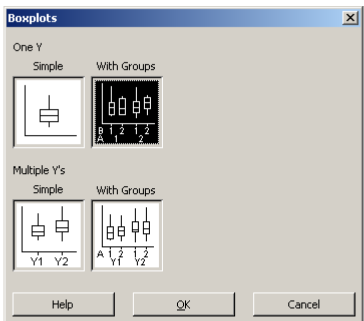
Example: Two-way ANOVA on Laundry Detergent, LRS pp.518-522

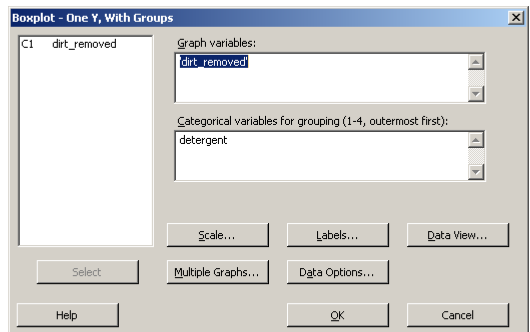
Load the data

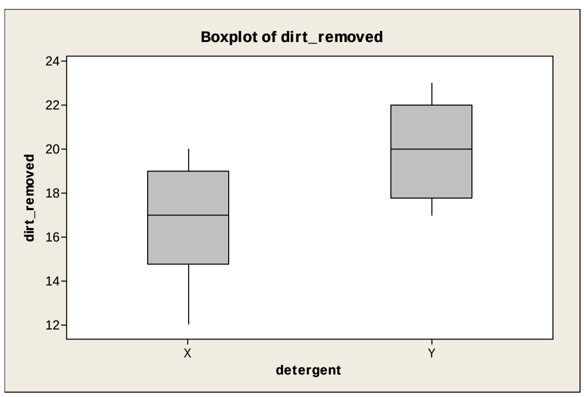
1. Open laundry2.txt. Don’t read first row as the header.
2. Label columns: C1: dirt_removed, C2: detergent, C3: temperature

Make BoxPlots

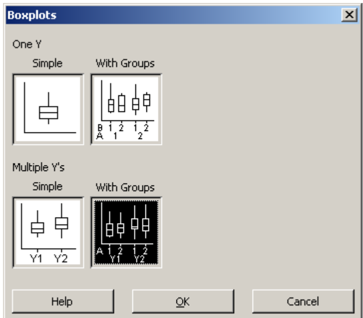
1. ***Wrong way***: Look at only Detergent brand

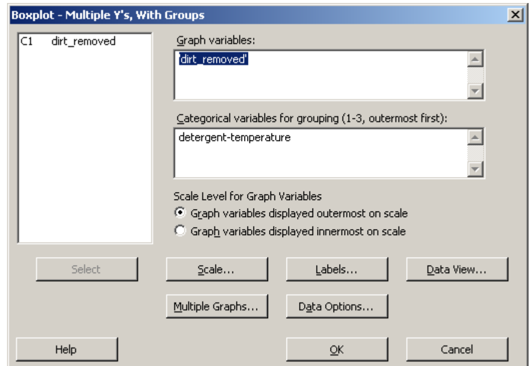


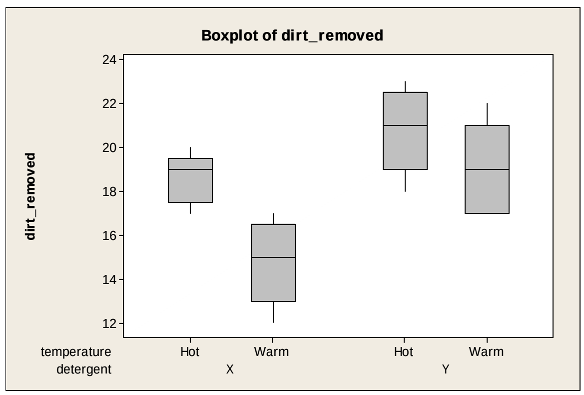




2. ***Correct way***: Boxplot with separate Factors







Two-way ANOVA

The three hypotheses to test are

1. There is no effect due to detergent brand. Let μ_A be the mean amount of dirt removed by brand A, and μ_B be the mean amount of dirt removed by brand B.

$$H_0: \ \mu_A = \mu_B$$
$$H_1: \ \mu_A \neq \mu_B$$

2. There is no effect due to temperature. Let μ_{T1} be the mean amount of dirt removed at temperature T_1, and μ_{T2} be the mean amount of dirt removed at temperature T_2.

