BIM 105
Probability and Statistics for Biomedical Engineers

David M. Rocke

Department of Biomedical Engineering

Factorial Experiments

- Regression type models can work with quantitative responses (and also qualitative ones) and predictors that are either quantitative or qualitative.
- Sometimes, the predictors and the response on a given unit are observed together as when we measure characteristics of a patient and the outcome of treatment.
- Sometimes the treatments are chosen by the experimenter and then the response is observed, as when some patients get a new treatment and others standard of care.

One-Factor Experiments

- We apply two or more treatments to experimental units, choosing the treatment for each unit by using random numbers.
- There needs to be in this case at least two units with each treatment or there is no basis for comparison.
- With exactly two treatments, this can be analyzed by the two-sample t-test.
- If there are more than two treatments, we use one-way ANOVA.
- Ideally, the number of units on each treatment is the same, but this cannot always be assured. When it is, the design is called *balanced*.

Red Cell Folate Study

- Red cell folate is a measure of folic acid (vitamin B9). It can be disrupted by anesthesia with nitrous oxide (N_2O).
- This study compared operations under three conditions:
  - N_2O (50%) +O_2 (50%) for 24 hours continuously up to and including the operation.
  - N_2O (50%) +O_2 (50%) only during the operation.
  - O_2 at 30%-50% before the operation, but no N_2O before the operation.
- There were 22 patients allocated 8/9/5 to the three treatments (unbalanced).
- The MATLAB function `fitlm` will be able to tell that ventilation is a factor because it does not consist of numbers. If it does, you have to tell it which variables are categorical. You can use `nominal` to convert numbers to categories of the Name-Value pair `'CategoricalVars'` in `fitlm`.

```
>> folatelm = fitlm(folate,'folate~ventilation')

Estimated Coefficients:
                                Estimate      SE       tStat       pValue
                                ________    ______    _______    __________

    (Intercept)                  316.62     16.164     19.588    4.6492e-14
    ventilation_"N2O+O2--op"    -60.181     22.216    -2.7089       0.01392
    ventilation_"O2--24h"       -38.625     26.064    -1.4819       0.15476


Number of observations: 22, Error degrees of freedom: 19
Root Mean Squared Error: 45.7
R-squared: 0.281,  Adjusted R-Squared 0.205
F-statistic vs. constant model: 3.71, p-value = 0.0436

These coefficients are comparisons between each of the two listed
treatments and the omitted comparison level, which is "N2O+O2--24h". This
is not a test of whether the factor as a whole is important. The F-test is
a valid test of the factor as a whole and is MS(ventilation)/MS(error) as
given on the next slide in more detail.
```

```
>> anova(folatelm)

                SumSq    DF    MeanSq    F         pValue
    ventilation 15516     2    7757.9    3.7113    0.043589
    Error       39716    19    2090.3
```

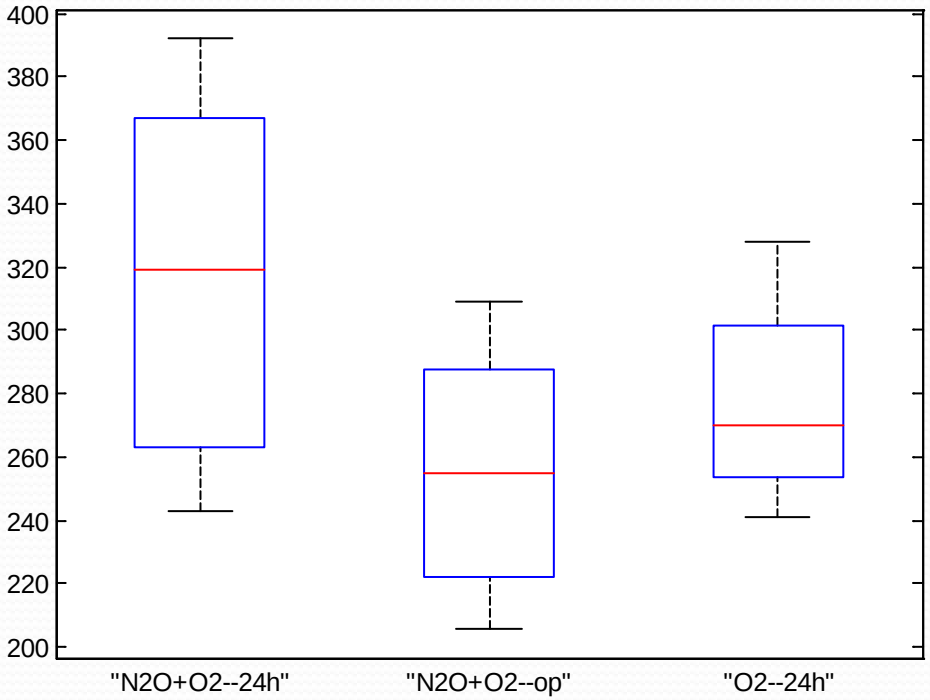
This test shows that ventilation has an effect on folate levels which is statistically significant at the 5% level. Note that the F ratio MS(ventilation)/MS(error) has df 2 and 19.

But which treatments are better?

This can be analyzed graphically with

```
>> boxplot(folate.folate,folate.ventilation)
```

And statistically with a more complex procedure for *multiple comparisons*, that is comparing each of the three procedures with the other two.


