

Post-selection inference

[selcorr & sciCORE]

Marco Cattaneo

Department of Clinical Research
University of Basel

8 February 2022

selcorr (R package)

- ▶ main issue of variable selection: naive inference after selection is usually wrong

selcorr (R package)

- ▶ main issue of variable selection: naive inference after selection is usually wrong
- ▶ `selcorr` solves this issue for (generalized) linear models and some standard selection procedures

selcorr (R package)

- ▶ main issue of variable selection: naive inference after selection is usually wrong
- ▶ selcorr solves this issue for (generalized) linear models and some standard selection procedures
- ▶ originally planned to use asymptotic distribution of likelihood ratios, but in realistic examples computation slow and asymptotic approximation poor: rough approximation of asymptotic distribution (fast) and bootstrap calibration (slow) performed better and are implemented

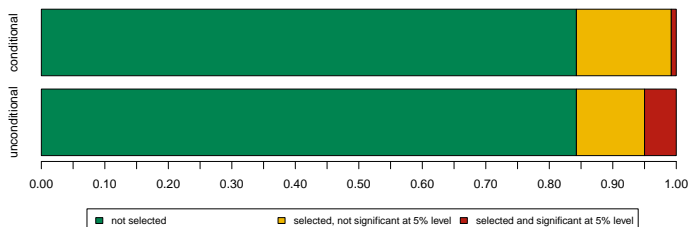
selcorr (R package)

- ▶ main issue of variable selection: naive inference after selection is usually wrong
- ▶ selcorr solves this issue for (generalized) linear models and some standard selection procedures
- ▶ originally planned to use asymptotic distribution of likelihood ratios, but in realistic examples computation slow and asymptotic approximation poor: rough approximation of asymptotic distribution (fast) and bootstrap calibration (slow) performed better and are implemented
- ▶ selcorr performs unconditional post-selection inference: if a variable has no effect, not selecting it is considered a correct inference

selcorr (R package)

- ▶ main issue of variable selection: naive inference after selection is usually wrong
- ▶ selcorr solves this issue for (generalized) linear models and some standard selection procedures
- ▶ originally planned to use asymptotic distribution of likelihood ratios, but in realistic examples computation slow and asymptotic approximation poor: rough approximation of asymptotic distribution (fast) and bootstrap calibration (slow) performed better and are implemented
- ▶ selcorr performs unconditional post-selection inference: if a variable has no effect, not selecting it is considered a correct inference

e.g. typical probabilities for a variable without effect:



example (qu19Egli)

```
> full.model = glm(paste("antibioticsadministered ~", paste(names(FAS)[-1], collapse = " + ")),
+                 binomial, FAS)
> dim(FAS)
[1] 807 26
> summary(full.model)
[...]
```

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	13.876076	842.547158	0.016	0.986860
age	0.000762	0.006125	0.124	0.900994
sexfemale	-0.069330	0.229684	-0.302	0.762768
length.of.stay	0.065690	0.020096	3.269	0.001080 **
hospital.careoutpatient	-1.270438	0.320324	-3.966	7.31e-05 ***
ILI.symptomsyes	0.214110	0.260737	0.821	0.411549
result.PCRinfluenzaA	-0.842044	0.268554	-3.135	0.001716 **
result.PCRinfluenzaB	-1.145874	0.345192	-3.320	0.000902 ***
result.blood.culturebacterialgrowth	0.298323	0.473529	0.630	0.528695
result.blood.cultureNA	-0.645101	0.244117	-2.643	0.008228 **
result.urine.culturebacterialgrowth	-0.386222	0.380683	-1.015	0.310320
result.urine.cultureNA	-0.816184	0.364451	-2.239	0.025124 *
result.antigen.testspostive	13.177484	648.203994	0.020	0.983781
result.antigen.testsnA	-1.145332	0.278215	-4.117	3.84e-05 ***
result.infection.testsbacterialgrowth	-12.562509	842.546773	-0.015	0.988104
result.infection.testsnA	-13.427966	842.546677	-0.016	0.987284
log.leucocytes	0.037695	0.520235	0.072	0.942238
log.c.react.protein	0.930584	0.195176	4.768	1.86e-06 ***
log.procalcitonin	1.763365	0.349039	5.052	4.37e-07 ***
log.procalcitonin.NATRUE	-0.409170	0.245360	-1.668	0.095389 .
log.UF.leucocytes	0.795687	0.224735	3.541	0.000399 ***
log.UF.bacteria	-0.185919	0.145300	-1.280	0.200703
log.UF.bacteria.NATRUE	0.329833	0.279812	1.179	0.238492
pneumoniapossible	1.167533	0.490400	2.381	0.017276 *
pneumoniayes	3.086041	0.898448	3.435	0.000593 ***
pneumoniaNA	0.022053	0.242929	0.091	0.927667

```
[...]
```

