Auswertung von Definitive Screening Designs

Bernd Heinen
SAS Institute GmbH
In der Neckarhelle
69118 Heidelberg
Bernd.heinen@jmp.com

Bradley Jones
SAS Institute Inc
SAS Campus Drive
27513 Cary, NC
Bradley.jones@jmp.com

Zusammenfassung

Definitive Screening Designs stellen eine neue Klasse von Versuchsplänen dar, die mit besonders wenigen Experimenten zur Lösung von Screening Problemen beitragen. Ihre Effizienz und Vielseitigkeit ist ungeschlagen, aber die Auswertung erfordert doch ein spezielles Vorgehen. Haupteffekte und Effekte 1. Ordnung sollten getrennt voneinander analysiert werden.

Schlüsselwörter: Definitive Screening Design, DSD, Dummy Faktoren, zweistufige Analyse

1 Grundlagen

Screening Experimente werden durchgeführt, um festzustellen, welche aus einer Kandidatenliste von möglichen Faktoren überhaupt einen Einfluss auf die Zielgröße(n) haben, ohne die Art dieses Zusammenhangs feststellen zu wollen. Daher sind Screeningdesigns besonders sparsam mit der Zahl der Experimente. Andererseits soll ein Faktor aber auch dann auffallen, wenn sein Effekt nicht linear ist oder er an einer Wechselwirkung beteiligt ist. Die meisten Designs nutzen dafür Experimente, die am Mittelpunkt (center point) des Versuchsraums durchgeführt werden. Definitive Screening Designs (DSD) [1] vereinigen einige wünschenswerte Eigenschaften auf sich: die Haupteffekte sind untereinander unabhängig und unabhängig von allen Termen höherer Ordnung, es ist das Design mit der geringsten Versuchszahl für Screeningexperimente, alle Haupteffekte und quadratischen Terme sind schätzbar und falls sich maximal drei Faktoren als relevant herausstellen, kann aus den Versuchsdaten für diese auch ein vollständiges Oberflächenmodell geschätzt werden. D.h. sie bergen das Potential, einen kompletten Entwicklungsschritt überflüssig zu machen.

Die Ergebnisse von Experimenten werden üblicherweise mittels eines linearen Modells ausgewertet. Den Bezugsrahmen für diesen Vortrag bildet ein Wirkungsflächenmodell (Response Surface Model, RSM), d.h. ein lineares Modell mit Haupteffekten, quadratischen Termen und einfachen Wechselwirkungen. Wenn m die Zahl der Faktoren bezeichnet und n die Zahl der Beobachtungen, dann hat das vollständige Modell die Form

$$y_i = \beta_i + \sum_{i=1}^{m} \beta_j x_{ij} + \sum_{j=1}^{m-1} \sum_{k=j+1}^{m} \beta_{jk} x_{ij}\, x_{jk} + \sum_{j=1}^{m} \beta_{jj}\, x_{ij}^2 + \varepsilon_i \qquad i = 1, \dots, n$$

Die Zahl der zu schätzenden Parameter beträgt somit

$$1 + \mathrm{m} + \frac{m\,(m-1)}{2} + \mathrm{m} = \frac{(m+1)(m+2)}{2}$$

Als Beispiel dient hier ein Versuchsplan mit sechs Faktoren. Für die Anpassung des vollständigen Modells müssen also 28 Parameter geschätzt werden. Ein DSD verlangt 2m + 1 Experimente, für sechs Faktoren also 13. Daher ist das DSD für ein Modell mit Haupteffekten und quadratischen Termen bereits gesättigt. Man kann also nicht hergehen, komplexere Modelle schätzen und dann Terme mit vernachlässigbaren Parametern eliminieren. Genausowenig ist es sinnvoll, ein reines Haupteffektmodell zu schätzen, wenn es Effekte höherer Ordnung gibt oder geben könnte. Zwar werden die Parameter korrekt geschätzt, aber ihre Standardfehler werden durch die Wechselwirkungen oder quadratischen Effekte stark aufgebläht. Somit können tatsächlich wirksame Effekte als nicht signifikant erscheinen.