400
380
360
340
320
300
280
260
240
220
200
"N2O+O2--24h"
"N2O+O2--op"
"O2--24h"

Multiple Comparisons

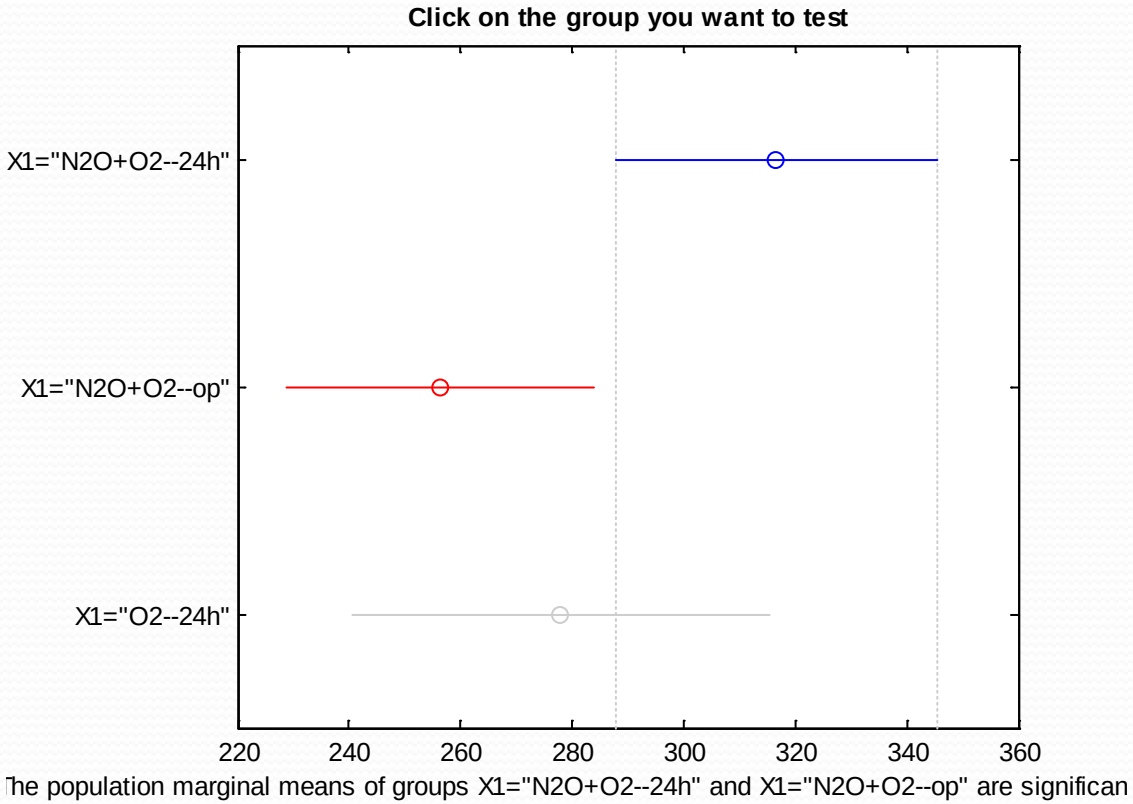
```
>> vent2 = num2cell(folate.ventilation,1)
>> [p,table,stats] = anovan(folate.folate,vent2)
p =
    0.0436
table =

    'Source'    'Sum Sq.'       'd.f.'    'Singular?'    'Mean Sq.'      'F'          'Prob>F'
    'X1'        [1.5516e+04]    [   2]    [         0]    [7.7579e+03]    [3.7113]    [0.0436]
    'Error'     [3.9716e+04]    [  19]    [         0]    [2.0903e+03]          []          []
    'Total'     [5.5232e+04]    [  21]    [         0]                []          []          []

stats =

         source: 'anovan'
          resid: [22x1 double]
         coeffs: [4x1 double]
            Rtr: [3x3 double]
      rowbasis: [3x4 double]
            dfe: 19
            mse: 2.0903e+03
    nullproject: [4x3 double]
          terms: 1
         nlevels: 3
     continuous: 0
         vmeans: 0
       termcols: [2x1 double]
     coeffnames: {4x1 cell}
           vars: [4x1 double]
       varnames: {'X1'}
       grpnames: {{3x1 cell}}
```

The `anovan` command does one or multiway ANOVA, but it needs the grouping variable to be in a cell array. The resulting ANOVA table and other statistics are the same as using `fitlm` and then `anova`, but we need the stats structure for the multiple comparisons.

```
>> multcompare(stats)

    1.0000    2.0000    3.7421   60.1806  116.6190    0.0355
    1.0000    3.0000  -27.5904   38.6250  104.8404    0.3215
    2.0000    3.0000  -86.3406  -21.5556   43.2295    0.6802
```

The first two columns show the groups being compared, the fourth column the difference, and the flanking third and fifth columns show a 95% CI adjusted for multiple comparisons. If this 95% CI does not include 0, then the groups are significantly different. In this case, this is only 1 vs. 2. The last column shows the associated p-value.

```
>> celldisp(getfield(stats,'grpnames'))

ans{1}{1} =
"N2O+O2--24h"

ans{1}{2} =
"N2O+O2--op"

ans{1}{3} =
"O2--24h"
```


Click on the group you want to test
X1="N2O+O2--24h"
X1="N2O+O2--op"
X1="O2--24h"
220
240
260
280
300
320
340
360
The population marginal means of groups X1="N2O+O2--24h" and X1="N2O+O2--op" are significan

