Optimal design of experiments

Session 6: Nonstandard design problems

Peter Goos

Universiteit Antwerpen

Heterogeneous variances

- $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$
- $\text{var}(Y_i) = \text{var}(\epsilon_i) = \sigma_i^2$ instead of constant σ^2
- instead of ordinary least squares, use weighted (generalized) least squares
- $\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{y}$
  where
  $$\mathbf{V} = \begin{bmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_n^2 \end{bmatrix}$$
- $\text{var}(\hat{\boldsymbol{\beta}}) = \left(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}\right)^{-1}$

Heterogeneous variances

- thus, optimal design depends on $\mathbf{V}$
  $\Rightarrow$ you need prior information about $\sigma_1^2, \ldots, \sigma_n^2$
- $\mathbf{M} = \mathbf{X}^T \mathbf{V}^{-1} \mathbf{X}$
  $= \sum_{i=1}^n \frac{1}{\sigma_i^2} \mathbf{f}(\mathbf{x}_i) \mathbf{f}^T(\mathbf{x}_i)$
- update formulas for adding a point $\mathbf{a}$ to the design
  - $\left(\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X}\right)_{\text{NEW}} = \left(\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X}\right)_{\text{OLD}} + \frac{\mathbf{f}(\mathbf{a}) \mathbf{f}^T(\mathbf{a})}{\sigma_a^2}$
  - $\left|\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X}\right|_{\text{NEW}} =$

$$\left|\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X}\right|_{\text{OLD}} \times \left(1 + \underbrace{\frac{\mathbf{f}^T(\mathbf{a}) \left(\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X}\right) \mathbf{f}(\mathbf{a})}{\sigma_a^2}}_{\text{prediction variance is now weighted}}\right)$$

Heterogeneous variances

Example: quadratic regression in one variable

$$Y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \epsilon_i$$

where

$$\text{var}(\epsilon_i) = \sigma^2 \sqrt{x_i}$$
$$n = 6$$

- see quadratic1het.xls

Correlated observations

$$\mathbf{V} = \mathrm{var}(\mathbf{Y}) \neq \begin{bmatrix} \sigma^2 & & & 0 \\ & \sigma^2 & & \\ & & \sigma^2 & \\ 0 & & & \ddots \end{bmatrix}$$

$$\neq \begin{bmatrix} \sigma_1^2 & & & 0 \\ & \sigma_2^2 & & \\ & & \sigma_3^2 & \\ 0 & & & \ddots \end{bmatrix}$$

$$= \begin{bmatrix} \sigma^2 & \sigma_{12} & \sigma_{13} & \dots \\ \sigma_{12} & \sigma^2 & \sigma_{23} & \dots \\ \sigma_{13} & \sigma_{23} & \ddots & \\ \vdots & & & \end{bmatrix}$$

Correlated observations

- generalized least squares

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}\right)\mathbf{X}^T\mathbf{V}^{-1}\mathbf{y}$$

$$\mathrm{var}(\hat{\boldsymbol{\beta}}) = \left(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}\right)^{-1}$$

- no simple update formulas any more
- *order* becomes important
  order in which you run the observations

Serial correlation

- 3^2 factorial design
- AR(1) correlation pattern

$$\mathbf{V} = \begin{bmatrix} \sigma^2 & \rho\sigma^2 & \rho^2\sigma^2 & \dots & \rho^8\sigma^2 \\ \rho\sigma^2 & \sigma^2 & \rho\sigma^2 & \dots & \rho^7\sigma^2 \\ \vdots & \ddots & \ddots & & \vdots \\ \rho^8\sigma^2 & \rho^7\sigma^2 & \rho^6\sigma^2 & \dots & \sigma^2 \end{bmatrix}$$

- usually $0 < \rho < 1$
- Excel file `correlation.xls` contains two sheets with the same design
  what do you observe?

