda_i}$$

The distribution of Λ depends on the distribution of the eigenvalues. Here are the usual methods of obtaining p -values.

1. Asymptotic: $-2 \log(\Lambda) \rightarrow \chi^2[pd_{2|1}]$.

2. Bartlett corrected with $n = \sum n_i$:

$$-(n - 1 + \frac{p + d_{2|1} + 1}{2}) \log(\Lambda) \rightarrow \chi^2(pd_{2|1})$$

3. The exact distribution is available in special cases with $d_{2|1}p \leq 3$.

4. Replace Λ by a different function of the λ s but still use asymptotics to get p -values. The Lawley-Hotelling method uses $\text{tr}(\mathbf{W}^{-1}\mathbf{B}) = \sum \lambda_i$. A small-sample F -approximation is used to get p -values.

5. Pillai's trace criterion appears to be the most popular and is often the default in computer programs, $\text{tr}(\mathbf{B}(\mathbf{B} + \mathbf{W})^{-1}) = \sum (\lambda_i / (1 + \lambda_i))$. A small-sample F -approximation is used to get p -values.

6. Roy's largest root is the maximum eigenvalue of $\mathbf{W}(\mathbf{B} + \mathbf{W})^{-1} = \lambda_1 / (1 + \lambda_1)$. It is equivalent to testing all possible univariate Anova's with response $\mathbf{Y}'\mathbf{a}$ for all \mathbf{a} , and then computing the p -value based on the maximum of all these tests. The distribution of this statistic is known and tables (or algorithms to compute tables) are available.

We use as an example data from Johnson and Wichern, *Applied Multivariate Statistical Analysis*, 6th Ed., p. 318. The data are concerned with a two-factor experiment done to study the properties of extruded plastic film. The two factors are rate of extrusion and amount of additive, both at two levels. There are three response variables, y_1 = tear resistance, y_2 = gloss, and y_3 = opacity. These three variables are on completely different scales and in different units.

```

loc <- "http://www.stat.umn.edu/~sandy/courses/8053/Data/Wichern_data/T6-4.dat"
data <- read.table(url(loc))
colnames(data) <- c("Rate", "Additive", "Y1", "Y2", "Y3")
data$Rate <- factor(data$Rate)
data$Additive <- factor(data$Additive)
data$RateAdd <- factor(with(data, paste("R", Rate, "A", Additive, sep="")))
data

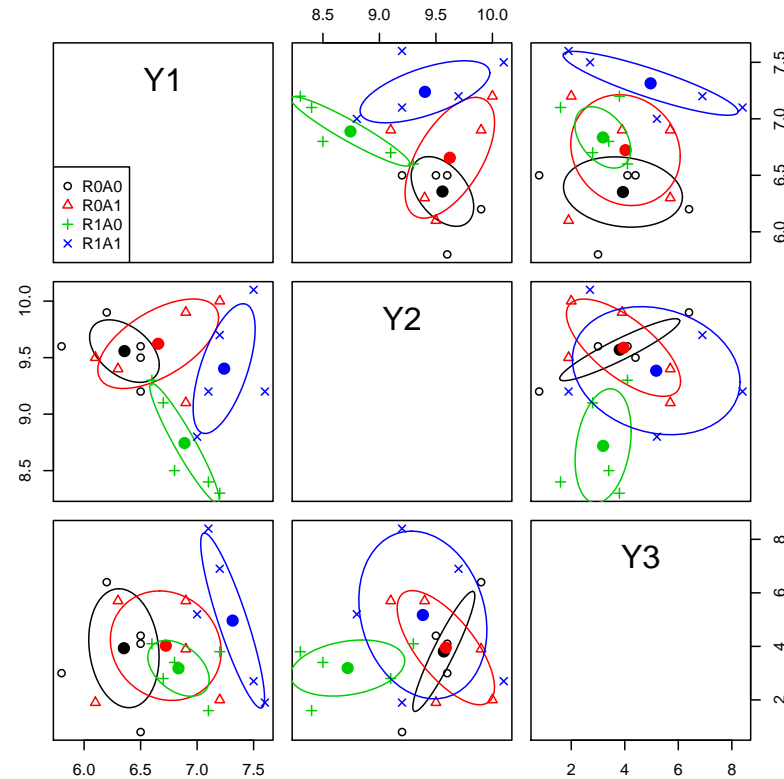
```