example (qu19Egli)

```
> summary(step(full.model, direction = "both"))
```

```
[...]
```

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	1.30900	0.59225	2.210	0.027092	*
length.of.stay	0.06451	0.01978	3.262	0.001108	**
hospital.careoutpatient	-1.19428	0.29471	-4.052	5.07e-05	***
result.PCRinfluenzaA	-0.79372	0.24385	-3.255	0.001134	**
result.PCRinfluenzaB	-1.02456	0.32432	-3.159	0.001582	**
result.blood.cultureNA	-0.73085	0.23644	-3.091	0.001994	**
result.urine.cultureNA	-0.40371	0.21843	-1.848	0.064570	.
result.antigen.testsNA	-1.17708	0.27338	-4.306	1.67e-05	***
result.infection.testsNA	-0.88555	0.41952	-2.111	0.034785	*
log.c.react.protein	0.94395	0.18924	4.988	6.10e-07	***
log.procalcitonin	1.78004	0.34219	5.202	1.97e-07	***
log.procalcitonin.NATRU	-0.42594	0.23696	-1.798	0.072250	.
log.UF.leucocytes	0.81857	0.22204	3.686	0.000227	***
log.UF.bacteria	-0.24014	0.13145	-1.827	0.067725	.
pneumoniapossible	1.25842	0.46081	2.731	0.006317	**
pneumoniayes	3.00205	0.85532	3.510	0.000448	***

```
[...]
```


example (qu19Egli)

```
> summary(step(full.model, direction = "both"))
[...]
```

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	1.30900	0.59225	2.210	0.027092	*
length.of.stay	0.06451	0.01978	3.262	0.001108	**
hospital.careoutpatient	-1.19428	0.29471	-4.052	5.07e-05	***
result.PCRinfluenzaA	-0.79372	0.24385	-3.255	0.001134	**
result.PCRinfluenzaB	-1.02456	0.32432	-3.159	0.001582	**
result.blood.cultureNA	-0.73085	0.23644	-3.091	0.001994	**
result.urine.cultureNA	-0.40371	0.21843	-1.848	0.064570	.
result.antigen.testsNA	-1.17708	0.27338	-4.306	1.67e-05	***
result.infection.testsNA	-0.88555	0.41952	-2.111	0.034785	*
log.c.react.protein	0.94395	0.18924	4.988	6.10e-07	***
log.procalcitonin	1.78004	0.34219	5.202	1.97e-07	***
log.procalcitonin.NATRUe	-0.42594	0.23696	-1.798	0.072250	.
log.UF.leucocytes	0.81857	0.22204	3.686	0.000227	***
log.UF.bacteria	-0.24014	0.13145	-1.827	0.067725	.
pneumoniapossible	1.25842	0.46081	2.731	0.006317	**
pneumoniayes	3.00205	0.85532	3.510	0.000448	***

```
[...]
```

```
> selcorr(full.model)
```