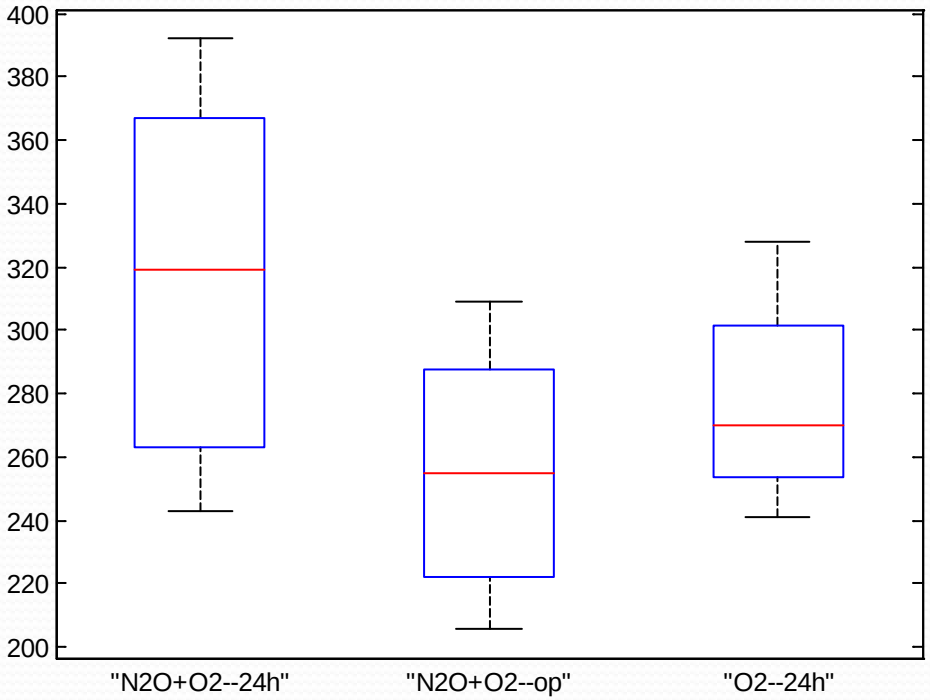
Multiple Comparisons

- `multcompare` can use different comparisons metrics.
- The default is the Tukey HSD or honest significant difference which is based on the studentized range, and attempts to declare any one or more differences significant only 5% of the time if all of the true group means are actually the same.
- An alternative is the least significant difference (lsd) which should only be used if the F-test is significant (protected lsd), but gives narrower intervals.

```
>> multcompare(stats)

ans =

    1.0000    2.0000    3.7421   60.1806  116.6190    0.0355
    1.0000    3.0000  -27.5904   38.6250  104.8404    0.3215
    2.0000    3.0000  -86.3406  -21.5556   43.2295    0.6802

>> multcompare(stats,'ctype','lsd')

ans =

    1.0000    2.0000   13.6821   60.1806  106.6791    0.0139
    1.0000    3.0000  -15.9285   38.6250   93.1785    0.1548
    2.0000    3.0000  -74.9306  -21.5556   31.8195    0.4085

The intervals are narrower, but the results are unchanged in this case.
```

Conclusions

- The form of ventilation appears to affect the folate level.
- The definitive conclusion of the study is that "N2O+O2--24h“ is better than "N2O+O2--op“, with "O2--24h“ in the middle and not definitively different from either one.
- A main assumption of ANOVA is that the groups have the same variance, and the boxplot does not strongly challenge that assumption.


400
380
360
340
320
300
280
260
240
220
200
"N2O+O2--24h"
"N2O+O2--op"
"O2--24h"

Two Factor Experiments

- In a two factor experiment, there are two sets of treatments and each experimental unit gets one treatment from each set.
- We can evaluate the effects of each factor separately, and also the interaction.

Coking Data

- This experiment is on time to coking (making coke from coal) in an experiment in which oven width and temperature were varied.
  - width = a factor with levels 4, 8, and 12 giving the oven width in inches.
  - temp = a factor with levels 1600 and 1900, giving the oven temperature in degrees Fahrenheit.
  - time = a numeric variable, time to coking
- This is a balanced two-way experiment with three replicates under each set of conditions (3x2 = 6 conditions), so n = 18.

Main Effects and Interactions

- The main effect of width is the change in time as width changes, averaged over temperatures. For example,

```
time(width = 12) – time(width = 8),
```

  which are differences of simple averages.
- The main effect of temperature is the change in time as temperature changes, averaged over widths.
- The interaction of width and time can be thought of as the change in the time between high and low temperature as width changes. For example,

```
 [time(width=12, temp=1900) – time(width=12, temp=1600)]
–[time(width=8, temp=1900) – time(width=8, temp=1600)]
```

  which are differences of differences.

```
>> cokinglm = fitlm(coking,'time~width*temp','CategoricalVars',[1,2])
Linear regression model:
    time ~ 1 + width*temp
Estimated Coefficients:
```

	Estimate	SE	tStat	pValue
(Intercept)	3.0667	0.30399	10.088	3.2569e-07
width_8	4.1	0.4299	9.5371	5.962e-07
width_12	7.7333	0.4299	17.989	4.7896e-10
temp_1900	-0.76667	0.4299	-1.7834	0.099819
width_8:temp_1900	-0.86667	0.60797	-1.4255	0.1795
width_12:temp_1900	-2.7	0.60797	-4.441	0.00080545