	Rate	Additive	Y1	Y2	Y3	RateAdd
1	0	0	6.5	9.5	4.4	R0A0
2	0	0	6.2	9.9	6.4	R0A0
3	0	0	5.8	9.6	3.0	R0A0
4	0	0	6.5	9.6	4.1	R0A0
5	0	0	6.5	9.2	0.8	R0A0
6	0	1	6.9	9.1	5.7	R0A1
7	0	1	7.2	10.0	2.0	R0A1
8	0	1	6.9	9.9	3.9	R0A1
9	0	1	6.1	9.5	1.9	R0A1
10	0	1	6.3	9.4	5.7	R0A1
11	1	0	6.7	9.1	2.8	R1A0
12	1	0	6.6	9.3	4.1	R1A0
13	1	0	7.2	8.3	3.8	R1A0
14	1	0	7.1	8.4	1.6	R1A0
15	1	0	6.8	8.5	3.4	R1A0
16	1	1	7.1	9.2	8.4	R1A1
17	1	1	7.0	8.8	5.2	R1A1
18	1	1	7.2	9.7	6.9	R1A1
19	1	1	7.5	10.1	2.7	R1A1
20	1	1	7.6	9.2	1.9	R1A1

```

library(car)
print(scatterplotMatrix( ~ Y1 + Y2 + Y3 | RateAdd, data=data, smooth=FALSE,
  reg.line=FALSE, ellipse=TRUE, levels= c(0.5), by.groups=TRUE,
  diagonal="none"))

```



The model assumes that the within group covariance matrices are equal. The ellipses are contours of constant probability based on the normal distribution taking the responses two at a time. Within a plot all the ellipses should be of the same shape and orientation, something not apparent here. However, the sample sizes are much too small to make much of this. Under the H_0 the larger dots on each plot should be the same. Y_1 seems to show larger differences than do the other two responses.

One-sample test

Let's start by a test that all the variable means are equal to zero, effectively a one-sample test:

```
# One sample test:
ybar <- colMeans(data[, 3:5])
n <- dim(data)[1] # sample size
p <- 3 # dimension of response
```

```

S_u <- var(data[, 3:5])
T2 <- n * t(ybar) %*% solve(S_u) %*% (ybar)
F <- (n - p)/((n-1)*p) * T2
c(T2=T2, F=F)

```

```

      T2      F
12686  3783

```

You can get `lm` to do the same test:

```

anova(m0 <- lm(cbind(Y1, Y2, Y3) ~ 1, data))

```

Analysis of Variance Table

```

              Df Pillai approx F num Df den Df Pr(>F)
(Intercept)   1  0.999      3783      3    17 <2e-16
Residuals    19

```

The computation by hand matches the F for the Pillai test because there is only one eigenvalue, so all tests are equivalent. A slight modification is to test $\boldsymbol{\mu} = \mathbf{c}$ for some vector \mathbf{c} :

```

C <- c(7, 9, 4)
anova(update(m0, I(cbind(Y1, Y2, Y3) - outer(rep(1, n), C)) ~ .))

```

Analysis of Variance Table

```

              Df Pillai approx F num Df den Df Pr(>F)
(Intercept)   1  0.348      3.02      3    17  0.059
Residuals    19

```

I set \mathbf{c} to equal `round(xbar)`.

One-way Manova

Let's fit a one-way Manova with four levels created all possible combinations of the factors:

```

(m1 <- lm(cbind(Y1, Y2, Y3) ~ RateAdd, data))

```



```
Call:
lm(formula = cbind(Y1, Y2, Y3) ~ RateAdd, data = data)
```

Coefficients:

	Y1	Y2	Y3
(Intercept)	6.30	9.56	3.74
RateAddR0A1	0.38	0.02	0.10
RateAddR1A0	0.58	-0.84	-0.60
RateAddR1A1	0.98	-0.16	1.28

The left-side of the model is a matrix, not a data-frame, with three columns; otherwise the use of `lm` is the same as for univariate models. The printed output is a table of coefficients for each of the responses. The `summary` method is illustrated later in this handout.