Blocked experiments

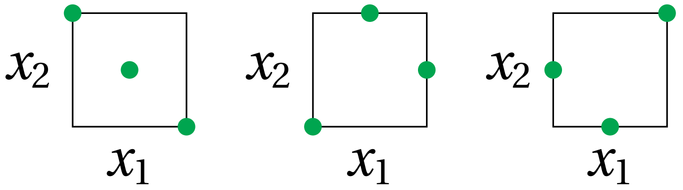
what? not all observations can be done in identical circumstances

e.g. more than one day
more than one batch
more than one operator
...
more than one block

→ response not just depends on experimental variables but also on block effects γ_i

e.g. $Y_{ij} = \beta_0 + \beta_1 x_{1j} + \beta_2 x_{ij}^2 + \gamma_i + \epsilon_{ij}$
$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\epsilon}$

Model for blocked experiments



$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

$$\mathbf{X} = \left[\begin{array}{ccc} 1 & -1 & +1 \\ 1 & 0 & 0 \\ 1 & +1 & -1 \\ \hline 1 & -1 & -1 \\ 1 & +1 & 0 \\ 1 & 0 & +1 \\ \hline 1 & -1 & 0 \\ 1 & 0 & -1 \\ 1 & +1 & +1 \end{array}\right] \qquad \mathbf{Z} = \left[\begin{array}{ccc} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ \hline 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ \hline 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{array}\right]$$

Orthogonal blocking

- definition orthogonal blocking
  "average level of regressors is the same in every block"
- consequence:
  incorporating the block effects in the model or not does not affect estimation of $\boldsymbol{\beta}$
- no information is lost because of blocking
- example: arrange 2^3 factorial design in 2 blocks

Orthogonally blocked 2^3 factorial design

$$
\begin{array}{rrrcrl}
-1 & -1 & -1 & \rightsquigarrow & -1 & \to \text{block } 1 \\
1 & -1 & -1 & & 1 & \to \text{block } 2 \\
-1 & 1 & -1 & & 1 & \to \text{block } 2 \\
1 & 1 & -1 & & -1 & \to \text{block } 1 \\
-1 & -1 & 1 & & 1 & \to \text{block } 2 \\
1 & -1 & 1 & & -1 & \to \text{block } 1 \\
-1 & 1 & 1 & & -1 & \to \text{block } 1 \\
1 & 1 & 1 & & 1 & \to \text{block } 2 \\
\downarrow & \downarrow & \downarrow & & & \\
x_1 & x_2 & x_3 & \rightsquigarrow & x_1 x_2 x_3 &
\end{array}
$$

Models for blocked experiments

MODEL 1

treat block effects as fixed:

- e.g. blocks are the machines you have
- you want conclusions just for those machines
- “intra-block analysis”

this was usually used in design literature (Atkinson & Donev (1989), Cook & Nachtsheim (1989))

MODEL 2

treat block effects as random:

- e.g. blocks are batches randomly drawn from warehouse
- you want conclusions for all batches
- “mixed model analysis” = “combined inter- & intra-block analysis” (Cheng)

this was done more recently (Goos & Vandebroek (2001))

Model 1: intra-block analysis

- $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\epsilon}$
- treat block effects as fixed
- estimate $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ using ordinary least squares
- $\begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\gamma}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^T\mathbf{X} & \mathbf{X}^T\mathbf{Z} \\ \mathbf{Z}^T\mathbf{X} & \mathbf{Z}^T\mathbf{Z} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}^T\mathbf{y} \\ \mathbf{Z}^T\mathbf{y} \end{bmatrix}$
- $\operatorname{var} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\gamma}} \end{bmatrix} = \sigma^2 \begin{bmatrix} \mathbf{X}^T\mathbf{X} & \mathbf{X}^T\mathbf{Z} \\ \mathbf{Z}^T\mathbf{X} & \mathbf{Z}^T\mathbf{Z} \end{bmatrix}^{-1}$
- D-optimal design maximizes $\det \begin{bmatrix} \mathbf{X}^T\mathbf{X} & \mathbf{X}^T\mathbf{Z} \\ \mathbf{Z}^T\mathbf{X} & \mathbf{Z}^T\mathbf{Z} \end{bmatrix}$

Optimal design for Model 1

- BLKL-algorithm: look what happens to the set of candidates
- run `block.exe` (fixed block effects)
  input file: `block.prn`
  - quadratic model in two variables
  - 3 blocks with 3 observations
- the projection of the three blocks on top of each other looks surprising

block.prn

```
 2 number of variables
 0 number of mixtures variables
 3 number of blocks
 9 number of observations
 3        3       3 number of observations in each block
 2 order of the model
 8 number of model parameters
 1        0   linear effect variable 1
 2        0   idem variable 2
 3        0   linear effect first dummy variable
 4        0   idem second dummy variable
 5        0   idem third dummy variable
 1        2   intercation effect
 1        1   quadratic effect variable 1
 2        2   quadratic effect variable 2
 3
 5
10
 0
 1
 0
```


