Chapter 5 in Everitt and Hothorn (2010) Analysis of Variance

Start R.

If you were not able to edit Rprofile.site, load the HSAUR2 and Rcmdr either using the commands: library(HSAUR2);library(Rcmdr)
or from the R Console using the menu Packages > Load package ... > select HSAUR2 and Rcmdr > Ok

We will be working with the R Commander menus.

Weight Gain in Rats

From the R Commander menus select Data > Data in packages > Read data set from an attached package... > double click on HSAUR2, select weightgain, and click ok.

To see a description, from the R commander menu select Data > Active data set > Help on active data set (if available)

Click View data set to view it.

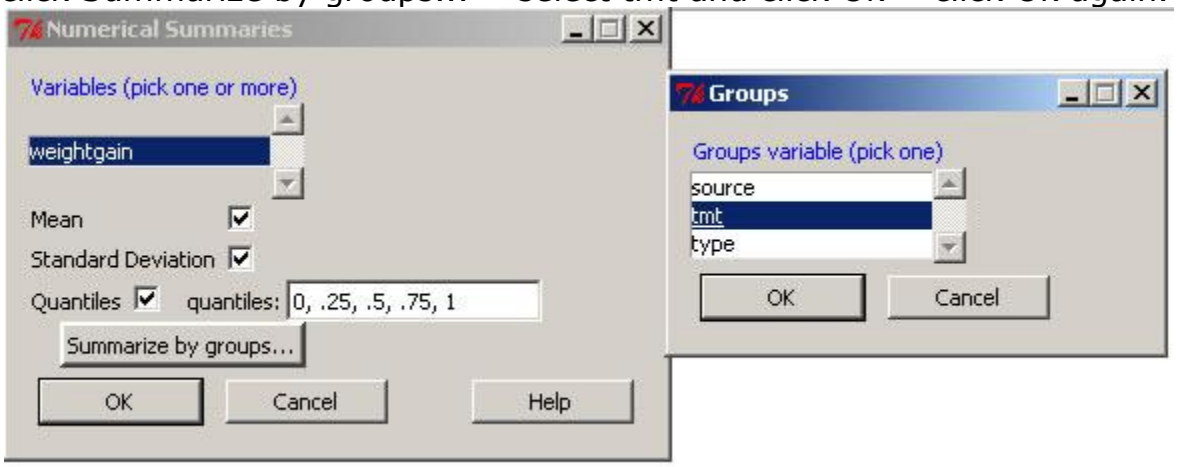
Create a new variable (tmt) combining both source and type by entering the following command in the Script Window and Submitting it.

weightgain$tmt=factor(paste(as.character(weightgain$source) , as.character(weightgain$type)))

You can also create a new variable from the menu Data > Manage variables in active data set > Compute new variable..., but it seems easier to just use a script command.

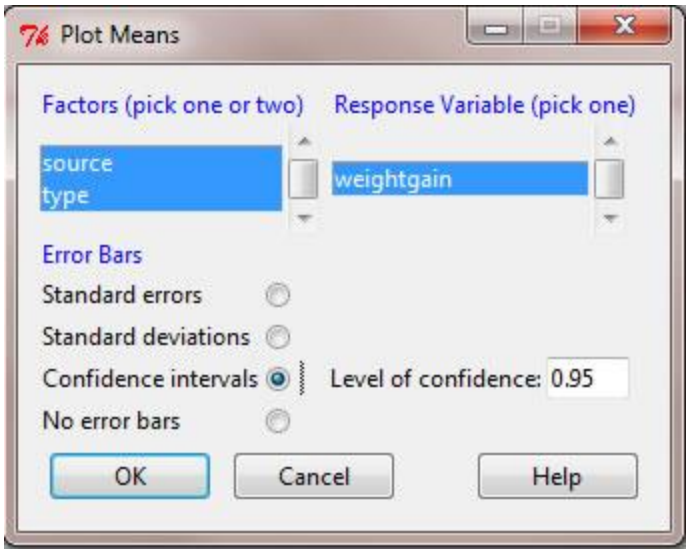
One of the quirks in R Commander is that you need to reset or refresh the active data set before it can recognize a new variable, so we will have to load another data set.
From the R Commander menus select Data > Active data set > Refresh data set.