$$H_0: \ \mu_{T1} = \mu_{T2}$$
$$H_1: \ \mu_{T1} \neq \mu_{T2}$$

3. There are no interactions.

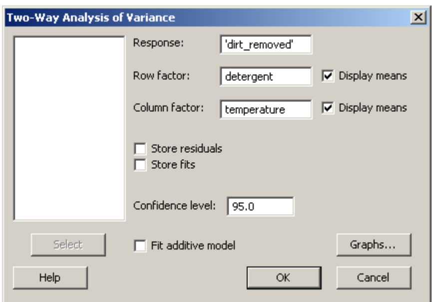
H_0: Interaction of detergent and temperature is zero

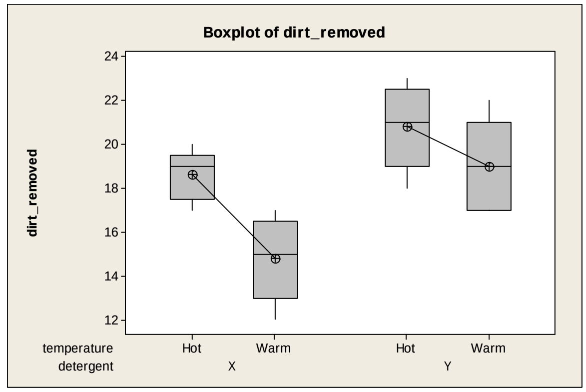
H_1: Interaction of detergent and temperature is *not* zero

Choose significance : $\alpha = 0.05$

Use MINITAB to perform the two-way ANOVA

1. Stat → ANOVA-two way
   a. Click *Graphs…*, then select *BoxPlots of data*
   b. Check *Display means* for both “detergent” and “temperature”






```
Two-way ANOVA: dirt_removed versus detergent, temperature

Source        DF     SS     MS      F      P
detergent      1   51.2   51.2  15.52  0.001
temperature    1   39.2   39.2  11.88  0.003
Interaction    1    5.0    5.0   1.52  0.236
Error         16   52.8    3.3
Total         19  148.2

S = 1.817   R-Sq = 64.37%   R-Sq(adj) = 57.69%


                 Individual 95% CIs For Mean Based on
                 Pooled StDev
detergent  Mean  -------+---------+---------+---------+--
X          16.7  (-------*-------)
Y          19.9                       (-------*-------)
                 -------+---------+---------+---------+--
                     16.5      18.0      19.5      21.0


                   Individual 95% CIs For Mean Based on
                   Pooled StDev
temperature  Mean  -----+---------+---------+---------+----
Hot          19.7                    (-------*-------)
Warm         16.9  (-------*-------)
                   -----+---------+---------+---------+----
                     16.5      18.0      19.5      21.0
```

Hypothesis 1: There is no effect of detergent

$$F_A = MSA/MSE = 15.52, \quad dfn = 1; \; dfd = 16, \quad p = 0.001$$

Since $p \approx 0.001$, the observed F is not with the region of acceptance. We reject the null hypothesis: there is strong evidence (at $p \approx 0.001$) that the kind of detergent has an effect on the amount of dirt removed. Or, we can say: there is a roughly 1/1000 chance that the choice of detergent has *no* effect.

How would you compute p-value for the observed F statistic if it was not included in the ANOVA report in the Session Window?

1. Calc → Probability Distributions → F…
2. Select Cumulative probability, leave Noncentrality parameter equal to 0
3. Numerator degrees of freedom: 1
4. Denominator degrees of freedom: 16
5. Input constant: 15.52

```
Cumulative Distribution Function

F distribution with 1 DF in numerator and 16 DF in denominator

    x  P( X <= x )
15.52     0.998828
```

P is the cumulative distribution, so the *p*-value is

$$1 - P(X <= x) = 1 – 0.9988 = 0.00112$$

Hypothesis 2: There is no effect of temperature

$$F_B = MSB/MSE = 11.88, \quad dfn = 1; \; dfd = 16, \; p = 0.003$$

Since $p = 0.003$, the observed F is not with the region of acceptance. We reject the null hypothesis. There is strong evidence (at $p = 0.003$) that the temperature has an effect on the amount of dirt removed. Or, we can say: there is a roughly 3/1000 chance that the choice of temperature has *no* effect.