Manova is computed using the `Anova` function in `car`

```
library(car)
a1 <- Anova(m1)
summary(a1)
```

Type II MANOVA Tests:

Sum of squares and products for error:

	Y1	Y2	Y3
Y1	1.764	0.020	-3.070
Y2	0.020	2.628	-0.552
Y3	-3.070	-0.552	64.924

Term: RateAdd

Sum of squares and products for the hypothesis:

	Y1	Y2	Y3
Y1	2.5015	-0.8055	2.831
Y2	-0.8055	2.4575	2.461
Y3	2.8305	2.4615	9.282

Multivariate Tests: RateAdd

	Df	test	stat	approx F	num Df	den Df	Pr(>F)
Pillai	3		1.146	3.295	9	48.00	0.003350
Wilks	3		0.178	3.925	9	34.22	0.001663
Hotelling-Lawley	3		2.817	3.965	9	38.00	0.001245
Roy	3		1.870	9.971	3	16.00	0.000603

The “Sum of squares and products for error” is the matrix **W** and “Sum of squares and products for the hypothesis” is **B**. The tests are based on the eigenvalues of $\mathbf{W}^{-1}\mathbf{B}$.

The usual print method for **Anova** uses some of this output:

```
print(a1, test.statistic="Pillai")
```

Type II MANOVA Tests: Pillai test statistic

	Df	test	stat	approx F	num Df	den Df	Pr(>F)
RateAdd	3		1.15	3.29	9	48	0.0034

Two-way Manova

```
summary(m2 <- lm(cbind(Y1, Y2, Y3) ~ Rate*Additive, data))
```

Response Y1 :

Call:

```
lm(formula = Y1 ~ Rate * Additive, data = data)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.580	-0.205	0.060	0.220	0.520

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	6.300	0.148	42.43	<2e-16
Rate1	0.580	0.210	2.76	0.014
Additive1	0.380	0.210	1.81	0.089

Rate1:Additive1 0.020 0.297 0.07 0.947

Residual standard error: 0.332 on 16 degrees of freedom
Multiple R-squared: 0.586, Adjusted R-squared: 0.509
F-statistic: 7.56 on 3 and 16 DF, p-value: 0.00227

Response Y2 :

Call:
lm(formula = Y2 ~ Rate * Additive, data = data)

Residuals:

Min	1Q	Median	3Q	Max
-0.600	-0.245	-0.070	0.325	0.700

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	9.560	0.181	52.75	<2e-16
Rate1	-0.840	0.256	-3.28	0.0047
Additive1	0.020	0.256	0.08	0.9388
Rate1:Additive1	0.660	0.362	1.82	0.0874

Residual standard error: 0.405 on 16 degrees of freedom
Multiple R-squared: 0.483, Adjusted R-squared: 0.386
F-statistic: 4.99 on 3 and 16 DF, p-value: 0.0125

Response Y3 :

Call:
lm(formula = Y3 ~ Rate * Additive, data = data)

Residuals:

Min	1Q	Median	3Q	Max
-----	----	--------	----	-----

-3.12 -1.61 0.22 1.18 3.38

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3.740	0.901	4.15	0.00075
Rate1	-0.600	1.274	-0.47	0.64403
Additive1	0.100	1.274	0.08	0.93841
Rate1:Additive1	1.780	1.802	0.99	0.33789

Residual standard error: 2.01 on 16 degrees of freedom

Multiple R-squared: 0.125, Adjusted R-squared: -0.039

F-statistic: 0.762 on 3 and 16 DF, p-value: 0.531

Anova from the `car` package does all the usual tests. If the data were unbalanced the correct tests are the Type II tests that obey the marginality principle.

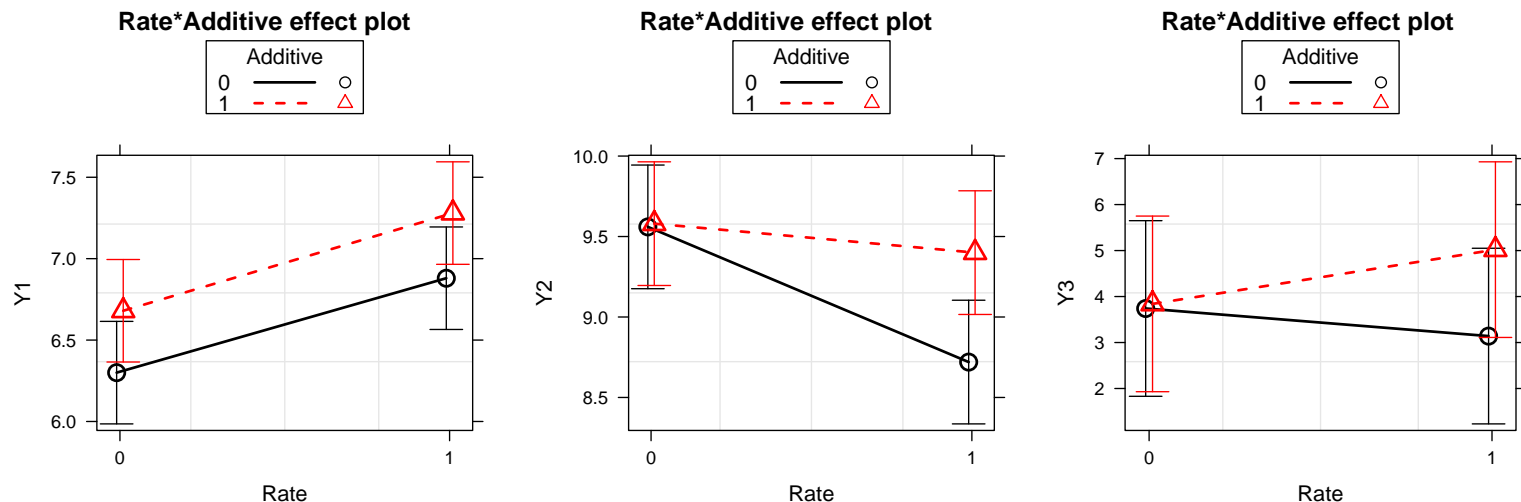
`Anova(m2)`

Type II MANOVA Tests: Pillai test statistic

	Df test	stat	approx F	num Df	den Df	Pr(>F)
Rate	1	0.618	7.55	3	14	0.003
Additive	1	0.477	4.26	3	14	0.025
Rate:Additive	1	0.223	1.34	3	14	0.302

In this problem all the tests are based on 1 eigenvalue, and hence have the same asymptotic distribution.