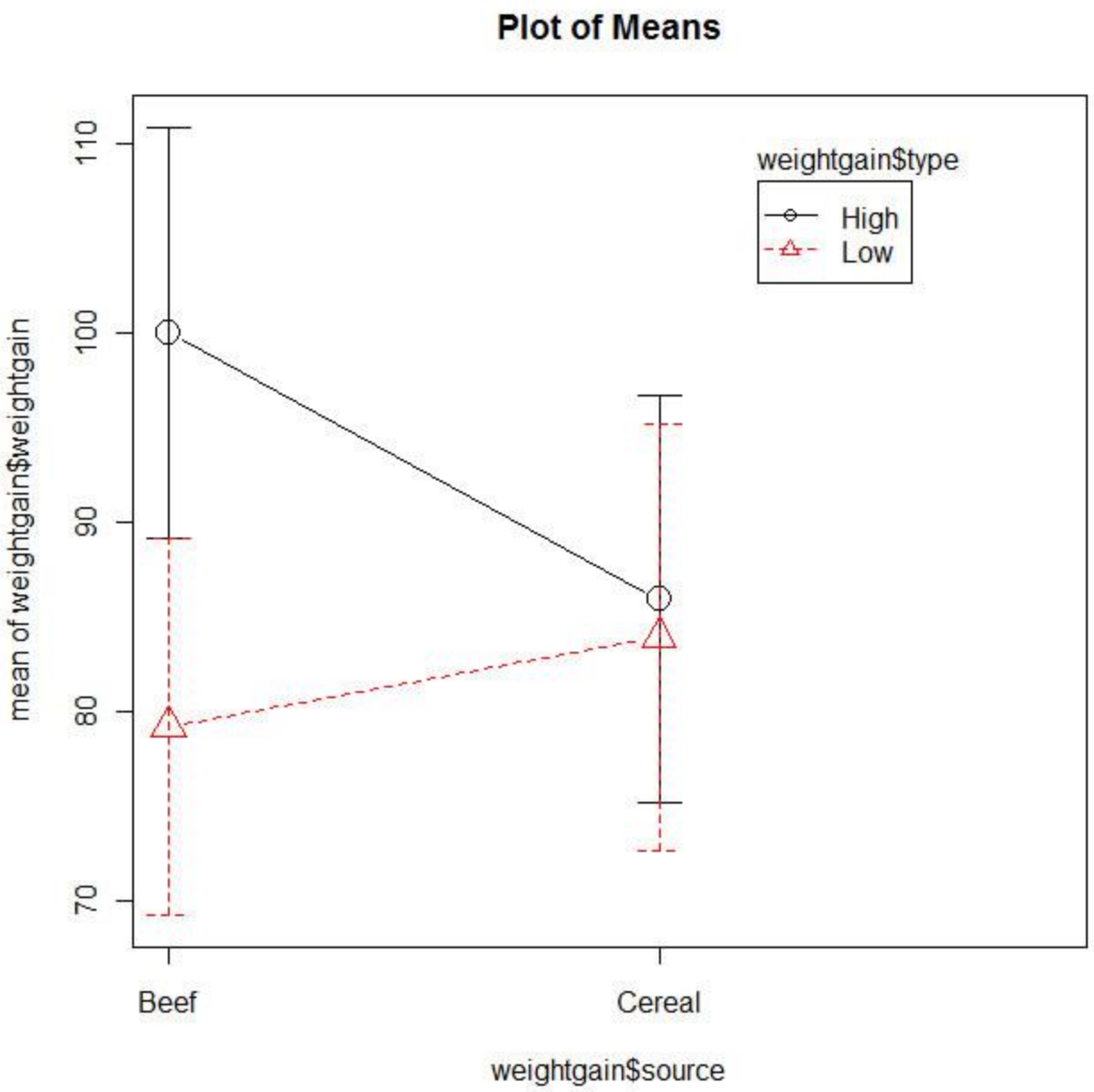
From the menu, select Statistics > Summaries > Numerical summaries... > click Summarize by groups... > select tmt and click Ok > click Ok again.



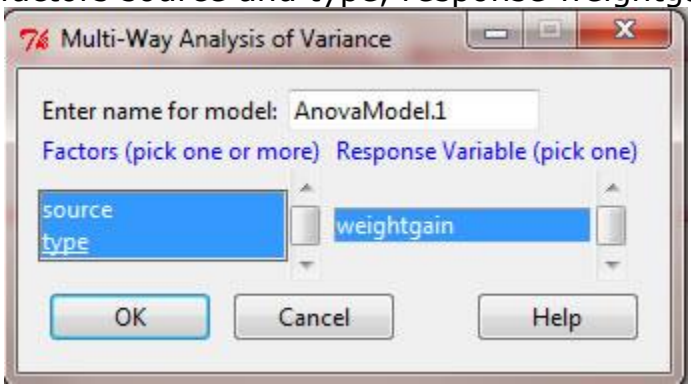
	mean	sd	0%	25%	50%	75%	100%	n
Beef High	100.0	15.13642	73	90.25	103.0	110.00	118	10
Beef Low	79.2	13.88684	51	73.00	82.0	90.00	95	10
Cereal High	85.9	15.02184	56	78.25	87.0	94.25	111	10
Cereal Low	83.9	15.70881	58	74.00	84.5	96.50	107	10

From the menu, select Graphs > Plot of means... > select factors: source and type, response: weightgain, confidence intervals and click Ok.