Parameterschätzer

Term	Schätzer	Std.-Fehler	t-Wert	Wahrsch. > |t|
Achsenabschnitt	94,875294	0,493569	192,22	<,0001*
X3	-2,201429	0,543886	-4,05	0,0019*
X4	-1,557143	0,543886	-2,86	0,0154*
X6	-2,93	0,543886	-5,39	0,0002*
X4*X6	4,643	0,595798	7,79	<,0001*
X3*X6	1,357	0,595798	2,28	0,0437*

Parameterschätzer

Term	Schätzer	Std.-Fehler	t-Wert	Wahrsch. > |t|
Achsenabschnitt	94,875294	1,37098	69,20	<,0001*
X3	-2,201429	1,510747	-1,46	0,1730
X4	-1,557143	1,510747	-1,03	0,3248
X6	-2,93	1,510747	-1,94	0,0785
X1	-0,027857	1,510747	-0,02	0,9856
X2	0,06	1,510747	0,04	0,9690

Abbildung 1: Veränderte Signifikanz durch Berücksichtigung von Wechselwirkungen

Die Schätzer für die wirksamen Haupteffekte X3, X4 und X6 sind in beiden Fällen identisch, aber bei dem reinen Haupteffektmodell (rechts) wird der Standardfehler fast dreimal so hoch geschätzt wie in dem Modell mit Wechselwirkungen. Keiner der Haupteffekt erscheint signifikant.

Eine gute Alternative ist die schrittweise Regression. Man spezifiziert das volle Oberflächenmodell und wählt Vorwärtsselektion mit Einschränkung, d.h. Terme höherer Ordnung gehen nur in das Modell ein, wenn die Haupteffekte, auf denen sie basieren, ebenfalls im Modell sind. Als Abbruchkriterium empfiehlt sich das AIC, das bei geringeren Fallzahlen besserer Ergebnisse liefert. Auch hier gibt es Einschränkungen: das Modell hat mehr Terme als der Datensatz Beobachtungen hat (im Beispiel 28/13), man kann also maximal 12 Terme schätzen. Außerdem werden Selektionsverfahren generell unzuverlässig, wenn die Zahl der Terme größer wird als n/2 (im Beispiel: 6). Simulationsstudien [2] zeigen, dass dieses Vorgehen dennoch eine hohe Power hat, wenn die tatsächlich einflussreichen Terme sich im Rahmen dieser Einschränkungen bewegen.

2 Lösungsansatz

DSDs sind ja bewusst so konstruiert, dass sie mit besonders wenigen Versuchen auskommen. Wenn dann aber mehrere Faktoren oder Effekte einflussreich sind, fehlt die Möglichkeit, die Fehlervarianz aus unabhängigen Beobachtungen zu schätzen. Dies kann nun erreicht werden, indem der Versuchsplan um zwei Dummy Faktoren erweitert wird, die inhaltlich keine Bedeutung für das Experiment haben. Dadurch wird die Versuchszahl moderat erhöht und die Auswertbarkeit des Versuchs deutlich gesteigert. Dummy Faktoren besitzen einige Vorzüge:

- sie sind orthogonal zu den Haupteffekten
- sie sind orthogonal zu allen Effekten zweiter Ordnung
- unter der Annahme, dass Effekte dritter und höherer Ordnung vernachlässigbar sind, können die durch sie induzierten Freiheitsgrade für eine erwartungstreue Schätzung der Fehlervarianz genutzt werden.

Die wesentliche Eigenschaft der DSDs ist, dass Haupteffekte und Effekte zweiter Ordnung voneinander unabhängig sind. Daher kann die Zielgröße (Y) in zwei unabhängige Größen aufgespalten werden:

- eine, die nur von den Haupteffekten abhängt (Yh)
- eine, die nur von den Effekten zweiter Ordnung abhängt (Yz)

Man erhält diese beiden Vektoren, indem man zunächst ein reines Haupteffektmodell ohne Konstante (!) für alle Faktoren inklusive der Dummy Faktoren anpasst. Dessen Vorhersagewerte bilden die neuen Zielgrößen Yh, die Residuen dieses Modells die Zielgrößen für die Effekte zweiter Ordnung Yz.

