

Design of Experiments – Multi-Factor Categorical Designs (Legacy Procedure)

Summary

The *Multi-Factor Categorical Designs* selection on the *Create Design* menu creates experimental designs for situations where primary interest centers on comparing levels of two or more categorical factors. The procedure will create a multilevel factorial design with runs at each combination of the levels of the factors. To analyze the results of the experiment once it has been performed, the data is input to the *Multifactor ANOVA* procedure.

Sample StatFolio: *doe multifactor.sgp*

Sample Data:

The example in this document comes from Kutner et al. (1996). They describe a situation in which researchers wished to conduct a stress test on a treadmill to determine the effects of three factors: smoking history, amount of body fat, and gender. They expected that the factors might interact, so they wished to amount more than one individual at each combination of the factors. The levels of the factors they selected were:

Amount of body fat: low or high

Gender: male or female

Smoking history: none, light, or heavy

They decided to select 3 individuals for each of the 12 combinations of the factor levels.

Design Creation

To create an experiment in which the primary goal is to analyze the effect of several categorical factors, select *Create Design* from the design of experiments menu and complete the dialog boxes as described below.

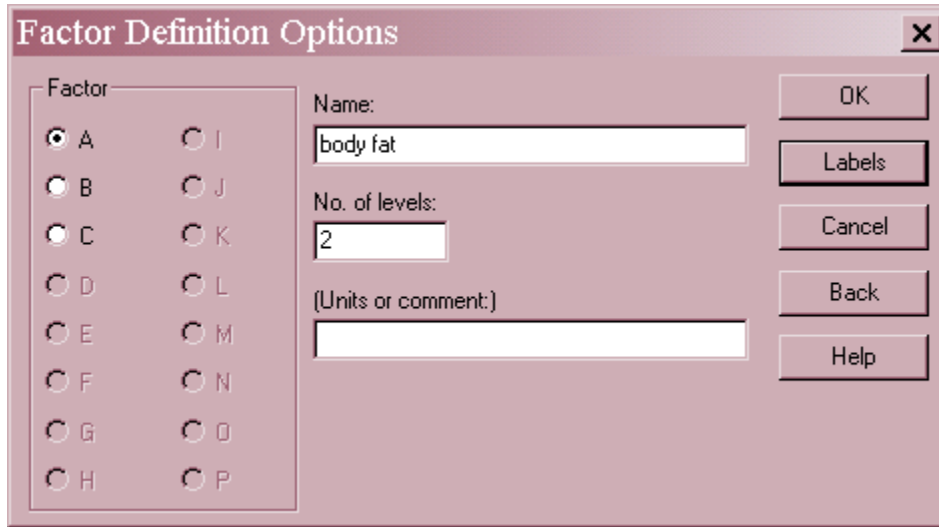
Dialog #1 – Design Type

The first dialog box displayed during design creation specifies the type of design to be created:

- **Design Class:** type of design to be created.
- **No. of Response Variables:** the number of response variables Y that will be measured during each experimental run. This number may range from 1 to 16.
- **No. of Experimental Factors:** the number of categorical factors X to be studied. This number may range from 2 to 8.
- **Comment:** a comment to appear on the output from the analysis procedures.

Dialog #2 – Experimental Factor

The second dialog box requests information about each experimental factor to be tested:



- **Name:** enter a name for the factor containing up to 32 characters. A column will be created in the datasheet for each factor.
- **No. of levels:** number of different levels of the factor at which experiments will be performed.
- **Units or comment** - an optional label or comment of up to 64 characters to be included on the experiment worksheet.
- **Labels button:** Press this button to enter identifiers for each level:



For the current example, the information needed to complete the dialog boxes is shown below:

| <i>Name</i> | <i>No. of levels</i> | <i>Labels</i> |
|-------------|----------------------|--------------------|
| body fat | 2 | low, high |
| gender | 2 | male, female |
| smoking | 3 | none, light, heavy |

Dialog #3 – Response Variables

The third dialog box requests information about each of the response variables:

Click on the numbers 1, 2, 3, ..., one at a time and enter the following information for each response variable in the experiment:

- **Name** - a name for each response containing up to 32 characters.
- **Units or comment** - an optional label or comment of up to 64 characters to be included on the experiment worksheet.