Plot Means
Factors (pick one or two)
Response Variable (pick one)
source
type
weightgain
Error Bars
Standard errors
Standard deviations
Confidence intervals
Level of confidence: 0.95
No error bars
OK
Cancel
Help


Plot of Means
weightgain$type
High
Low
mean of weightgain$weightgain
110
100
90
80
70
Beef
Cereal
weightgain$source

From the menu, select Statistics > Means > Multi-way Anova... > select factors source and type, response weightgain and click ok.




```
< Anova(AnovaModel.1)
Anova Table (Type II tests)

Response: weightgain
            Sum Sq Df F value  Pr(>F)
source       220.9  1  0.9879 0.32688
type        1299.6  1  5.8123 0.02114 *
source:type  883.6  1  3.9518 0.05447 .
Residuals   8049.4 36
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

< tapply(weightgain$weightgain, list(source=weightgain$source,
type=weightgain$type), mean, na.rm=TRUE) # means
        type
source    High  Low
  Beef   100.0 79.2
  Cereal  85.9 83.9

< tapply(weightgain$weightgain, list(source=weightgain$source,
type=weightgain$type), sd, na.rm=TRUE) # std. deviations
        type
source       High      Low
  Beef   15.13642 13.88684
  Cereal 15.02184 15.70881

< tapply(weightgain$weightgain, list(source=weightgain$source,
type=weightgain$type), function(x) sum(!is.na(x)))
+   # counts
        type
source   High Low
  Beef     10  10
  Cereal   10  10
```

From the menu, select Models > Summarize model.

```
Call:
lm(formula = weightgain ~ source * type, data = weightgain)

Residuals:
   Min     1Q Median     3Q    Max
-29.90  -9.90   2.05  10.85  25.10
```

```
Coefficients:
                            Estimate Std. Error t value Pr(>|t|)
(Intercept)                  100.000      4.729  21.148  < 2e-16 ***
source[T.Cereal]             -14.100      6.687  -2.109  0.04201 *
type[T.Low]                  -20.800      6.687  -3.110  0.00364 **
source[T.Cereal]:type[T.Low]  18.800      9.457   1.988  0.05447 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 14.95 on 36 degrees of freedom
Multiple R-squared:  0.23,    Adjusted R-squared: 0.1658
F-statistic: 3.584 on 3 and 36 DF,  p-value: 0.02297
```