X1	X2	X3	X4	X5	X6	Dummy 1	Dummy 2	Y	Yh	Yz
0	1	1	1	1	1	1	1	94,51	-6,53	101,04
0	-1	-1	-1	-1	-1	-1	-1	107,57	6,53	101,04
1	0	1	1	-1	1	-1	-1	94,36	-6,81	101,18
-1	0	-1	-1	1	-1	1	1	107,99	6,81	101,18
1	-1	0	1	1	-1	1	-1	91,80	1,27	90,53
-1	1	0	-1	-1	1	-1	1	89,25	-1,27	90,53
1	-1	-1	0	1	1	-1	1	93,70	-0,78	94,49
-1	1	1	0	-1	-1	1	-1	95,27	0,78	94,49
1	1	-1	-1	0	1	1	-1	89,55	0,84	88,71
-1	-1	1	1	0	-1	-1	1	87,87	-0,84	88,71
1	-1	1	-1	-1	0	1	1	94,58	-0,66	95,24
-1	1	-1	1	1	0	-1	-1	95,89	0,66	95,24
1	1	-1	1	-1	-1	0	1	93,23	3,65	89,58
-1	-1	1	-1	1	1	0	-1	85,93	-3,65	89,58
1	1	1	-1	1	-1	-1	0	98,11	2,30	95,82
-1	-1	-1	1	-1	1	1	0	93,52	-2,30	95,82
0	0	0	0	0	0	0	0	99,75	0,00	99,75

Abbildung 2: Zerlegung in Hauptfaktoren und Terme zweiter Ordnung

In der Darstellung der Versuchsergebnisse sieht man gut, dass das DSD durch Faltung erzeugt wird. Die Zerlegung des Beobachtungsvektors Y in die orthogonalen Vektoren Yh und Yz zeigen sehr schön die paarweise Abhängigkeit der Ergebnisse.

Korrelationen

	Y	Yh	Yz
Y	1,0000	0,6223	0,7828
Yh	0,6223	1,0000	0,0000
Yz	0,7828	0,0000	1,0000

Abbildung 3: Korrelation der Zielgrößen

Die vier durch die Dummy Faktoren erzeugten Zusatzversuche liefern zwei Freiheitsgrade für die Schätzung der Fehlervarianz. Berechnet man die Korrelationen zwischen den Faktoren, zeigt sich die Unabhängigkeit der berechneten Vektoren.

2.1 Betrachtung der Haupteffekte Yh

Die Ergebnisse jedes Faltungspaares addieren sich zu Null. Das Ergebnis für den Mittelpunkt ist immer Null. Es gibt insgesamt 17 Beobachtungen, die aber nur 8 unabhängige Werte (= Freiheitsgrade) liefern. Sechs Faktoren sollen untersucht werden, 8 Freiheitsgrade stehen zur Verfügung, so bleiben 2 Freiheitsgrade für die Schätzung der Fehlervarianz.

2.2 Betrachtung der Effekte zweiter Ordnung Yz

Für jedes Faltungspaar sind die Ergebnisse identisch. Von den 17 Zeilen sind nur neun unabhängige Ergebnisse (= Freiheitsgrade). Nach der Schätzung des Achsenabschnitts bleiben noch acht Freiheitsgrade für die Schätzung von Effekten.

2.3 Die Schätzung

Die neue Zielgröße Yh wird nun herangezogen, um eine reines Haupteffektmodell nur für die tatsächlich interessierenden Faktoren anzupassen. Die Residuen aus dieser Anpassung haben, wie gezeigt, zwei Freiheitsgrade.