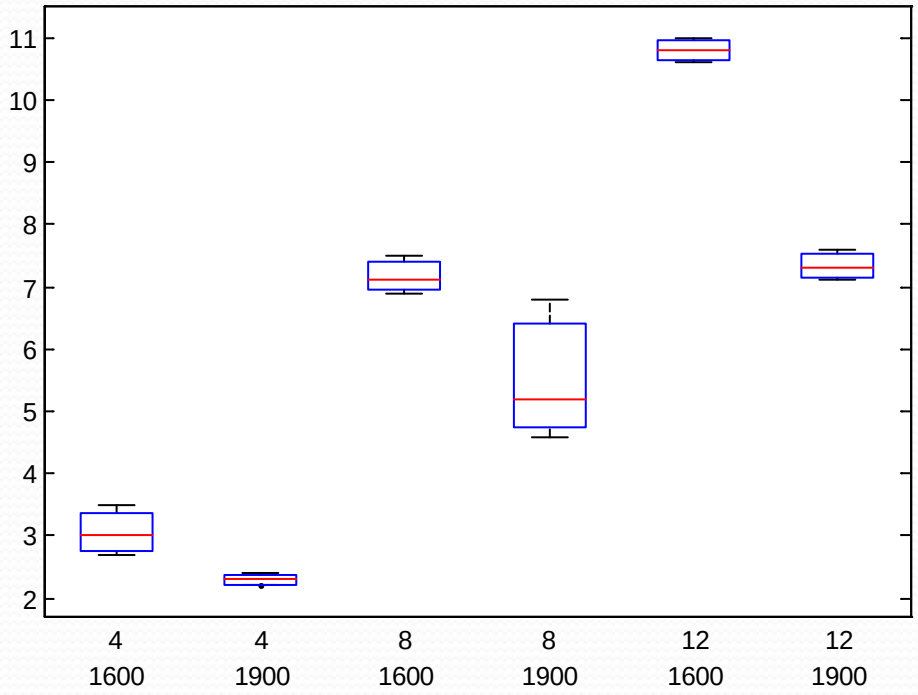
```
Number of observations: 18, Error degrees of freedom: 12
Root Mean Squared Error: 0.527
R-squared: 0.978,  Adjusted R-Squared 0.968
F-statistic vs. constant model: 105, p-value = 1.74e-09
```

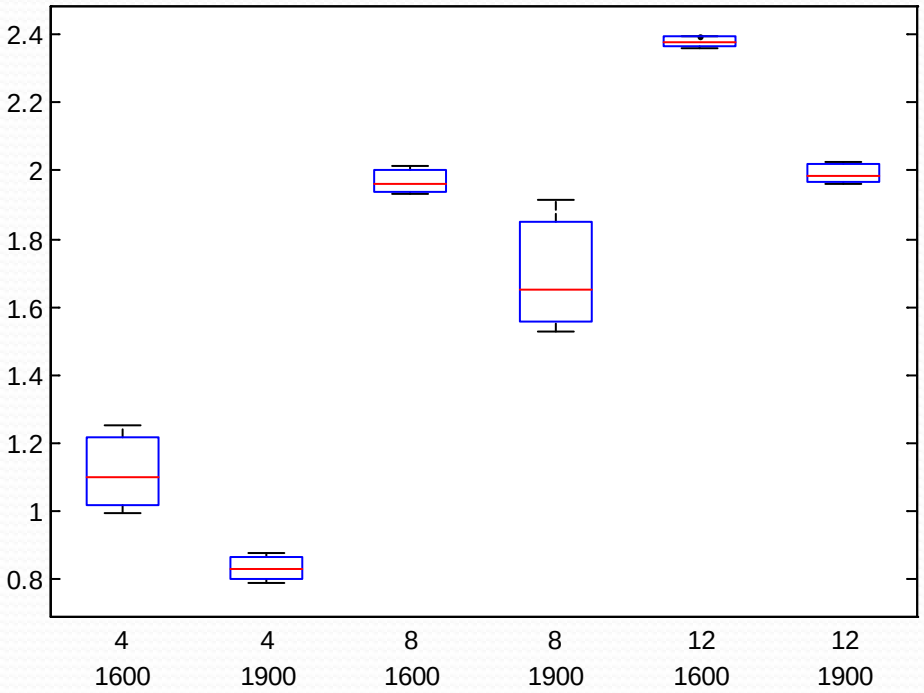
```
>> anova(cokinglm)
```

	SumSq	DF	MeanSq	F	pValue
width	123.14	2	61.572	222.1	3.3123e-10
temp	17.209	1	17.209	62.076	4.3942e-06
width:temp	5.7011	2	2.8506	10.283	0.0025036
Error	3.3267	12	0.27722		

The interaction term is significant, which means that the effect of temperature is different at different levels of width. This makes it hard to interpret the main effects of width and temperature.

```
>> boxplot(coking.time,[coking.width coking.temp])
>> boxplot(log(coking.time),[coking.width coking.temp])
```


11
10
9
8
7
6
5
4
3
2
4
1600
4
1900
8
1600
8
1900
12
1600
12
1900


2.4
2.2
2
1.8
1.6
1.4
1.2
1
0.8
4
1600
4
1900
8
1600
8
1900
12
1600
12
1900

```
>> ltime = log(coking.time)
>> coking2 = [coking table(ltime)]
>> coking2.time = []

coking2 =

    width    temp     ltime
    _____    ____    _______

     4       1600     1.2528
     4       1600     1.0986
     4       1600    0.99325
    ..........................

>> coking2lm = fitlm(coking2,'ltime~width*temp','CategoricalVars',[1,2])
>> anova(coking2lm)

                SumSq      DF     MeanSq        F         pValue
               ________    __    ________    _______    __________

    width         4.6648     2      2.3324     224.99    3.0714e-10
    temp         0.44332     1     0.44332     42.764    2.7718e-05
    width:temp  0.012252     2    0.006126    0.59094       0.56915
    Error         0.1244    12    0.010367
```

On the log scale, only the main effects are significant, which makes the interpretation much easier.

Parameterizations for One-Way ANOVA

- If we have one factor with k levels, predictions need k parameters, one for each group.
  - We can make this k group means or
  - We can make this a grand mean and differences from the grand mean or
  - We can make this the differences from a base level of the factor, usually the first one.
- All are equally valid, but interpretation of the coefficients is different.