Model 2: mixed model analysis

- treat block effects as random
- $\gamma_i \sim N(0, \sigma_\gamma^2)$
  all independent: $\boldsymbol{\gamma} \sim N(\mathbf{0}, \sigma_\gamma^2 \mathbf{I}_b)$
- $\epsilon_i \sim N(0, \sigma_\epsilon^2)$
  all independent: $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma_\epsilon^2 \mathbf{I}_n)$
- $\mathbf{V} = \mathrm{var}(\mathbf{Y}) = \sigma_\epsilon^2 \mathbf{I}_n + \sigma_\gamma^2 \mathbf{Z}\mathbf{Z}^T$
- $\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X}\right)^{-1} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{y}$
- $\mathrm{var}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1}$
- note that $\mathbf{V} = \sigma_\epsilon^2 \left(I_n + \underbrace{\frac{\sigma_\gamma^2}{\sigma_\epsilon^2}}_{\eta} \mathbf{Z}\mathbf{Z}^T\right) = \sigma_\epsilon^2 \mathbf{D}_\eta$
- $\mathrm{var}(\hat{\boldsymbol{\beta}}) = \sigma_\epsilon^2 (\mathbf{X}^T \mathbf{D}_\eta^{-1} \mathbf{X})^{-1}$

Optimal designs for Model 2

- run `blklbis.exe` (for random block effects)
  input file: `block1.prn`
  - 3 blocks of 3 observations
  - quadratic model in 2 variables
  - optimal designs are computed for two values of η, 1 and 10
- have a look at the projections now too
- $\eta = 1$ yields different design than $\eta = 10$
- $\eta = 10$ yields same design as BLKL-algorithm (fixed block effects)

block1.prn

```
2 number of variables
9 observations
3 blocks
1 all blocks have same number of observations
2 eta values
1. 10.    these are the two eta values for which you want a d
2         order of the model
6         number of beta-parameters
0 0       intercept
1 0       linear term in variable 1
2 0       linear term in variable 2
1 2       interaction
1 1       quadr var 1
2 2       quadr var 2
100 number of tries, the rest is technical stuff
100
100
0
0
0
```

Optimal designs for Model 2

$$\mathbf{V} = \sigma_\epsilon^2(\mathbf{I}_n + \eta \mathbf{Z}\mathbf{Z}^T)$$

$$= \sigma_\epsilon^2 \begin{bmatrix} 1+\eta & \eta & \eta & 0 & 0 & 0 & 0 & 0 & 0 \\ \eta & 1+\eta & \eta & 0 & 0 & 0 & 0 & 0 & 0 \\ \eta & \eta & 1+\eta & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1+\eta & \eta & \eta & 0 & 0 & 0 \\ 0 & 0 & 0 & \eta & 1+\eta & \eta & 0 & 0 & 0 \\ 0 & 0 & 0 & \eta & \eta & 1+\eta & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1+\eta & \eta & \eta \\ 0 & 0 & 0 & 0 & 0 & 0 & \eta & 1+\eta & \eta \\ 0 & 0 & 0 & 0 & 0 & 0 & \eta & \eta & 1+\eta \end{bmatrix}$$

Another example

run `block.exe`

input file: `block2.prn`

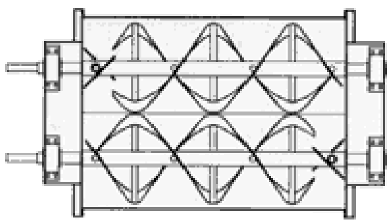
- 2 blocks of 4 observations
- 3 variables
- linear effects + two-factor interactions
- what would a researcher do when (s)he had never heard of optimal design?
- $\eta = 0.01$, $\eta = 1$, $\eta = 10$

block2.prn

```
3 variables
8 eight observations
2 two blocks
1 number of observations in each block is the same
3 number of eta's
0.01 1. 10.    eta-values
2       order of the model
7       number of beta parameters
0 0       intercept
1 0       lin 1
2 0       lin 2
3 0       lin 3
1 2       interaction 1 2
1 3       int 1 3
2 3       int 2 3
50       number of tries, the rest is technical stuff
100
100
0
0
0
```


Pastry dough experiment

(Trinca & Gilmour 2000)



- 7 days of experimentation
- 4 runs/day

purpose: increase quality

3 factors:
- initial moisture content (18–21–24%)
- feed flow rate (30–37.5–45 tons/h)
- screw speed (300–350–400 rpm)