Term	Schätzer	t-Test berechnet
Achsenabschnitt	0,000	0,5000
X1	-0,028	0,4421
X2	0,060	0,6218
X3	-2,201	0,0029
X4	-1,557	0,0058
X5	0,011	0,5224
X6	-2,930	0,0017

Abbildung 4: t-Tests für Haupteffekte

Also kann man die Fehlervarianz schätzen, indem man die Summe der Residuenquadrate durch zwei teilt. Mit diesem Varianzschätzer führt man für jeden der Parameter einen t-Test durch und wählt die Effekte mit signifikanten Ergebnissen aus, für das Beispiel also X3, X4 und X6.

Für das weitere Vorgehen nehmen wir ein hierarchisches Modell an, d.h. dass Effekte höherer Ordnung nur dann einen Einfluss auf das Ergebnis haben, wenn sie aus signifikanten Haupteffekten zusammengesetzt sind. Das ist nicht mathematisch zwingend, aber in der Praxis häufig und besonders in Screening Situationen sinnvoll.
Nun bildet man alle Terme zweiter Ordnung und führt für Yz eine Regression über alle Teilmengen durch, bis die mittlere Fehlerquadratsumme nur unwesentlich größer ist als der berechnete Varianzschätzer. Für das Beispiel erweisen sich alle Terme zweiter Ordnung, die aus den Haupteffekten X3, X4 und X6 gebildet werden können, als signifikant. Das endgültige Modell mit den korrekten Schätzern für den Standardfehler setzt sich dann aus allen Haupteffekten und den signifikanten Termen höherer Ordnung zusammen. In dem Beispiel führt das zu folgender Parameterschätzung:

Parameterschätzer

Term	Schätzer	Std.-Fehler	t-Wert	Wahrsch. > |t|
Achsenabschnitt	99,807488	0,123141	810,51	<,0001*
X1	-0,027857	0,039271	-0,71	0,5173
X2	0,06	0,039271	1,53	0,2013
X3	-2,201429	0,039271	-56,06	<,0001*
X4	-1,557143	0,039271	-39,65	<,0001*
X5	0,0107143	0,039271	0,27	0,7985
X6	-2,93	0,039271	-74,61	<,0001*
X3*X3	-1,329698	0,106172	-12,52	0,0002*
X3*X4	1,0993372	0,047798	23,00	<,0001*
X4*X4	-2,172198	0,106172	-20,46	<,0001*
X3*X6	1,5343372	0,047798	32,10	<,0001*
X4*X6	4,6518372	0,047798	97,32	<,0001*
X6*X6	-2,487198	0,106172	-23,43	<,0001*

Abbildung 5: Endgültiges Modell
Jetzt werden die signifikanten Effekte auch korrekt identifiziert.

3 Empfehlung

Kein Versuchsplan kann alle denkbaren Aufgaben auf Anhieb erreichen. Automatische Modellselektionsverfahren versagen, wenn mehr als n/2 Terme aktiv sind. Die Power für die Entdeckung quadratischer Terme ist gering, für eine Power > 0,9 muss der Effekt größer als 3σ sein. Wenn die Analysen entsprechend an ihre Grenzen stoßen, sind zusätzlich Versuche zur Klärung notwendig. Für DSDs empfiehlt sich:

- anstelle zusätzlicher Mittelpunktversuche lieber Dummy Faktoren einfügen
- eine Hierarchie der Effekte unterstellen, solange keine starken sachlichen Gründe dagegen sprechen
- Aufteilen der der Zielgröße in zwei Vektoren für eine separate Anpassung der Haupteffekte und der Effekte zweiter Ordnung

Literatur

[1] Jones, Bradley and Nachtsheim, C. J. (2011) "A Class of Three-Level Designs for Definitive Screening in the Presence of Second-Order Effects" *Journal of Quality Technology*, **43**. 1-15.

[2] Philip Ramsey, Maria Weese, Douglas Montgomery, Model Selection Strategies for Definitive Screening Designs Using JMP Pro and R, JMP Discovery Summit 2015 San Diego (https://community.jmp.com/docs/DOC-7734)