```
# install.packages("effects", repos="http://r-forge.r-project.org")
library(effects)
plot(Effect(c("Rate", "Additive"), m2), rows=1, cols=3, multiline=TRUE, grid=TRUE, ci.style="bars")
```



The effects plots suggest a *hint* of an interaction for Y2; how could you explore this further? Are tests believable?

Testing Linear Hypotheses

The multivariate general linear model is

$$\mathbf{Y}|\mathbf{X} = \mathbf{X}\boldsymbol{\beta} + \mathbf{E} \quad (1)$$

where \mathbf{Y} is $n \times m$, \mathbf{X} is $n \times p$, $\boldsymbol{\beta}$ is $p \times m$ and rows of \mathbf{E} have $N_p(\mathbf{0}, \boldsymbol{\Sigma})$ distributions.

Consider testing a general linear hypothesis that places restrictions on the coefficient matrix $\boldsymbol{\beta}$,

$$H_0 : \mathbf{L}\boldsymbol{\beta} = \mathbf{C}$$

where \mathbf{L} is a $q \times p$ full rank *hypothesis matrix*, and \mathbf{C} is a $q \times m$ matrix of constants, usually zeroes. The sum of squares and cross-products matrix for this hypothesis is

$$\mathbf{B}_H = (\mathbf{L}\hat{\boldsymbol{\beta}} - \mathbf{C})' [\mathbf{L}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}']^{-1} (\mathbf{L}\hat{\boldsymbol{\beta}} - \mathbf{C})$$

where $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ is the ols estimator. Using Cochran's theorem again, tests are based on the eigenvalues of $\mathbf{W}^{-1}\mathbf{B}_H$.

For example, consider the `Rate:Additive` interaction. Using model `m2`, this corresponds to testing the coefficient for `Rate1:Additive1 = 0`, so $\mathbf{L} = (0, 0, 0, 1)$,

