Building a transfer function from several smaller tests

Drs. Jan Willem Bikker PDEng

The scenario

Early in the development of a new product, the effects of design parameters are not well known. In fact, the ways in which the design can be adapted change over time, so that the list of design parameters also evolves.

A project may use several small empirical studies to investigate the effect of the design parameters. Other activities and insights, together with the results from the empirical test, increase one's knowledge iteratively. During the sequence of studies, new design features and parameters are introduced along the way, while other design parameters may at some point be fixed and no longer vary in the later tests.

The tests are typically performed to study the impact of design parameters on a CtQ (Critical to Quality) parameter that describes the performance of the prototype product. The impact is described by a transfer function, which is typically built using statistics. The techniques used could be regression analysis and/or design of experiments — topics commonly taught in DfSS and statistical training courses.

Below we give an example based on an actual project. The structure of the data is the same, the transfer function is fictional.

Dataset example

The table gives the history of tests in a product development project. The rows correspond to the 8 tests, the columns are the design parameters and the cells list the values of the design parameters used in the test. The coloured cells are those with multiple values for the design parameter (i.e. the parameter was varied during the test).

	Design parameters				
Test	A	B	C	D	E
1	t	2	60	1	A, none
2	t	2	60	1	A
3	t	1,2,4	0,20,40	1,2	none
4	t	2	20, 60	1	A
5	t	2	60	1	A,B
6	t,r	2,3	4,24,26,56	1	none
7	t	2	40	1	none
8	r,q	2	26	1	none

Having a transfer function is very beneficial, as it summarizes the knowledge of the impact of a large set of considered technical parameters on product performance.

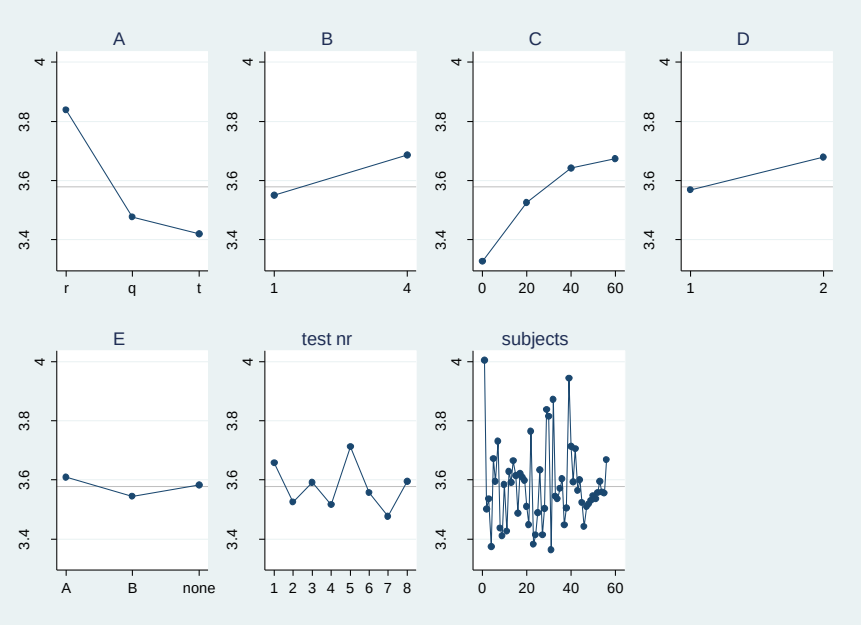
The best way to build such a transfer function empirically is to perform a new test that varies all parameters. The simplest test design would consist of all possible combinations of design parameters, which can turn out to be a very extensive test. ‘Design of experiments’ (DoE) can help limit the size of the design; but in practice, budget and time may still not permit such a test. In which case, you can attempt to build a rough transfer function based on the combined datasets of the smaller sets.

Building the transfer function from imperfect data

The transfer function describes the relation between the design parameter and the relevant product property, the CtQ. The unexplained element in this relation is the error. In an equation:

$$CtQ = f(A,B,C,D,E) + error$$

The resulting transfer function can be represented graphically as in the example below. Each design parameter gets a model term. Some of the design parameters are categorical factors, some continuous factors.



The outcome measure is given on the vertical axes (all with the same range): each panel has one factor plotted on the horizontal axis.

The last two panels display the variations between tests and between subjects. The variation between tests in the model is not explained by the factors A to E, and could be due to small changes in protocol, different operators, seasonal effects, etc. Likewise, the variation between subjects (human volunteers) isn't explained by the factors A to E and is due to inherent differences between the subjects.

In other projects, you could similarly use batches of raw material rather than human subjects. Compared to the effects of the design factors A to E, we see from the graph that the variation between tests is moderate, but the variation between subjects moderate to large.