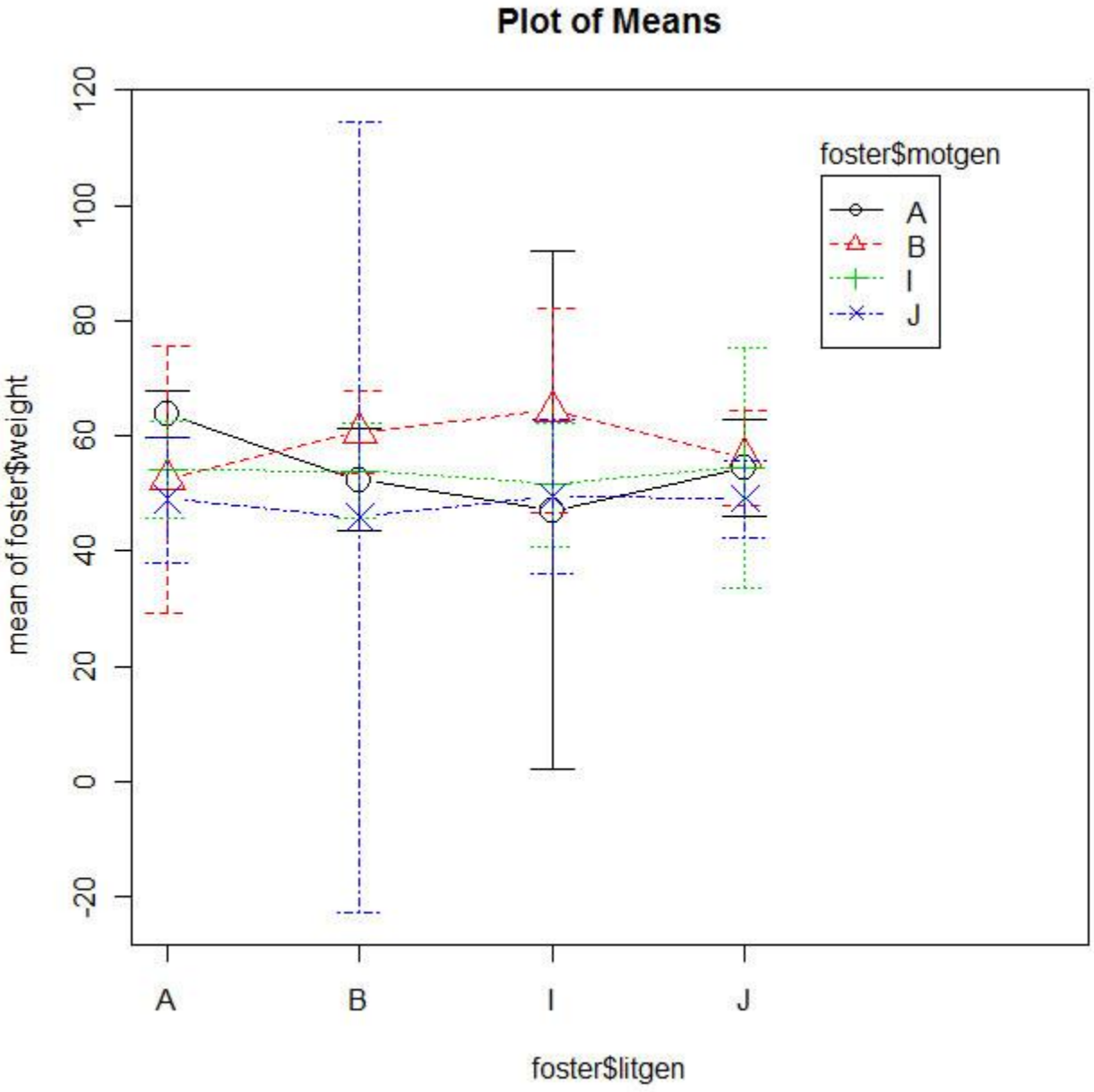
Foster Feeding of Rats of Different Genotype

From the R Commander menus select Data > Data in packages > Read data set from an attached package... >
Double click on HSAUR2 and select foster, then click OK.

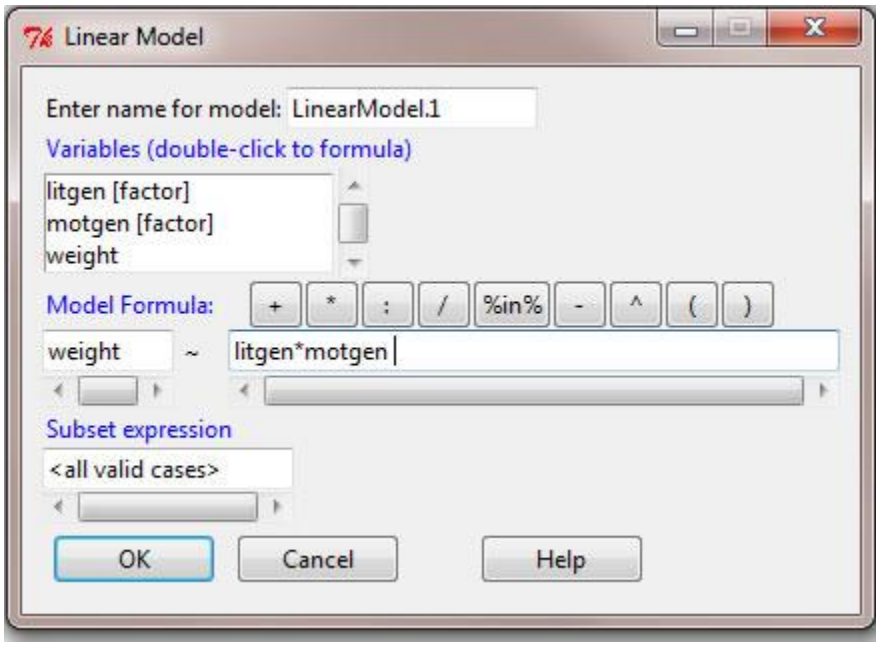
To see a description, from the R commander menu select Data > Active data set > Help on active data set (if available)

Click View data set to view it.