```
linearHypothesis(m2, c(0, 0, 0, 1))
```

Sum of squares and products for the hypothesis:

	Y1	Y2	Y3
Y1	0.0005	0.0165	0.0445
Y2	0.0165	0.5445	1.4685
Y3	0.0445	1.4685	3.9605

Sum of squares and products for error:

	Y1	Y2	Y3
Y1	1.764	0.020	-3.070
Y2	0.020	2.628	-0.552
Y3	-3.070	-0.552	64.924

Multivariate Tests:

	Df	test stat	approx F	num Df	den Df	Pr(>F)
Pillai	1	0.2229	1.339	3	14	0.302
Wilks	1	0.7771	1.339	3	14	0.302
Hotelling-Lawley	1	0.2868	1.339	3	14	0.302
Roy	1	0.2868	1.339	3	14	0.302

In model m1, the interaction is given by $\mathbf{L} = (0, -1, -1, 1)$,

`linearHypothesis(m1, c(0, -1, -1, 1))`

Sum of squares and products for the hypothesis:

	Y1	Y2	Y3
Y1	0.0005	0.0165	0.0445
Y2	0.0165	0.5445	1.4685
Y3	0.0445	1.4685	3.9605

Sum of squares and products for error:

	Y1	Y2	Y3
Y1	1.764	0.020	-3.070
Y2	0.020	2.628	-0.552
Y3	-3.070	-0.552	64.924

Multivariate Tests:

	Df	test	stat	approx F	num Df	den Df	Pr(>F)
Pillai	1		0.2229	1.339	3	14	0.302
Wilks	1		0.7771	1.339	3	14	0.302
Hotelling-Lawley	1		0.2868	1.339	3	14	0.302
Roy	1		0.2868	1.339	3	14	0.302

`linearHypothesis` has several shortcuts for specifying **L**: see <http://z.umn.edu/ffwmult>.

Repeated Measures

You probably already know two methods for the analysis of repeated measures on the same variable on the same subject: using mixed-effects models with a random subject effects, and using split-plot type analysis of variance with multiple error terms. A third approach uses the multivariate linear model (1), where the columns of **Y** are measures of the same variable over time. In this case hypotheses of the form

$$H_0 : \mathbf{L}\boldsymbol{\beta}\mathbf{P} = \mathbf{C}$$

may make sense. Here **L** is a hypothesis matrix concerning contrasts of columns of $\boldsymbol{\beta}$, or between-subject contrasts, while the $m \times v$ matrix **P** defines contrasts over the repeated measures, or within-subjects contrasts. In this case the **B_H** matrix is

$$\mathbf{B}_H = (\mathbf{L}\hat{\boldsymbol{\beta}}\mathbf{P} - \mathbf{C})' [\mathbf{L}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}]^{-1} (\mathbf{L}\hat{\boldsymbol{\beta}}\mathbf{P} - \mathbf{C})$$

Testing is based on the eigenvalues of $\mathbf{W}^{-1}\mathbf{B}_H$.

The data in the `car` data set `OBrienKaiser` consists of 16 subjects, with between-subjects effects `gender` with two levels and `treatment` with three levels:

```
xtabs(~ gender + treatment, OBrienKaiser)
```

```

      treatment
gender control A B
F           2  2  4
M           3  2  3

```

The file also includes 15 response variables, 5 hourly measurements in a pretest phase, 5 5 hourly measurements in a post test phase, and finally 5 5 hourly measurements in a follow-up, so there are 15 repeated observations. We create a within-subject design matrix:

```
head(OBrienKaiser)
```

	treatment	gender	pre.1	pre.2	pre.3	pre.4	pre.5	post.1	post.2	post.3	post.4	post.5	fup.1
1	control	M	1	2	4	2	1	3	2	5	3	2	2
2	control	M	4	4	5	3	4	2	2	3	5	3	4
3	control	M	5	6	5	7	7	4	5	7	5	4	7
4	control	F	5	4	7	5	4	2	2	3	5	3	4
5	control	F	3	4	6	4	3	6	7	8	6	3	4
6	A	M	7	8	7	9	9	9	9	10	8	9	9
	fup.2	fup.3	fup.4	fup.5									
1	3	2	4	4									
2	5	6	4	1									
3	6	9	7	6									
4	4	5	3	4									
5	3	6	4	3									
6	10	11	9	6									