Breakdown

Additionally, a breakdown can be given of how important the model terms are in explaining the outcome for this combined dataset.

parameter	settings	% variation
A	r, q, t	24%
B	1, 4	3%
C	0, 20, 40, 60	13%
D	1, 2	2%
E	A, B, none	1%
tests		8%
subjects		20%
within subjects	"=100%-R2"	29%

In this example, the within-subject variation accounts for 29%, which is the largest component. This can be linked to the so-called R^2, which is a well-known model diagnostic feature, though this link does get a bit technical:

> In a standard analysis after a DoE, statistics software usually displays the R^2, which is equal to the percentage of variation in the CtQ explained by the model. Note that a typical model for any of the single tests here would use *blocks* for the human subjects, and the explanatory value of the blocks is included in the R^2. In such cases, 100%- R^2 corresponds to the part of unexplained variation. This unexplained variation corresponds to the "within subjects" component in this table, so that the corresponding R^2 would be 100%-29%=71%. In total, the design parameters explain 24+3+13+2+1 = 43% of the variation[1].

[1] A very technical point here: in DoE, the R^2 describes how much of the variation is explained *in the current dataset.* In constructing this table, the percentages indicate how much variation is explained, not in the unbalanced dataset, but in a situation where the levels of the design parameters, tests and subjects are chosen randomly. In practice, the difference would be minimal.

Technical complications

There are several statistical challenges in building such a function compared to a standard workflow for single tests.

- Many **interactions** cannot be included. For instance, from the table we see that the combination D=2, E="A" has never been tested. By 'interaction' we mean: "where there is an interaction effect of D and E on the outcome, the effect of changing D from D=1 to D=2 would vary between the different values for E". However, using this dataset we can't test whether this is the case, so we have to assume there are no strong interactions.
- **Test conditions** (e.g. the test protocol) probably varied between the tests and/or there might have been seasonal effects. This can be remedied to a certain extend by including a test-dependent intercept to the model, so that outcomes may be structurally higher in one test than the other (over and above what is expected from the technical parameters A to E). But we must assume that the effect of the parameters is roughly the same across tests.
- The smaller tests may be designs that involve **blocks** (e.g. there are multiple observations per subject or per batch of material). The blocks should be incorporated into the model as an intercept term. Some data preprocessing might be necessary to insure that subject=1 from one test is regarded as a different person to subject=1 from another test.
- If you want **statistical inferences such as p-values and confidence intervals**, the complex and unbalanced structure often prohibits the use of the standard models generally used for analyzing single tests. For instance, the General Linear Models from Minitab 17 would not suffice. To see this, note how in this example parameter E is varied both within and between tests. The standard "General linear model" techniques from Minitab are based on least-squares estimates and cannot easily handle this feature. A convenient alternative is to use *linear mixed models*[2] estimated with maximum likelihood: they use the same or very similar model formulae, but different techniques, to fit the model.
- With **simpler statistical techniques** of the type used for single experiments, one can often make a transfer function describing the trend (e.g. using Minitab 17's general linear model or other regression techniques). However, p-values and confidence intervals cannot be trusted.
- The **combined dataset** may be so **unbalanced** that effects are not identified. For example, if two factors are changed simultaneously, the analysis might need to be split into several analyses, assumptions then made, and the model formula finally 'glued together'. We have come across examples of this.

Conclusion

In some development projects, a series of smaller tests is performed as knowledge increases. At a later stage, it is sometimes possible to build a rough transfer function, based on the combination of the datasets from the smaller tests. This transfer function then serves as a summary of the empirical knowledge gained to date.

[2] You can estimate linear mixed models (also known as 'random effects models' or 'hierarchical models') using the open source and freely available tool R, with the packages lme4 and multcomp, though some inferences are not easy to make. At CQM, we use Stata, which offers more support here. In the example above, we had three random effects: tests, subjects within tests, and sessions within subjects.

A rough picture can often be made using statistical techniques that are not overly complicated. A more refined picture is obtained by using a detailed analysis that produces p-values and confidence intervals. These help address the question: “could the trend we see just be the result of noise?” For such an unbalanced and ‘cobbled together’ dataset, more complicated techniques are usually required.

In our example, the data was still relatively balanced and large interactions between the design factors were not expected. If the situation deviates from these assumptions, the statistics needed to build the model become more complex, more assumptions may sometimes need to be made and the transfer function becomes less reliable.

For more information, please contact CQM
T +31 40 750 23 23
info@cqm.nl

Copyright © 2014 by CQM B.V.
All rights reserved. No part of this publication may be reproduced, distributed, or transmitted in any form or by any means, without the prior written permission of the publisher.