Suppose we have one factor with k levels. Group i, $1 \le i \le k$ has n_i replicates and $n = n_1 + n_2 + \cdots + n_k$ Let the observation y_{ij} be the j^{th} replicate from group i, $1 \le i \le k, 1 \le j \le n_i$

Cell Means Model (k coefficients, k df)

$y_{ij} = \mu_i + \epsilon_{ij}$

Each of the k μ_i values is the true population mean of group i, estimated by $\overline{y}_{i\cdot}$

Variation Around the Mean Model ($k+1$ coefficients, k df)

$y_{ij} = \mu + \alpha_i + \epsilon_{ij}$

where μ is the grand mean, which is the mean of the expected values of the responses, $\mu = n^{-1}\sum_i n_i \mu_i$ estimated by $\overline{y}_{\cdot\cdot}$ and α_i is the difference between μ_i and μ, estimated by $\overline{y}_{i\cdot} - \overline{y}_{\cdot\cdot}$

$\alpha_i = \mu_i - \mu$ so $\sum \alpha_i = 0$. The set of $\{\alpha_i\}$ has k elements, but they add to 0, so only $k-1$ df

Omitted Levels Model, as used by MATLAB ($k+1$ coefficients, one of which is 0, so k df)

$y_{ij} = \mu_0 + \alpha_i + \epsilon_{ij}$

where μ_0 is the population mean of group 1 estimated by $\overline{y}_{1\cdot}$, $\alpha_1 = 0$, and $\alpha_i = \mu_i - \mu_1$

The coefficient estimates $\hat{\alpha}_i$ are estimates of the difference between group i and group 1.

$\hat{\mu}_0 = \overline{y}_{1\cdot}$ and $\hat{\alpha}_i = \overline{y}_{i\cdot} - \overline{y}_{1\cdot}$

Two Factor Experiments

- In a two factor experiment, there are two sets of treatments and each experimental unit gets one treatment from each set.
- In the tensile strength data, there are treatment and control (one factor) evaluated at two weeks and four weeks (a second factor).
- We can evaluate the effects of each factor separately, and also the interaction.

Parameterizations for Two-Way ANOVA

- We have the same choices for a two-way ANOVA as for a one-way ANOVA
  - Cell means
  - Grand mean and differences from the grand mean
  - Differences from base levels of the factors, usually the first level of each factor.
- All are equally valid, but interpretation of the coefficients is different.

Suppose we have one factor A with a levels and one factor B with b levels.

The group with A at level i and B at level j, $1 \le i \le a, 1 \le j \le b$ has m replicates and $n = abm$

Let the observation y_{ijk} be the k^{th} replicate from the group with A at i, and B at j, $1 \le k \le m$

Cell Means Model

$y_{ijk} = \mu_{ij} + \epsilon_{ijk}$

Each of the μ_{ij} values is the true population mean of group i,j, estimated by $\bar{y}_{ij\cdot}$

Variation Around the Mean Model

$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk}$

where μ is the grand mean, which is the mean of the expected values of the responses,

$\mu = n^{-1} \sum_{ij} m\mu_{ij}$ estimated by $\bar{y}_{\cdots}$

α_i is the difference between $\mu_{i\cdot}$ and μ, estimated by $\bar{y}_{i\cdot\cdot} - \bar{y}_{\cdots}$, where $\mu_{i\cdot} = b^{-1} \sum_{j=1}^{b} \mu_{ij}$

$\alpha_i = \mu_{i\cdot} - \mu$ so $\sum \alpha_i = 0$. The set of $\{\alpha_i\}$ has a elements, but they add to 0, so only $a-1$ df

β_j is the difference between $\mu_{\cdot j}$ and μ, estimated by $\bar{y}_{\cdot j\cdot} - \bar{y}_{\cdots}$, where $\mu_{\cdot j} = a^{-1} \sum_{i=1}^{a} \mu_{ij}$ which has $b-1$ df

γ_{ij} is the difference between μ_{ij} and $\mu + \alpha_i + \beta_j = \mu_{i\cdot} + \mu_{\cdot j} - \mu$ estimated by $\bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot j\cdot} + \bar{y}_{\cdots}$

Suppose we have one factor A with a levels and one factor B with b levels.

The group with A at level i and B at level j, $1 \le i \le a, 1 \le j \le b$ has m replicates and $n = abm$

Let the observation y_{ijk} be the k^{th} replicate from the group with A at i, and B at j, $1 \le k \le m$

Cell Means Model

$y_{ijk} = \mu_{ij} + \epsilon_{ijk}$

Each of the μ_{ij} values is the true population mean of group i,j, estimated by $\overline{y}_{ij\cdot}$

Omitted Levels Model as used by MATLAB

$y_{ijk} = \mu_0 + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk}$

where μ_0 is population mean at the first level of each factor, estimated by $\overline{y}_{11\cdot}$

α_i is the difference between $\mu_{i\cdot}$ and $\mu_{1\cdot}$, estimated by $\overline{y}_{i\cdot\cdot} - \overline{y}_{1\cdot\cdot}$, where $\mu_{i\cdot} = b^{-1}\sum_{j=1}^{b} \mu_{ij}$ so $\alpha_1 = 0$

β_j is the difference between $\mu_{\cdot j}$ and $\mu_{\cdot 1}$, estimated by $\overline{y}_{\cdot j\cdot} - \overline{y}_{\cdot 1\cdot}$, where $\mu_{\cdot j} = a^{-1}\sum_{i=1}^{a} \mu_{ij}$ and $\beta_1 = 0$

γ_{ij} is the difference between μ_{ij} and $\mu_0 + \alpha_i + \beta_j = \mu_{11} + \mu_{i\cdot} - \mu_{1\cdot} + \mu_{\cdot j} - \mu_{\cdot 1}$

Comparing Cell Means

- In the coking data set, we have six groups defined by two variables, `width` and `temperature`.
- If we want to compare the six groups, we can reformulate this as a one-way ANOVA.
- Then we can use `anovan` and `multcompare` to compare the groups as a whole.
- We concatenate the variable values, converting to strings if the values are numbers, then use {} to make it a cell array to use as a grouping variable in `anovan`.

```
>> widthtemp = [int2str(coking.width) int2str(coking.temp)]

 41600
 41600
 41600
 41900
 41900
 41900
 81600
 81600
 81600
 81900
 81900
 81900
121600
121600
121600
121900
121900
121900
```