	Estimate	CI lower	CI upper	p
(Intercept)	1.30899792	0.17088146	2.49995322	2.383593e-02
length.of.stay	0.06451065	0.02758028	0.10521110	4.082291e-04
hospital.careoutpatient	-1.19428438	-1.78823265	-0.62072298	3.664392e-05
result.PCRinfluenzaA	-0.79372284	-1.27711683	-0.31974667	9.836783e-04
result.PCRinfluenzaB	-1.02456274	-1.67551532	-0.40111090	1.137154e-03
result.blood.cultureNA	-0.73085326	-1.19808941	-0.26968448	1.869773e-03
result.urine.cultureNA	-0.40370917	-0.83175343	0.02571881	6.531766e-02
result.antigen.testsNA	-1.17707602	-1.72737660	-0.64583125	1.078679e-05
result.infection.testsNA	-0.88555294	-2.15262937	0.22112554	1.203560e-01
log.c.react.protein	0.94394906	0.57863340	1.32164194	2.508300e-07
log.procalcitonin	1.78003675	1.12488947	2.50224866	6.236563e-09
log.procalcitonin.NATRUe	-0.42593871	-0.89283509	0.03757126	7.169846e-02
log.UF.leucocytes	0.81856587	0.38344459	1.27586653	1.790322e-04
log.UF.bacteria	-0.24014011	-0.51217098	0.02823039	7.956539e-02
pneumoniapossible	1.25841965	0.38625980	2.20532074	4.191692e-03
pneumoniayes	3.00205186	1.52178766	4.99707270	7.222524e-06

?selcorr

Post-Selection Inference for Generalized Linear Models

Description:

selcorr returns (unconditional) post-selection confidence intervals and p-values for the coefficients of (generalized) linear models.

Usage:

```
selcorr(  
  object,  
  fixed.vars = NULL,  
  further.vars = NULL,  
  boot.repl = 0,  
  k = 2,  
  conf.level = 0.95,  
  quiet = FALSE  
)
```

Arguments:

object: an object representing a model of an appropriate class. This is used as the initial model in a (bidirectional) stepwise model selection.

fixed.vars: the names of all independent variables that must be included in the selected model. The default is none.

further.vars: the names of all independent variables that can be included in the selected model, but are not part of object. The default is none.

boot.repl: a number or list of bootstrap replicates. The default is no bootstrapping. See Details and Examples for clarification.

k: the multiple of the number of degrees of freedom used as penalty in the model selection. The default $k = 2$ corresponds to the AIC.

conf.level: the level of the confidence intervals.

quiet: if TRUE, then selcorr does not generate an output.

?selcorr

Details:

When `boot.repl = 0`, an approximate asymptotic distribution of the test statistic is used to calculate p-values and calibrate the profile-likelihood confidence intervals. This approach is faster, but p-values and confidence intervals can be more precisely calibrated by parametrically bootstrapping the test statistic (with `boot.repl` the number of replicates). Parallel computing can be used to speed up the bootstrapping: see Examples.

Value:

the selected model is returned, without correction for model-selection, but with up to two additional components. There is an output component corresponding to the post-selection inference, which is also printed unless `quiet = TRUE`. When `boot.repl` is not 0, there is also a `boot.repl` component corresponding to the bootstrap replicates.

Examples:

```
## linear regression:
selcorr(lm(Fertility ~ ., swiss))

## logistic regression:
swiss.lr = within(swiss, Fertility <- (Fertility > 70))
selcorr(glm(Fertility ~ ., binomial, swiss.lr))

## parallel bootstrapping:
## Not run:

library(future.apply)
plan(multisession)
boot.repl = future_replicate(8, selcorr(lm(Fertility ~ ., swiss), boot.repl = 1000,
                                       quiet = TRUE)$boot.repl, simplify = FALSE)

plan(sequential)
selcorr(lm(Fertility ~ ., swiss), boot.repl = do.call("rbind", boot.repl))
## End(Not run)
```

sciCORE (HPC)

- ▶ already had a sciCORE account (otherwise to be created: see support.scicore.unibas.ch)

sciCORE (HPC)

- ▶ already had a sciCORE account (otherwise to be created: see support.scicore.unibas.ch)
- ▶ installed and started VPN (see its.unibas.ch/en/manuals/network-access/instructions-vpn)

sciCORE (HPC)