From the menu, select Graphs > Plot of means... > select factors: litgen and motgen, response: weight, confidence intervals and click Ok.



We could use the multi-way ANOVA menu as above, but a more flexible approach is:
From the menu, select Statistics > Fit models > Linear model... > click on weight to add it to the left side of the model equation, click on litgen, then *, then morgen to add them to the right side of the equation, and click Ok.



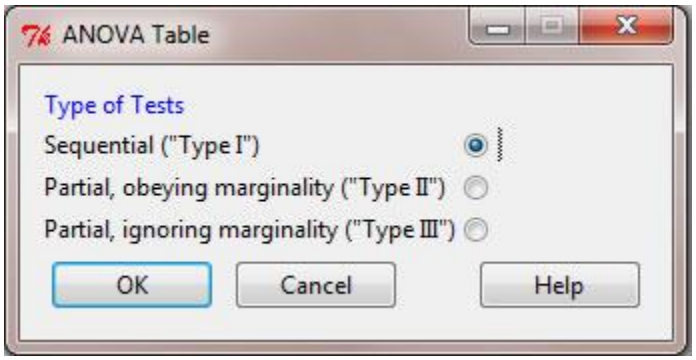

```
Call:
lm(formula = weight ~ litgen * motgen, data = foster)

Residuals:
     Min       1Q   Median       3Q      Max
-11.9000  -4.6250   0.5667   4.1000  20.9000

Coefficients:
                        Estimate Std. Error t value Pr(>|t|)
(Intercept)               63.680      3.294  19.334  < 2e-16 ***
litgen[T.B]              -11.355      4.940  -2.298 0.026241 *
litgen[T.I]              -16.580      5.378  -3.083 0.003498 **
litgen[T.J]               -9.330      4.940  -1.888 0.065417 .
motgen[T.B]              -11.280      5.378  -2.097 0.041622 *
motgen[T.I]               -9.555      4.940  -1.934 0.059413 .
motgen[T.J]              -14.720      4.658  -3.160 0.002818 **
litgen[T.B]:motgen[T.B]   19.595      7.303   2.683 0.010169 *
litgen[T.I]:motgen[T.B]   28.547      8.068   3.538 0.000948 ***
litgen[T.J]:motgen[T.B]   13.030      7.783   1.674 0.101018
litgen[T.B]:motgen[T.I]   11.155      7.178   1.554 0.127194
litgen[T.I]:motgen[T.I]   14.055      7.303   1.925 0.060628 .
litgen[T.J]:motgen[T.I]    9.738      7.487   1.301 0.199956
litgen[T.B]:motgen[T.J]    8.295      7.898   1.050 0.299196
litgen[T.I]:motgen[T.J]   17.053      7.606   2.242 0.029937 *
litgen[T.J]:motgen[T.J]    9.430      6.790   1.389 0.171728
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7.365 on 45 degrees of freedom
Multiple R-squared: 0.4047,    Adjusted R-squared: 0.2063
F-statistic: 2.039 on 15 and 45 DF,  p-value: 0.03333
```