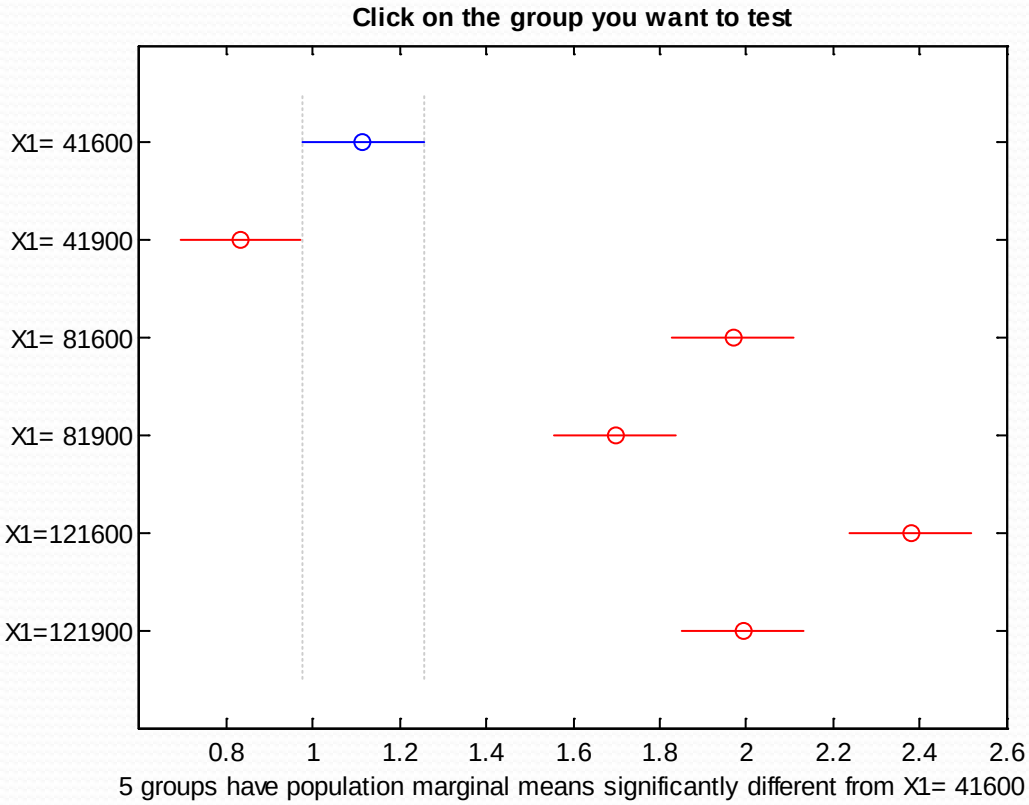
```
>> [p table stats] = anovan(coking2.logtime,{widthtemp})
p =
   2.5316e-09

table =
    'Source'    'Sum Sq.'    'd.f.'    'Singular?'    'Mean Sq.'    'F'          'Prob>F'
    'X1'        [ 5.1203]    [   5]    [        0]    [  1.0241]    [98.7856]    [2.5316e-09]
    'Error'     [ 0.1244]    [  12]    [        0]    [  0.0104]           []             []
    'Total'     [ 5.2447]    [  17]    [        0]           []           []             []

stats =
         source: 'anovan'
          resid: [18x1 double]
         coeffs: [7x1 double]
            Rtr: [6x6 double]
       rowbasis: [6x7 double]
            dfe: 12
            mse: 0.0104
    nullproject: [7x6 double]
          terms: 1
         nlevels: 6
     continuous: 0
         vmeans: 0
       termcols: [2x1 double]
     coeffnames: {7x1 cell}
           vars: [7x1 double]
       varnames: {'X1'}
       grpnames: {{6x1 cell}}
```