```

phase <- factor(rep(c("pretest", "posttest", "followup"), each=5),
  levels=c("pretest", "posttest", "followup"))
hour <- ordered(rep(1:5, 3))
idata <- data.frame(phase, hour)
idata

```

	phase	hour
1	pretest	1
2	pretest	2
3	pretest	3
4	pretest	4
5	pretest	5
6	posttest	1
7	posttest	2
8	posttest	3
9	posttest	4
10	posttest	5
11	followup	1
12	followup	2
13	followup	3


```
14 followup      4
15 followup      5
```

The within-subject effects are called **phase** and **hour**.

We first fit the between-subjects model only:

```
mod.ok <- lm(cbind(pre.1, pre.2, pre.3, pre.4, pre.5, post.1, post.2, post.3, post.4, post.5,
                  fup.1, fup.2, fup.3, fup.4, fup.5) ~ treatment*gender, data=OBrienKaiser)
```

This fit has $n = 16$, $m = 15$, and $p = 5$. Manova cannot be computed because the matrix **W** has no degrees of freedom. The **Anova** function takes additional arguments **idata** and **idesign** to specify the within-subject effects:

```
Anova(mod.ok, idata=idata, idesign=~phase*hour)
```

Type II Repeated Measures MANOVA Tests: Pillai test statistic

	Df	test	stat	approx	F	num	Df	den	Df	Pr(>F)
(Intercept)	1		0.970		318		1	10		6.5e-09
treatment	2		0.481		5		2	10		0.03769
gender	1		0.204		3		1	10		0.14097
treatment:gender	2		0.364		3		2	10		0.10447
phase	1		0.851		26		2	9		0.00019
treatment:phase	2		0.685		3		4	20		0.06674
gender:phase	1		0.043		0		2	9		0.82000
treatment:gender:phase	2		0.311		1		4	20		0.47215
hour	1		0.935		25		4	7		0.00030
treatment:hour	2		0.301		0		8	16		0.92952
gender:hour	1		0.293		1		4	7		0.60237
treatment:gender:hour	2		0.570		1		8	16		0.61319
phase:hour	1		0.550		0		8	3		0.83245
treatment:phase:hour	2		0.664		0		16	8		0.99144
gender:phase:hour	1		0.695		1		8	3		0.62021
treatment:gender:phase:hour	2		0.793		0		16	8		0.97237

Reading the Manova table as usual from bottom to top, only the within-subject main-effects of **phase** and **hour**, and possibly the between-subject **treatment** main effect seem to be important.

Further details of this example are given by <http://z.umn.edu/ffwmult>.

For completeness, here is the same problem treated using **lmer**:

```

OBrien.long <- reshape(OBrienKaiser,
  varying=c("pre.1", "pre.2", "pre.3", "pre.4", "pre.5", "post.1", "post.2", "post.3",
    "post.4", "post.5", "fup.1", "fup.2", "fup.3", "fup.4", "fup.5"),
  v.names="score", timevar="phase.hour", direction="long")
OBrien.long$phase <- ordered(c("pre", "post", "fup")[1 + ((OBrien.long$phase.hour - 1) %% 5)],
  levels=c("pre", "post", "fup"))
OBrien.long$hour <- ordered(1 + ((OBrien.long$phase.hour - 1) %% 5))
library(lme4)
l1 <- lmer(score ~ treatment*gender*phase*hour + (1|id), OBrien.long)
Anova(l1)

```

Analysis of Deviance Table (Type II Wald chisquare tests)

Response: score

	Chisq	Df	Pr(>Chisq)
treatment	9.26	2	0.0097
gender	2.56	1	0.1099
phase	98.14	2	< 2e-16
hour	62.28	4	9.6e-13
treatment:gender	5.71	2	0.0575
treatment:phase	46.09	4	2.4e-09
gender:phase	0.98	2	0.6135
treatment:hour	0.68	8	0.9996
gender:hour	1.50	4	0.8268
phase:hour	6.49	8	0.5921
treatment:gender:phase	5.99	4	0.2000
treatment:gender:hour	4.54	8	0.8050
treatment:phase:hour	3.67	16	0.9994
gender:phase:hour	3.89	8	0.8671
treatment:gender:phase:hour	8.29	16	0.9397

```

plot(Effect(c("treatment", "phase", "hour"), l1), multiline=TRUE, rows=1, cols=3, grid=TRUE)

```