Paired comparison experiments

option I	option II
sloping frame	classic frame
Shimano	Campagnolo
Campagnolo Hyperon	Mavic Ksyrium

↑
slider

Paired comparison experiments

- only one response is recorded for every set of options
- this is essentially a comparison within a block
  - → "intra-block analysis" like in the fixed block effects case
  - → BLKL-algorithm can be used to design paired comparison experiments
- run `blkbis.exe`
  input file: `paired.prn`

paired.prn

```
2             number of expl variables
0             no mixture variables
4             four blocks
8             eight observations
2   2   2    2        all blocks have 2 observations
2             order of the model
7             number of parameters
1       0     linear term expl var 1
2       0     linear term expl var 2
1       2     interaction
3       0     first block effect
4       0     second
5       0     third
6       0     fourth block effect
3     k
5     l
10    number of times you try
0     technical stuff
1
0
```


Hard-to-change variables

- doing experiments is time-consuming
- especially when some of the factors are "hard" to change
- example: temperature of a furnace

$$\begin{gathered}\text{high temperature}\\ \downarrow\\ \text{low temperature}\\ \downarrow\\ \text{intermediate temperature}\\ \downarrow\\ \text{low temperature}\\ \vdots\end{gathered}$$

Hard-to-change variables

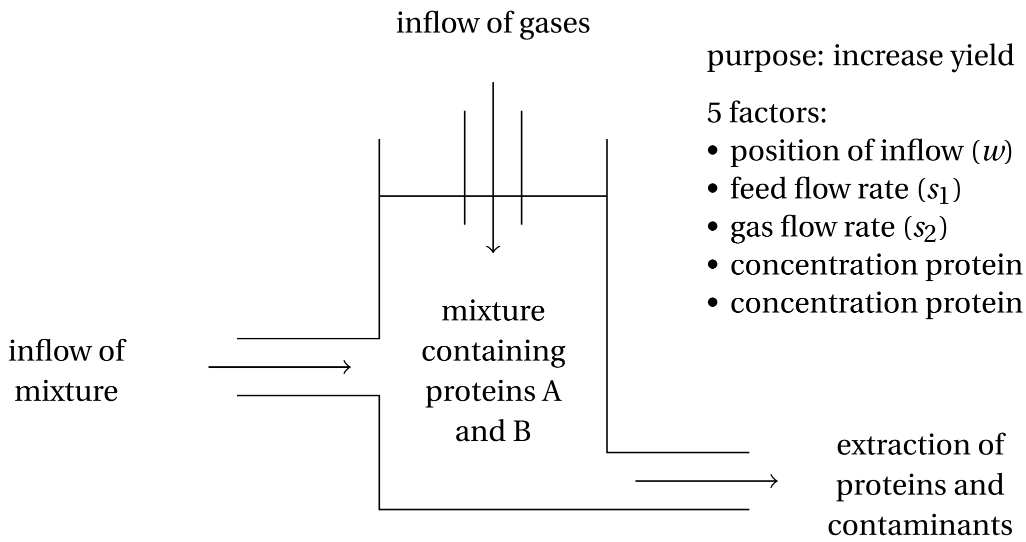
- researchers don't want to change the levels of such factors all the time
  - e.g. keep temperatures fixed during a day and try other temperature on other days
- such experiments resemble blocked experiments
  why? observations are partitioned in groups
- the difference is that the level(s) of hard-to-change variable(s) is (are) held constant in every group

Hard-to-change variables

- resulting designs are called split-plot designs
- groups: whole plots, main plots
- hard-to-change variable(s)
  = whole-plot variable(s)
- other variables
  = sub-plot variables
- D-optimal split-plot designs can be generated using JMP software

Protein extraction experiment

(Trinca & Gilmour 2001)



purpose: increase yield

5 factors:
- position of inflow (w)
- feed flow rate (s_1)
- gas flow rate (s_2)
- concentration protein A (s_3)
- concentration protein B (s_4)

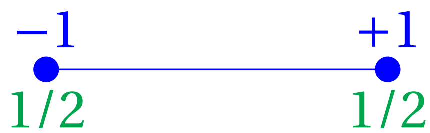
- 2 runs/day
- 21 days of experimentation
- full quadratic model in the 5 variables

Model uncertainty

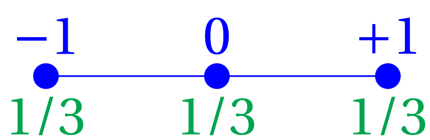
- a lot of work has been done on that topic
- two competing models:

$$Y = \beta_0 + \beta_1 x + \epsilon$$
$$Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \epsilon$$