From the menu, select Models > Hypothesis tests > Anova table > Type I



Repeat, selecting Type II and then Type III tests.
Fit another model, changing the order, putting mortgen before litgen, and printout the three types of tests.

```
Analysis of Variance Table
Response: weight
             Df  Sum Sq Mean Sq F value   Pr(>F)
litgen        3   60.16  20.052  0.3697 0.775221
motgen        3  775.08 258.360  4.7632 0.005736 **
litgen:motgen 9  824.07  91.564  1.6881 0.120053
Residuals    45 2440.82  54.240
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Anova Table (Type II tests)
Response: weight
              Sum Sq Df F value   Pr(>F)
litgen         63.63  3  0.3911 0.760004
motgen        775.08  3  4.7632 0.005736 **
litgen:motgen 824.07  9  1.6881 0.120053
Residuals    2440.82 45
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Anova Table (Type III tests)
Response: weight
              Sum Sq Df  F value  Pr(>F)
(Intercept)  20275.7  1 373.8122 < 2e-16 ***
litgen         591.7  3   3.6362 0.01968 *
motgen         582.3  3   3.5782 0.02099 *
litgen:motgen  824.1  9   1.6881 0.12005
Residuals     2440.8 45
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Analysis of Variance Table
Response: weight
             Df  Sum Sq Mean Sq F value   Pr(>F)
motgen        3  771.61 257.202  4.7419 0.005869 **
litgen        3   63.63  21.211  0.3911 0.760004
```

```
motgen:litgen  9  824.07  91.564  1.6881 0.120053
Residuals     45 2440.82  54.240
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Anova Table (Type II tests)
Response: weight
              Sum Sq Df F value   Pr(>F)
motgen         775.08  3  4.7632 0.005736 **
litgen          63.63  3  0.3911 0.760004
motgen:litgen  824.07  9  1.6881 0.120053
Residuals     2440.82 45
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Anova Table (Type III tests)
Response: weight
              Sum Sq Df  F value  Pr(>F)
(Intercept)   20275.7  1 373.8122 < 2e-16 ***
motgen          582.3  3   3.5782 0.02099 *
litgen          591.7  3   3.6362 0.01968 *
motgen:litgen   824.1  9   1.6881 0.12005
Residuals      2440.8 45
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

What is going on here?
All the treatment combinations do not have the same number of observations, so it is not possible to partition the variation into non-overlapping or orthogonal sums or squares.
Type I tests use sequential tests of litgen, motgen adjusted for litgen, and interaction.
Type II tests test each main effect adjusted for the other main effects such as litgen adjusted for motgen, motgen adjusted for litgen, and interaction.
Note that these definitions of types of tests differs from the types used by SAS.
The order does not make any difference for type II and III tests.
The documentation says:
"The designations 'type-II' and 'type-III' are borrowed from SAS, but the definitions used here do not correspond precisely to those employed by SAS. Type-II tests are calculated according to the principle of marginality, testing each term after all others, except ignoring the term's higher-order relatives; so-called type-III tests violate marginality, testing each term in the model after all of the others. This definition of Type-II tests corresponds to the tests produced by SAS for analysis-of-variance models, where all of the predictors are factors, but not more generally (i.e., when there are quantitative predictors). Be very careful in formulating the model for type-III tests, or the hypotheses tested will not make sense." In almost all cases, I suggest type II tests.

Effect	Sum of Squares
litgen	60.16
litgen adjusted for morgen	63.63
morgen	771.61
morgen adjusted for litgen	775.08
interaction adjusted for litgen and morgen	824.07
intercept adjusted for everything else	20275.7
litgen adjusted for everything else	591.7
motgen adjusted for everything else	582.3
interaction adjusted for everything else	824.1

Multiple comparisons are only available using the aov(Analysis of Variance) command.
TukeyHSD will not work if RcmdrPlugin.HH is loaded.
Enter the following commands into the Script Window and Submit them.

*m1=aov(weight ~ litgen*motgen, data=foster)*
TukeyHSD(m1,"motgen")

```
Tukey multiple comparisons of means
    95% family-wise confidence level
Fit: aov(formula = weight ~ litgen * motgen, data = foster)
$motgen
         diff        lwr        upr     p adj
B-A  3.330369  -3.859729 10.5204672 0.6078581
I-A -1.895574  -8.841869  5.0507207 0.8853702
J-A -6.566168 -13.627285  0.4949498 0.0767540
I-B -5.225943 -12.416041  1.9641552 0.2266493
J-B -9.896537 -17.197624 -2.5954489 0.0040509
J-I -4.670593 -11.731711  2.3905240 0.3035490
```

Water Hardness and Mortality

From the R Commander menus select Data > Data in packages > Read data set from an attached package... >
Double click on HSAUR2 and select water, then click OK.

To see a description, from the R commander menu select Data > Active data set > Help on active data set (if available)

Click View data set to view it.

Multivariate analysis of variance is not available from the menu, so enter the following commands in the Script Window and Submit them.

m1=manova(cbind(hardness,mortality) ~ location, data=water)
summary(m1, test="Hotelling-Lawley")

```
Df Hotelling-Lawley approx F num Df den Df    Pr(>F)
location   1          0.90021   26.106      2     58 8.217e-09 ***
Residuals 59
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The test shows a significant difference in hardness, mortality or some combination of them.

From the menu, select Statistics > Summaries > Numerical summaries... > select variables hardness and mortality, click Summarize by groups > select location and click Ok > click Ok again.

```
Variable: hardness
          mean       sd 0%   25%  50%   75% 100%  n
North 30.40000 26.13449  6 12.50 17.0 44.00   94 35
South 69.76923 40.36068  5 40.25 75.5 99.75  138 26

Variable: mortality
          mean       sd   0%     25%  50%     75% 100%  n
North 1633.600 136.9369 1378 1557.50 1637 1718.00 1987 35
South 1376.808 140.2692 1096 1259.25 1364 1485.75 1627 26
```

Male Egyptian Skulls

From the R Commander menus select Data > Data in packages > Read data set from an attached package... >
Double click on HSAUR2 and select skulls, then click OK.