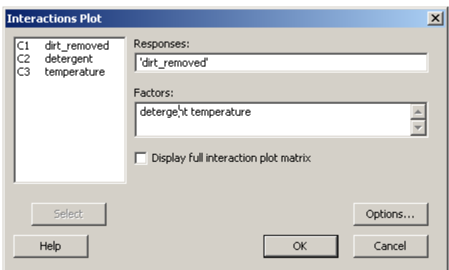
Hypothesis 3: There is no interaction

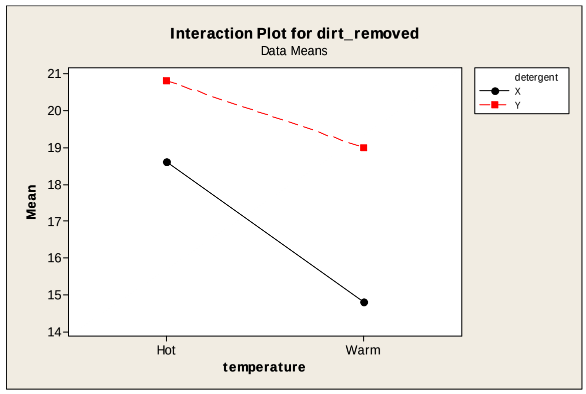
$$F_{AB} = MSAB/MSE = 1.52, \quad dfn = 1; \; dfd = 16, \quad p = 0.236$$

Since $p>0.05$, the observed F is in the region of acceptance. We do not reject the null hypothesis. There is no evidence that there is an interaction between temperature and brand. Or, we can say: the small effect of interaction between detergent brand and temperature is likely due to chance.

Interaction Plots

1. Stat → ANOVA → Interaction Plots
2. Responses: dirt_removed
3. Factors: detergent and temperature
   a. Click in the empty *Factors* box
   b. Select *C2: detergent* and *C3: temperature* in the larger box on the left
   c. Click the *Select* button





Example: Two-level, Three-factor ANOVA on Laundry Detergent, pp.525-531

The data set laundry3.txt contains data on a two-level, three factor experiment. The levels are fixed, so there are $2^3 = 8$ experimental conditions. Each condition is replicated twice, giving 16 experimental observations.

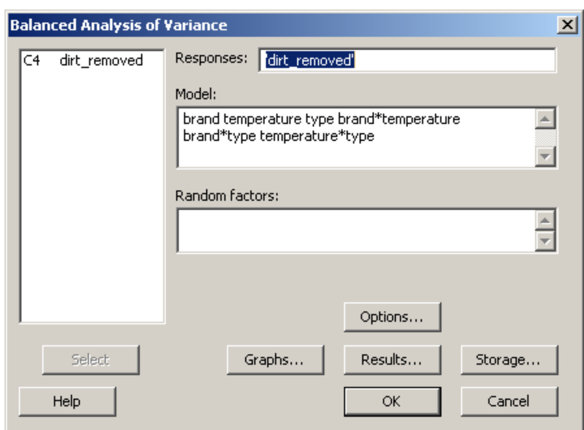
The response is the amount of dirt removed. The factors are detergent brand (X or Y), detergent type (powder or liquid), and washing temperature (hot or warm). Therefore, there are three main effects — brand, type and temperature — and three possible interactions — brand × type, brand × temperature, and type × temperature. This leads to six F tests for six null hypotheses. The significance is chosen to be $\alpha = 0.5$.

Load the data

1. Open laundry3.tx t. Don't read first row as the header.
2. Label columns: C1: brand, C2: temperature, C3: type, C4: dirt_removed

Use Balanced ANOVA

1. Stat → ANOVA → Balanced ANOVA
2. Make the following selections in the *Balance Analysis of Variance* dialog box
   a. In the *Responses* box enter dirt_removed
   b. In the *Model* box enter: brand temperature type brand*temperature brand*type temperature*type