- linear model



- quadratic model



Model uncertainty

- some trade-off between the optimal design for the linear model and the quadratic model
- a simple approach is the Bayesian approach proposed by Dumouchel & Jones (1995)
- idea:
  - for some of the model terms, you know "for sure" that they are in the model: $\boldsymbol{\beta}_1$
  - for other terms you are not so sure: $\boldsymbol{\beta}_2$

Bayesian D-optimal designs

- Bayesian idea: use a prior distribution for β's for which we're not so sure

$$\boldsymbol{\beta}_2 \sim \text{NORMAL}(\ \mathbf{0}\ ,\ \tau^2\sigma^2\mathbf{I}\ )$$

this means: "I think these parameters are not in the model"
or "I think their effects are zero"

this means: "I'm not totally sure of that"

the smaller τ^2, the less likely $\boldsymbol{\beta}_2$ is in the model
the larger τ^2, the more likely $\boldsymbol{\beta}_2$ is in the model

Bayesian estimator versus OLS

Ordinary least squares

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

$$\begin{bmatrix} \hat{\boldsymbol{\beta}}_1 \\ \hat{\boldsymbol{\beta}}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1^T\mathbf{X}_1 & \mathbf{X}_1^T\mathbf{X}_2 \\ \mathbf{X}_2^T\mathbf{X}_1 & \mathbf{X}_2^T\mathbf{X}_2 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}_1^T\mathbf{y} \\ \mathbf{X}_2^T\mathbf{y} \end{bmatrix}$$

Bayesian approach

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X} + \mathbf{K})^{-1}\mathbf{X}^T\mathbf{y}$$

$$\begin{bmatrix} \hat{\boldsymbol{\beta}}_1 \\ \hat{\boldsymbol{\beta}}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1^T\mathbf{X}_1 & \mathbf{X}_1^T\mathbf{X}_2 \\ \mathbf{X}_2^T\mathbf{X}_1 & \mathbf{X}_2^T\mathbf{X}_2 + \frac{\mathbf{I}}{\tau^2} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{X}_1^T\mathbf{y} \\ \mathbf{X}_2^T\mathbf{y} \end{bmatrix}$$

(this is mean of posterior distribution of $\boldsymbol{\beta}$)

Bayesian D-optimality criterion

Classical D-optimality

$$\text{maximize det} \begin{bmatrix} \mathbf{X}_1^T\mathbf{X}_1 & \mathbf{X}_1^T\mathbf{X}_2 \\ \mathbf{X}_2^T\mathbf{X}_1 & \mathbf{X}_2^T\mathbf{X}_2 \end{bmatrix}$$

Bayesian D-optimality

$$\text{maximize det} \begin{bmatrix} \mathbf{X}_1^T\mathbf{X}_1 & \mathbf{X}_1^T\mathbf{X}_2 \\ \mathbf{X}_2^T\mathbf{X}_1 & \mathbf{X}_2^T\mathbf{X}_2 + \frac{\mathbf{I}}{\tau^2} \end{bmatrix}$$

Illustration

- $Y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + \epsilon_i$
- primary terms: β_0, $\beta_1 x_i$
- potential term: $\beta_2 x_i^2$, $\beta_2 \sim \text{NORMAL}(0, \tau^2\sigma^2)$
- $\hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix} = \begin{bmatrix} n & \sum x_i & \sum x_i^2 \\ \sum x_i & \sum x_i^2 & \sum x_i^3 \\ \sum x_i^2 & \sum x_i^3 & \sum x_i^4 + \frac{1}{\tau^2} \end{bmatrix}^{-1} \mathbf{X}^T y$
- Bayesian D-optimal design

$$\text{maximizes} \begin{vmatrix} n & \sum x_i & \sum x_i^2 \\ \sum x_i & \sum x_i^2 & \sum x_i^3 \\ \sum x_i^2 & \sum x_i^3 & \sum x_i^4 + \frac{1}{\tau^2} \end{vmatrix}$$

→ see file `bayesian1.sas`

⇝ uses `prior` option in the `optex` procedure

bayesian1.sas

```
proc factex;
factors x / nlev = 3;
output out=can x nvals = (-1 0 1);
proc print data=can;
run;
proc optex data=can seed=57922;
model x, x*x / prior = 0, 10;
generate n=30 method=fedorov;
output out=des;
proc print data=des;
run;
```