To see a description, from the R commander menu select Data > Active data set > Help on active data set (if available)

Click View data set to view it.

From the menu, select Statistics > summaries > Numerical summaries... > select variables bh, bl, mb and nh, and click Summarize by groups > select epoch and click Ok > click Ok again.

```
Variable: bh
            mean       sd  0%    25%   50%    75% 100%  n
c4000BC 133.6000 4.469051 121 131.25 134.0 136.00  143 30
c3300BC 132.7000 4.647209 124 129.25 133.0 136.00  145 30
c1850BC 133.8000 4.978575 123 131.00 133.5 137.00  145 30
```

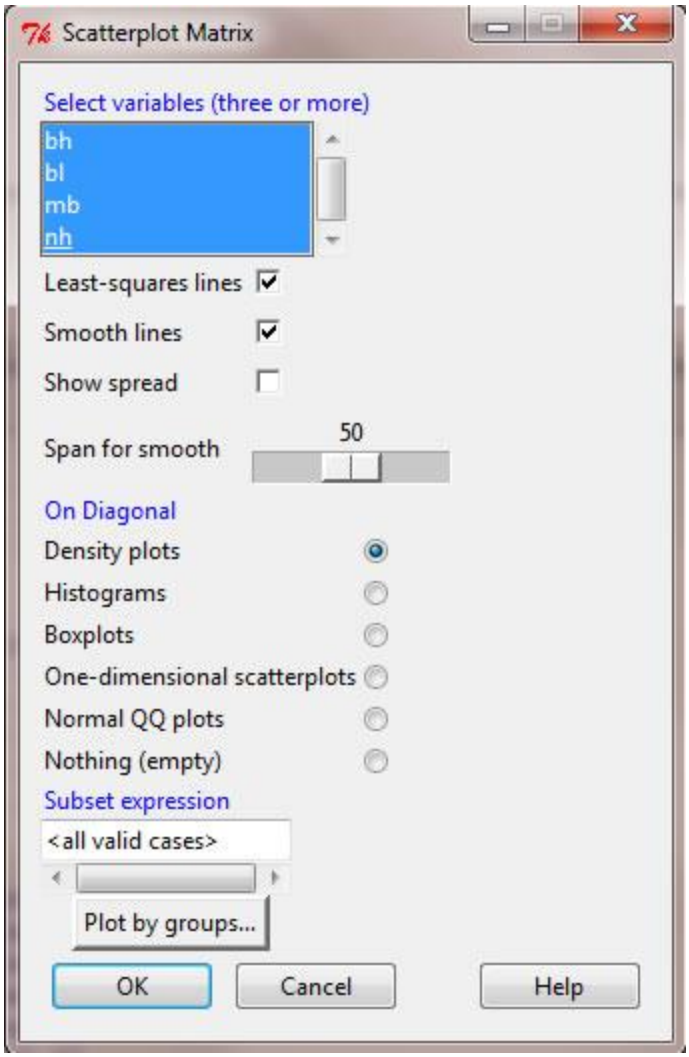
```
c200BC  132.3000 5.133729 120 130.00 132.0 135.75  142 30
cAD150  130.3333 4.971181 120 126.00 130.0 135.00  138 30

Variable: bl
            mean       sd 0%   25%   50%    75% 100%  n
c4000BC 99.16667 5.884423 89 95.00 100.0 102.75  114 30
c3300BC 99.06667 4.346488 90 97.00  98.5 101.75  107 30
c1850BC 96.03333 4.552251 87 92.25  96.0  99.75  106 30
c200BC  94.53333 4.591847 86 91.25  94.5  97.75  107 30
cAD150  93.50000 5.056576 81 91.00  94.0  97.00  103 30

Variable: mb
            mean       sd  0%    25% 50%    75% 100%  n
c4000BC 131.3667 5.129249 119 128.00 131 134.75  141 30
c3300BC 132.3667 4.810071 123 130.00 132 134.75  148 30
c1850BC 134.4667 3.481313 126 132.25 136 137.00  140 30
c200BC  135.5000 3.919448 129 132.25 135 138.75  144 30
cAD150  136.1667 5.350368 126 132.25 137 139.00  147 30

Variable: nh
            mean       sd 0%   25%  50%   75% 100%  n
c4000BC 50.53333 2.763473 44 49.00 50.0 53.00   56 30
c3300BC 50.23333 2.955805 45 48.00 50.5 52.75   56 30
c1850BC 50.56667 3.549486 45 48.25 50.0 52.75   60 30
c200BC  51.96667 2.822121 46 50.25 52.0 53.75   60 30
cAD150  51.36667 3.718392 44 48.25 52.0 54.00   58 30
```

From the menu, select Graphs > Scatterplot matrix... > select variables bh, bl, mb and nh, and click Ok.