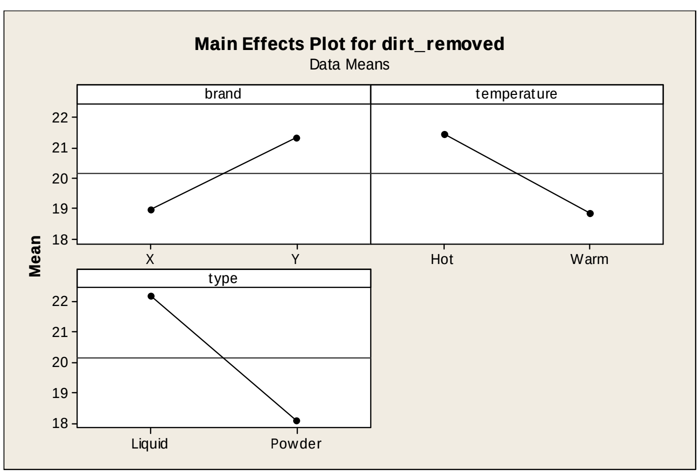


3 Click OK

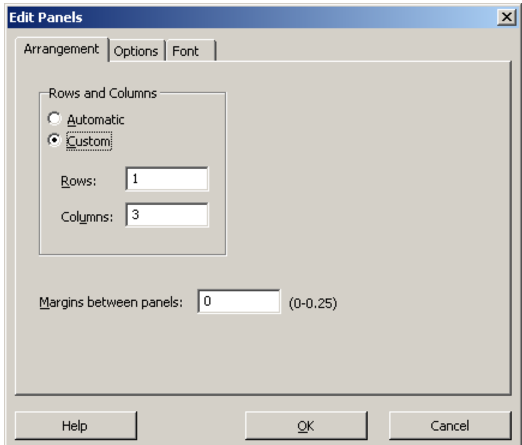
```
ANOVA: dirt_removed versus brand, temperature, type

Factor       Type   Levels  Values
brand        fixed       2  X, Y
temperature  fixed       2  Hot, Warm
type         fixed       2  Liquid, Powder


Analysis of Variance for dirt_removed

Source             DF       SS      MS       F      P
brand               1   22.090  22.090  156.54  0.000
temperature         1   27.040  27.040  191.62  0.000
type                1   67.240  67.240  476.50  0.000
brand*temperature   1    0.490   0.490    3.47  0.095
brand*type          1    3.610   3.610   25.58  0.001
temperature*type    1    0.040   0.040    0.28  0.607
Error               9    1.270   0.141
Total              15  121.780


S = 0.375648   R-Sq = 98.96%   R-Sq(adj) = 98.26%
```

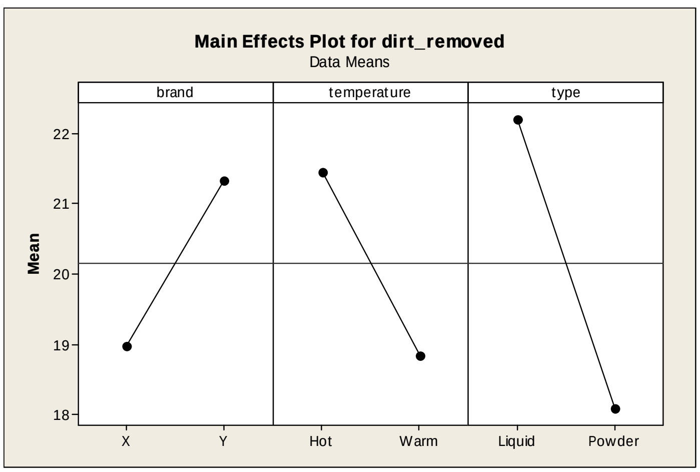
Show plots of Main Effects

1. Stat →ANOVA → Main Effects Plot
   a. In the Responses box enter dirt_removed
   b. In the Factors box enter brand temperature type



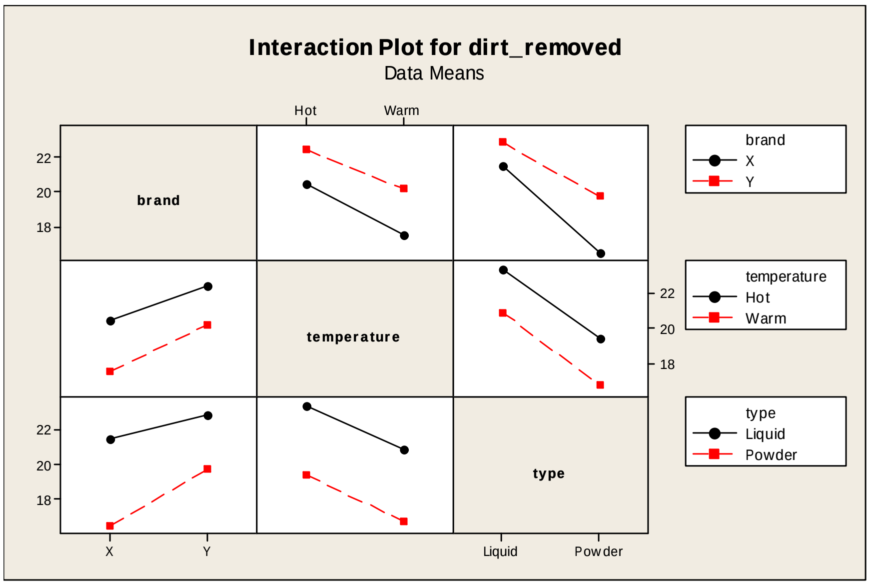
2. You can change the alignment so that all three factors are on the same row, which makes visual comparison of the response easier. This step is optional
   a. With the plot in the foreground, select Editor → Panel…
   b. In the Arrangement tab, select the *Custom* radio button in the *Rows and Columns* region. Set the display to 1 *Rows* and 3 *Columns*