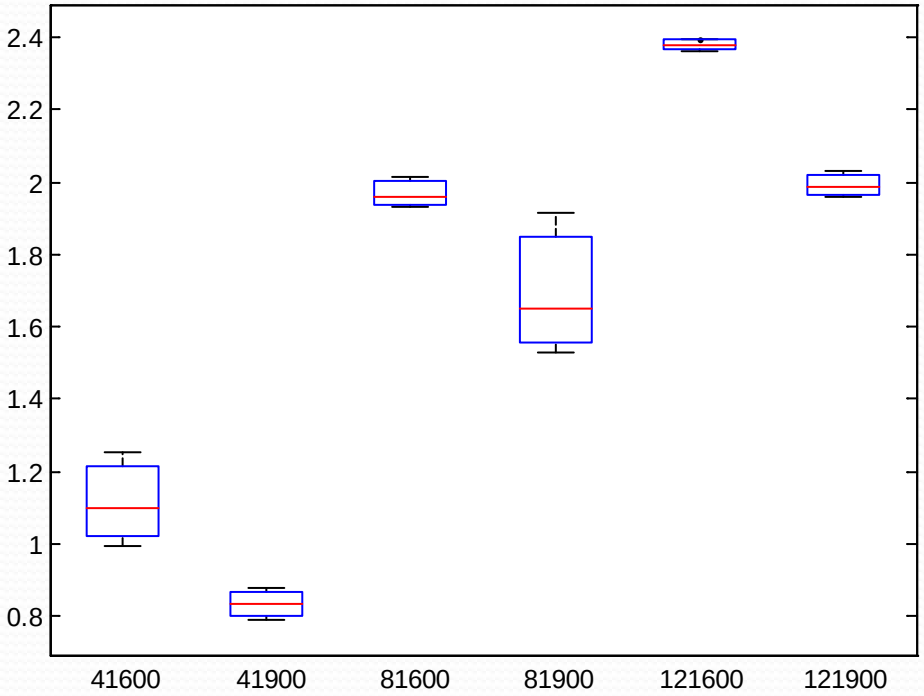
```
>> multcompare(stats)

    1.0000    2.0000    0.0034    0.2826    0.5618
    1.0000    3.0000   -1.1332   -0.8540   -0.5747
    1.0000    4.0000   -0.8616   -0.5823   -0.3031
    1.0000    5.0000   -1.5438   -1.2646   -0.9853
    1.0000    6.0000   -1.1564   -0.8772   -0.5979
    2.0000    3.0000   -1.4158   -1.1366   -0.8573
    2.0000    4.0000   -1.1442   -0.8649   -0.5857
    2.0000    5.0000   -1.8264   -1.5472   -1.2679
    2.0000    6.0000   -1.4390   -1.1598   -0.8805
    3.0000    4.0000   -0.0076    0.2716    0.5509
    3.0000    5.0000   -0.6898   -0.4106   -0.1314
    3.0000    6.0000   -0.3024   -0.0232    0.2560
    4.0000    5.0000   -0.9615   -0.6822   -0.4030
    4.0000    6.0000   -0.5741   -0.2948   -0.0156
    5.0000    6.0000    0.1082    0.3874    0.6666
```

```
>> multcompare(stats,'ctype','lsd')

ans =

    1.0000    2.0000    0.1015    0.2826    0.4637
    1.0000    3.0000   -1.0351   -0.8540   -0.6728
    1.0000    4.0000   -0.7635   -0.5823   -0.4012
    1.0000    5.0000   -1.4457   -1.2646   -1.0834
    1.0000    6.0000   -1.0583   -0.8772   -0.6960
    2.0000    3.0000   -1.3177   -1.1366   -0.9554
    2.0000    4.0000   -1.0461   -0.8649   -0.6838
    2.0000    5.0000   -1.7283   -1.5472   -1.3660
    2.0000    6.0000   -1.3409   -1.1598   -0.9786
    3.0000    4.0000    0.0905    0.2716    0.4528
    3.0000    5.0000   -0.5917   -0.4106   -0.2295
    3.0000    6.0000   -0.2043   -0.0232    0.1579
    4.0000    5.0000   -0.8633   -0.6822   -0.5011
    4.0000    6.0000   -0.4760   -0.2948   -0.1137
    5.0000    6.0000    0.2063    0.3874    0.5685
```

```
>> getfield(stats,'grpnames')

    {6x1 cell}
>> celldisp(getfield(stats,'grpnames'))

ans{1}{1} =
 41600
ans{1}{2} =
 41900
ans{1}{3} =
 81600
ans{1}{4} =
 81900
ans{1}{5} =
121600
ans{1}{6} =
121900
```

```
>> boxplot(coking2.logtime,widthtemp)
>> grpstats(coking2.logtime,widthtemp)

    1.1149
    0.8323
    1.9688
    1.6972
    2.3794
    1.9920

>> exp(grpstats(coking2.logtime,widthtemp))

    3.0492  #  3:03   4″   1600
    2.2985  #  2:18   4″   1900
    7.1624  #  7:10   8″   1600
    5.4587  #  5:28   8″   1900
   10.7988  # 10:48  12″   1600
    7.3305  #  7:20  12″   1900
```


Click on the group you want to test
X1= 41600
X1= 41900
X1= 81600
X1= 81900
X1=121600
X1=121900
0.8
1
1.2
1.4
1.6
1.8
2
2.2
2.4
2.6
5 groups have population marginal means significantly different from X1= 41600


2.4
2.2
2
1.8
1.6
1.4
1.2
1
0.8
41600
41900
81600
81900
121600
121900

Additive Model

	SumSq	DF	MeanSq	F	pValue
width	4.6648	2	2.3324	238.95	1.5125e-11
temp	0.44332	1	0.44332	45.419	9.4813e-06
Error	0.13665	14	0.0097608		

Interaction Model

	SumSq	DF	MeanSq	F	pValue
width	4.6648	2	2.3324	224.99	3.0714e-10
temp	0.44332	1	0.44332	42.764	2.7718e-05
width:temp	0.012252	2	0.006126	0.59094	0.56915
Error	0.1244	12	0.010367		

Cell Means Model

	SumSq	DF	MeanSq	F	pValue
widthtemp	5.1203	5	1.0241	98.786	2.5316e-09
Error	0.1244	12	0.010367		

When moving from the interaction model to the additive model, the error term for the latter is the sum of the SSE and the SS(width:temp) from the interaction model. 0.1244 + 0.0123 = 0.1367

The cell means model has the same SSE as the interaction model. The SS(widthtemp) is the sum of the sums of squares of width, temp, and width:temp. 4.6648 + 0.4433 + 0.0123 = 5.1203.

Some of these depend on balanced data.

Heart Rate Data

- Six subjects with congestive heart failure are given a drug (enalaprilat) which is meant to lower blood pressure and heart rate. It is in the class of angiotensin converting enzyme (ACE) inhibitors.
  - hr = heart rate in beats per minute
  - subj = subject number (1–9) as a factor
  - time = a factor with levels 0 (before), 30, 60, 120 (minutes after administration).
- This is a balanced two-way study with one observation per cell.
- This means that the interaction term cannot be estimated.

```
>> fitlm(heartrate,'hr~subj+time','CategoricalVars',[2,3])

ans =


Linear regression model:
    hr ~ 1 + subj + time

Estimated Coefficients:
                    Estimate    SE        tStat      pValue
    (Intercept)      94.917     2.0302     46.751     4.366e-25
    subj_2              18      2.4865      7.239     1.7662e-07
    subj_3            -5.75     2.4865    -2.3125      0.029644
    subj_4              -8      2.4865    -3.2173     0.0036828
    subj_5            30.5      2.4865     12.266     7.9273e-12
    subj_6             6.5      2.4865     2.6141      0.015212
    subj_7             -22      2.4865    -8.8476     5.0746e-09
    subj_8             -16      2.4865    -6.4346     1.1819e-06
    subj_9            11.5      2.4865     4.6249     0.00010779
    time_30             -4      1.6577     -2.413      0.023822
    time_60         -5.4444     1.6577    -3.2844     0.0031291
    time_120        -4.2222     1.6577     -2.547      0.017693


Number of observations: 36, Error degrees of freedom: 24
Root Mean Squared Error: 3.52
R-squared: 0.968,  Adjusted R-Squared 0.954
F-statistic vs. constant model: 67, p-value = 2.88e-15
```