Dialog #4 – Design Options

The fourth dialog box is used to specify the number of replicates at each combination of the factors levels and whether the order of the runs should be randomized:

- **Replicate Design:** number of additional runs at each combination of the factor levels (not counting the first run).
- **Randomize:** check to put the runs in random order within each set of replicates.

The above dialog box requests a total of 36 runs, 3 each at the $2 \times 2 \times 3 = 12$ combinations of the factor levels.

Design Attributes

Once the dialog boxes have been completed, the *Design Attributes* window will be displayed:

| <u>Multi-factor Categorical Design Attributes</u> | | |
|---|---------------|--------------|
| Design class: Multi-factor Categorical | | |
| File name: <Untitled> | | |
| Comment: Stress test | | |
| Base Design | | |
| Number of experimental factors: 3 | | |
| Number of responses: 1 | | |
| Number of runs: 36 | | |
| Error degrees of freedom: 26 | | |
| Randomized: No | | |
| <i>Factors</i> | <i>Levels</i> | <i>Units</i> |
| body fat | 2 | |
| gender | 2 | |
| smoking | 3 | |
| <i>Responses</i> | <i>Units</i> | |
| minutes | | |

It contains a summary of the generated design. To save the design, select *Save Design File* from the main *File* menu. The example is saved in the *stresstest.sfx* file.

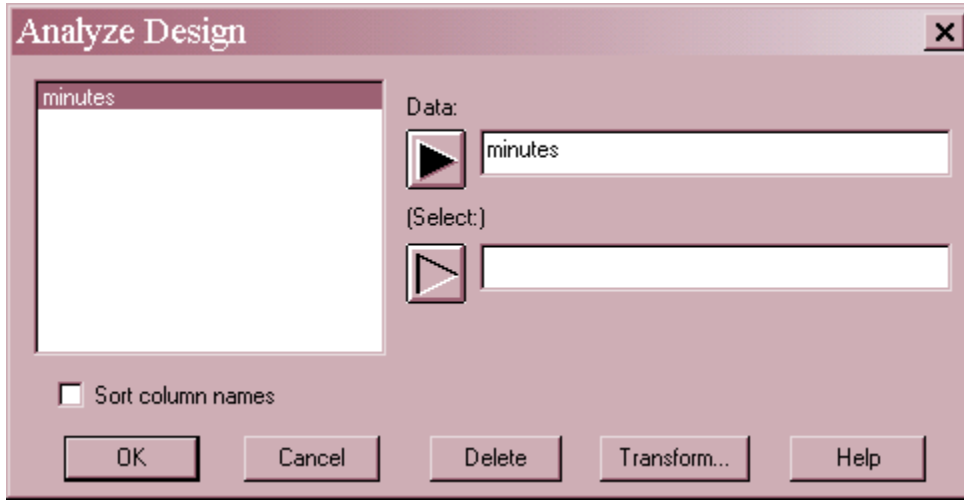
Worksheet

The experimental runs are entered into the datasheet automatically. They may also be displayed in the *Worksheet*, a portion of which is shown below:

| Worksheet for stresstest.sfx - Stress test | | | | |
|--|-----------------|---------------|----------------|----------------|
| <i>run</i> | <i>body fat</i> | <i>gender</i> | <i>smoking</i> | <i>minutes</i> |
| 1 | low | male | none | |
| 2 | low | male | none | |
| 3 | low | male | none | |
| 4 | low | male | light | |
| 5 | low | male | light | |
| 6 | low | male | light | |
| 7 | low | male | heavy | |
| 8 | low | male | heavy | |
| 9 | low | male | heavy | |
| 10 | low | female | none | |
| 11 | low | female | none | |
| 12 | low | female | none | |
| 13 | low | female | light | |
| 14 | low | female | light | |
| 15 | low | female | light | |
| 16 | low | female | heavy | |
| 17 | low | female | heavy | |
| 18 | low | female | heavy | |
| 19 | high | male | none | |
| 20 | ... | ... | ... | |

Analyzing the Data

After the results of the experimental runs have been entered, *Analyze Data* can be selected from the DOE menu. A dialog box will be displayed requesting the column containing the response to be analyzed:



- **Data:** column containing the values of the response variable to be analyzed.
- **Select:** subset selection.

When OK is pressed, the program will invoke the *Multifactor ANOVA* procedure. Full details of that analysis is contained in the *Multifactor ANOVA* documentation.