Show plots of Interactions

1. Stat → ANOVA → Interaction Plots
   a. Response: dirt_removed
   b. Factors: brand temperature type
   c. Check: Display full interaction plot matrix



Interactions are evident when the response lines are not parallel. From the plots in the upper right corner and lower left corner, we suspect that there is an interaction between brand and type. The plots are consistent with the ANOVA which suggests that we reject the null hypotheses for the brand/type interaction at p=0.001.

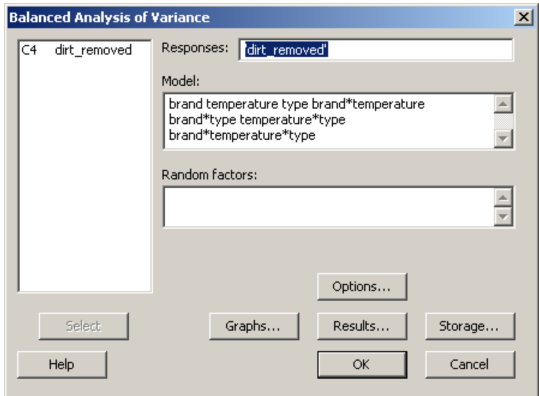
Including the Higher Order Interaction Term

The preceding analysis did not include the three-way interaction between brand, temperature and type. What can we expect if that interaction is included?

Adding the interaction to the model allows for more of the variance to be explained by factors under control of the person who designs the experiment. The variation is already present in the result.

Add the three-way interaction and re-run the analysis.

1. Stat → ANOVA → Balanced ANOVA
2. Make the following selections in the *Balance Analysis of Variance* dialog box
   a. In the *Model* box add the interaction: brand*temperature*type



3. Click OK

ANOVA: dirt_removed versus brand, temperature, type

Factor	Type	Levels	Values
brand	fixed	2	X, Y
temperature	fixed	2	Hot, Warm
type	fixed	2	Liquid, Powder

Analysis of Variance for dirt_removed

Source	DF	SS	MS	F	P
brand	1	22.090	22.090	140.25	0.000
temperature	1	27.040	27.040	171.68	0.000
type	1	67.240	67.240	426.92	0.000
brand*temperature	1	0.490	0.490	3.11	0.116
brand*type	1	3.610	3.610	22.92	0.001
temperature*type	1	0.040	0.040	0.25	0.628
brand*temperature*type	1	0.010	0.010	0.06	0.807
Error	8	1.260	0.157		
Total	15	121.780			

S = 0.396863   R-Sq = 98.97%   R-Sq(adj) = 98.06%

The following exerpts from the MINITAB output show the differences with the analysis without the three way interaction term:

Without the three way interaction: (partial results)

brand*type	1	3.610	3.610	25.58	0.001
temperature*type	1	0.040	0.040	0.28	0.607
Error	9	1.270	0.141		
Total	15	121.780			

With the three way interaction: (partial results)

brand*type	1	3.610	3.610	22.92	0.001
temperature*type	1	0.040	0.040	0.25	0.628
brand*temperature*type	1	0.010	0.010	0.06	0.807
Error	8	1.260	0.157		
Total	15	121.780			

The sum of squares (SS) and mean sum of squares (MS) for the main effects and the two-way interactions are unchanged. Adding the three way interaction creates a row with values for SS, MS, F, and p values of the three way interaction. Also changed is the magnitude of SSE, which has decreased, and the magnitude of MSE, which has increased. The value of SSE is decreased exactly by the amount of the SS attributed to the three-way interaction. The value of MSE increases because the number of degrees of freedom for MSE decreases. In this case the decrease in SSE is more than offset by the decrease in the number of degrees of freedom: (1.26/8) > (1.27/9).

This example shows that by explicitly including more terms in the model for the variance, the amount of unexplained error (attributed to SSE) is decreased. Note that changing MSE will change the F and p values for the main effects because the F statistic in those terms has MSE in the denominator.