Differences between subjects

Differences between times

```
>> anova(fitlm(heartrate,'hr~subj+time','CategoricalVars',[2,3]))

ans =

            SumSq     DF    MeanSq    F          pValue
    subj    8966.6     8    1120.8    90.639     4.8627e-16
    time    150.97     3    50.324    4.0696        0.01802
    Error   296.78    24    12.366
```

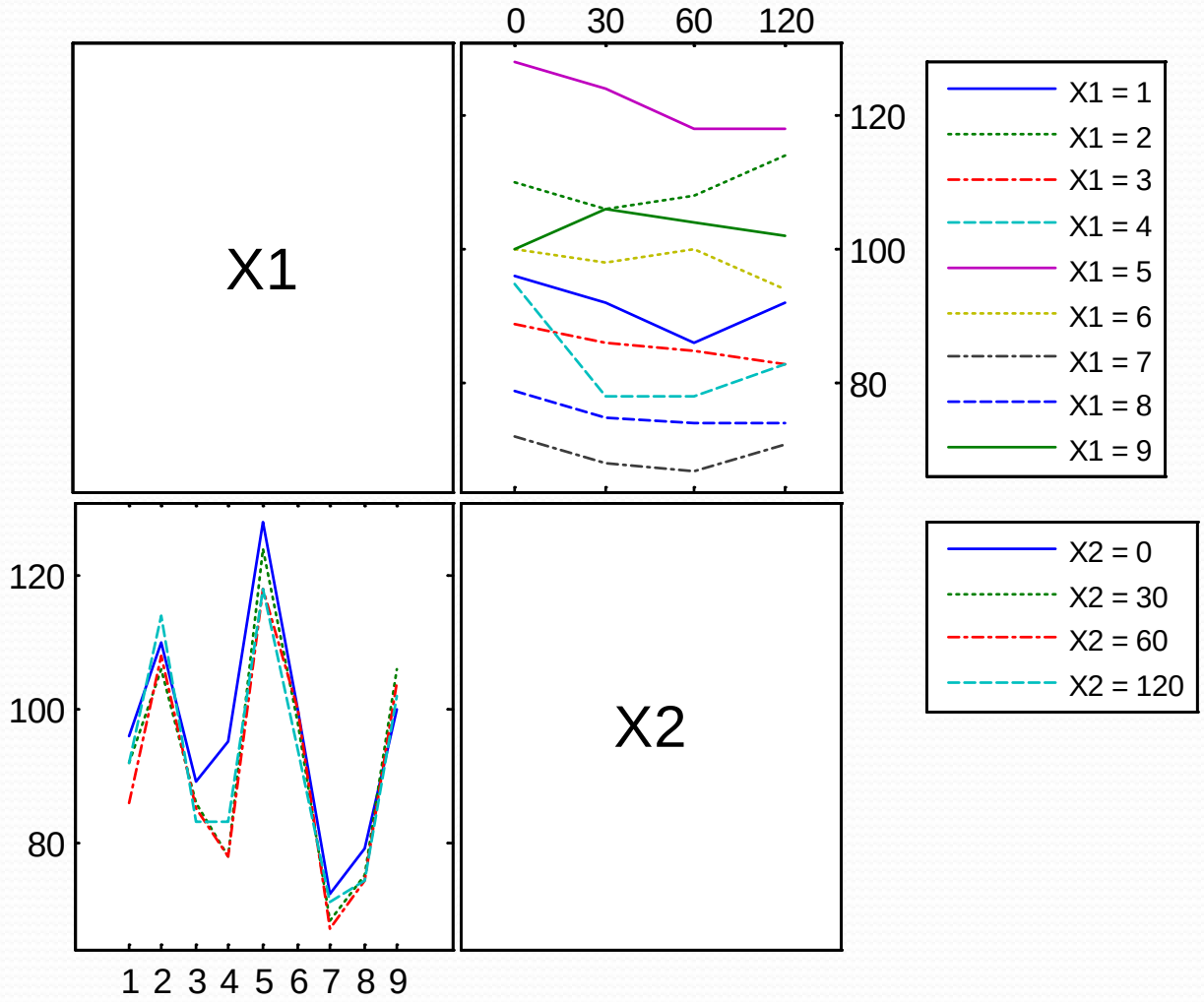
There is a very strong difference among subjects (expected).

The drug changes the heart rate from the baseline over time.

The coefficient tests show that the heart rate is lowered at 30 minutes and that this persists at 60 and 120 minutes.

We cannot estimate the interaction directly—which is the difference across individuals in the time course of changes in heart rate over time.

```
>> interactionplot(heartrate.hr,{heartrate.subj heartrate.time})
```


0
30
60
120
X1
120
100
80
X1 = 1
X1 = 2
X1 = 3
X1 = 4
X1 = 5
X1 = 6
X1 = 7
X1 = 8
X1 = 9
120
100
80
X2
X2 = 0
X2 = 30
X2 = 60
X2 = 120
1 2 3 4 5 6 7 8 9