Multivariate analysis of variance is not available from the menu, so enter the following commands in the Script Window and Submit them.

m1=manova(cbind(mb,bh,bl,nh) ~ epoch, data=skulls)
summary(m1,test="Pillai")
summary(m1,test="Wilks")
summary(m1,test="Hotelling-Lawley")
summary(m1,test="Roy")
summary.aov(m1)

```
          Df  Pillai approx F num Df den Df    Pr(>F)
epoch      4 0.35331    3.512     16    580 4.675e-06 ***
Residuals 145
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

          Df   Wilks approx F num Df den Df   Pr(>F)
epoch      4 0.66359   3.9009     16 434.45 7.01e-07 ***
Residuals 145
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

           Df Hotelling-Lawley approx F num Df den Df    Pr(>F)
epoch       4          0.48182    4.231     16    562 8.278e-08 ***
Residuals 145
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

           Df    Roy approx F num Df den Df    Pr(>F)
epoch       4 0.4251   15.410      4    145 1.588e-10 ***
Residuals 145
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

 Response mb :
             Df  Sum Sq Mean Sq F value    Pr(>F)
epoch         4  502.83 125.707  5.9546 0.0001826 ***
Residuals   145 3061.07  21.111
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

 Response bh :
             Df Sum Sq Mean Sq F value  Pr(>F)
epoch         4  229.9  57.477  2.4474 0.04897 *
Residuals   145 3405.3  23.485
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

 Response bl :
             Df Sum Sq Mean Sq F value    Pr(>F)
epoch         4  803.3 200.823  8.3057 4.636e-06 ***
Residuals   145 3506.0  24.179
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

 Response nh :
             Df Sum Sq Mean Sq F value Pr(>F)
epoch         4   61.2  15.300   1.507 0.2032
Residuals   145 1472.1  10.153
```

To look at pairwise multivariate tests, enter the following commands in the Script Window and Submit them.

summary(manova(cbind(mb,bh,bl,nh) ~ epoch, data=skulls, subset= epoch %in% c("c4000BC", "c3300BC")))
summary(manova(cbind(mb,bh,bl,nh) ~ epoch, data=skulls, subset= epoch %in% c("c4000BC", "c1850BC")))
summary(manova(cbind(mb,bh,bl,nh) ~ epoch, data=skulls, subset= epoch %in% c("c4000BC", "c200BC")))
summary(manova(cbind(mb,bh,bl,nh) ~ epoch, data=skulls, subset= epoch %in% c("c4000BC", "cAD150")))

```
> summary(manova(cbind(mb,bh,bl,nh) ~ epoch, data=skulls, subset= epoch %in%
c("c4000BC", "c3300BC")))
          Df   Pillai approx F num Df den Df Pr(>F)
```

```
epoch      1 0.027674  0.39135      4     55  0.814
Residuals 58

> summary(manova(cbind(mb,bh,bl,nh) ~ epoch, data=skulls, subset= epoch %in%
c("c4000BC", "c1850BC")))
          Df  Pillai approx F num Df den Df  Pr(>F)
epoch      1 0.18757   3.1744      4     55 0.02035 *
Residuals 58
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> summary(manova(cbind(mb,bh,bl,nh) ~ epoch, data=skulls, subset= epoch %in%
c("c4000BC", "c200BC")))
          Df  Pillai approx F num Df den Df    Pr(>F)
epoch      1 0.30297   5.9766      4     55 0.0004564 ***
Residuals 58
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> summary(manova(cbind(mb,bh,bl,nh) ~ epoch, data=skulls, subset= epoch %in%
c("c4000BC", "cAD150")))
          Df  Pillai approx F num Df den Df    Pr(>F)
epoch      1 0.36182   7.7956      4     55 4.736e-05 ***
Residuals 58
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> summary(manova(cbind(mb,bh,bl,nh) ~ epoch, data=skulls, subset= epoch %in%
c("c4000BC", "c3300BC")))
          Df   Pillai approx F num Df den Df Pr(>F)
epoch      1 0.027674  0.39135      4     55  0.814
Residuals 58
```