Of particular interest in the current example are several tables and graphs:

ANOVA Table

This table is used to judge whether or not there are statistically significant differences between the levels of the experimental factors:

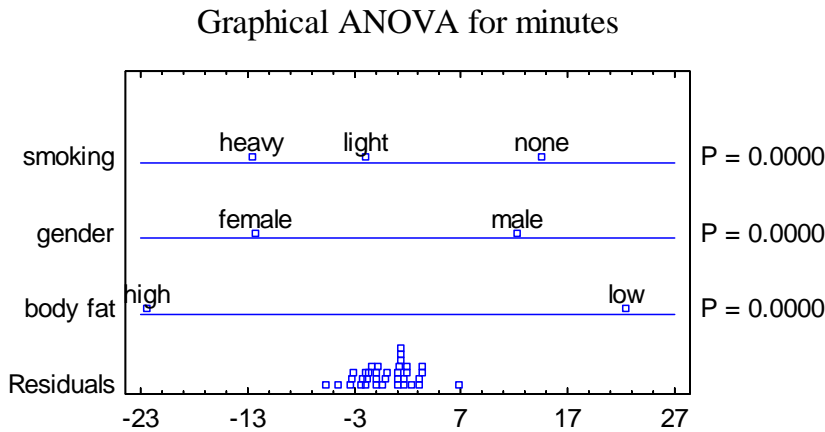
| Source | Sum of Squares | Df | Mean Square | F-Ratio | P-Value |
|---------------------|----------------|----|-------------|---------|---------|
| MAIN EFFECTS | | | | | |
| A:body fat | 702.25 | 1 | 702.25 | 79.10 | 0.0000 |
| B:gender | 210.25 | 1 | 210.25 | 23.68 | 0.0000 |
| C:smoking | 343.056 | 2 | 171.528 | 19.32 | 0.0000 |
| INTERACTIONS | | | | | |
| AB | 2.25 | 1 | 2.25 | 0.25 | 0.6189 |
| AC | 204.167 | 2 | 102.083 | 11.50 | 0.0003 |
| BC | 21.5 | 2 | 10.75 | 1.21 | 0.3142 |
| RESIDUAL | 230.833 | 26 | 8.87821 | | |
| TOTAL (CORRECTED) | 1714.31 | 35 | | | |

All F-ratios are based on the residual mean square error.

A small P-value for any main effect or interaction (less than 0.05 if operating at the 5% significance level) indicates that the corresponding factor has a significant effect on the response. If the current example, all 3 factors have significant main effects, and there is a significant interaction between *body fat* and *smoking*.

Graphical ANOVA

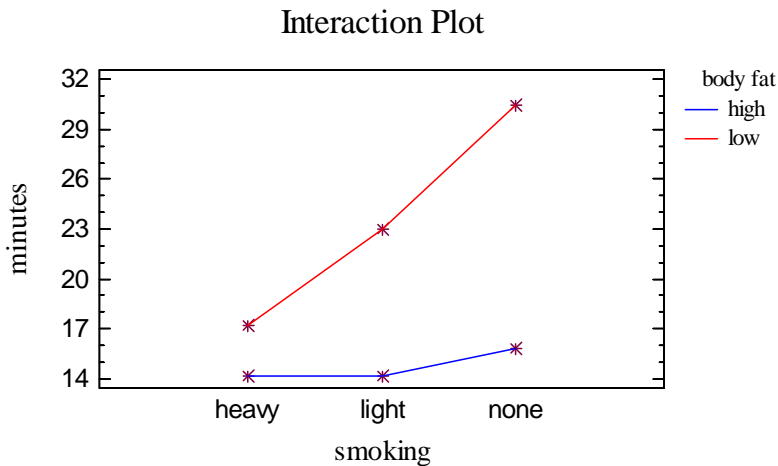
A new method for illustrating the results of an analysis of variance, from Hunter (2005), is shown below:



The plot shows the scaled deviations of the block and treatment averages from the grand mean, together with the model residuals. Scaling is such that, if a factor has no effect, the variation observed for that factor should be comparable to that of the residuals. Note that the variation for all factors is considerably greater than that of the residuals.

Interaction Plot

The *Interaction Plot* is particularly important when two factors show a significant interaction:



For example, the above plot shows the average response at each combination of *smoking* and *body fat*. The larger difference between the lines for non-smokers than for heavy smokers indicates that the effect of body fat is much greater for individuals who do not smoke.