- ▶ already had a sciCORE account (otherwise to be created: see support.scicore.unibas.ch)
- ▶ installed and started VPN (see its.unibas.ch/en/manuals/network-access/instructions-vpn)
- ▶ added sciCORE storage as network drive in Nautilus (Other Locations / Connect to Server / `ssh://username@login.scicore.unibas.ch`)

sciCORE (HPC)

- ▶ already had a sciCORE account (otherwise to be created: see support.scicore.unibas.ch)
- ▶ installed and started VPN (see its.unibas.ch/en/manuals/network-access/instructions-vpn)
- ▶ added sciCORE storage as network drive in Nautilus (Other Locations / Connect to Server / `ssh://username@login.scicore.unibas.ch`)
- ▶ logged in (in terminal: `ssh username@login.scicore.unibas.ch`)

sciCORE (HPC)

- ▶ already had a sciCORE account (otherwise to be created: see support.scicore.unibas.ch)
- ▶ installed and started VPN (see its.unibas.ch/en/manuals/network-access/instructions-vpn)
- ▶ added sciCORE storage as network drive in Nautilus (Other Locations / Connect to Server / `ssh://username@login.scicore.unibas.ch`)
- ▶ logged in (in terminal: `ssh username@login.scicore.unibas.ch`)
- ▶ loaded R (ml load R), started R (R), and installed required packages (alternative: use rstudio.scicore.unibas.ch)

sciCORE (HPC)

- ▶ already had a sciCORE account (otherwise to be created: see `support.scicore.unibas.ch`)
- ▶ installed and started VPN (see `its.unibas.ch/en/manuals/network-access/instructions-vpn`)
- ▶ added sciCORE storage as network drive in Nautilus (Other Locations / Connect to Server / `ssh://username@login.scicore.unibas.ch`)
- ▶ logged in (in terminal: `ssh username@login.scicore.unibas.ch`)
- ▶ loaded R (`m1 load R`), started R (R), and installed required packages (alternative: use `rstudio.scicore.unibas.ch`)
- ▶ run test code (`Rscript -vanilla testcode.R > testcode.Rout &`), but very slow (to estimate required time better use own computer)

sciCORE (HPC)

- ▶ already had a sciCORE account (otherwise to be created: see support.scicore.unibas.ch)
- ▶ installed and started VPN (see its.unibas.ch/en/manuals/network-access/instructions-vpn)
- ▶ added sciCORE storage as network drive in Nautilus (Other Locations / Connect to Server / `ssh://username@login.scicore.unibas.ch`)
- ▶ logged in (in terminal: `ssh username@login.scicore.unibas.ch`)
- ▶ loaded R (ml load R), started R (R), and installed required packages (alternative: use rstudio.scicore.unibas.ch)
- ▶ run test code (`Rscript -vanilla testcode.R > testcode.Rout &`), but very slow (to estimate required time better use own computer)
- ▶ wrote SLURM script (see wiki.biozentrum.unibas.ch/display/scicore/SLURM+user+guide), replacing parallel computing in R with array job (100 times 1-core R code, changing only seed and output)

sciCORE (HPC)

- ▶ already had a sciCORE account (otherwise to be created: see `support.scicore.unibas.ch`)
- ▶ installed and started VPN (see `its.unibas.ch/en/manuals/network-access/instructions-vpn`)
- ▶ added sciCORE storage as network drive in Nautilus (Other Locations / Connect to Server / `ssh://username@login.scicore.unibas.ch`)
- ▶ logged in (in terminal: `ssh username@login.scicore.unibas.ch`)
- ▶ loaded R (ml load R), started R (R), and installed required packages (alternative: use `rstudio.scicore.unibas.ch`)
- ▶ run test code (`Rscript -vanilla testcode.R > testcode.Rout &`), but very slow (to estimate required time better use own computer)
- ▶ wrote SLURM script (see `wiki.biozentrum.unibas.ch/display/scicore/SLURM+user+guide`), replacing parallel computing in R with array job (100 times 1-core R code, changing only seed and output)
- ▶ run SLURM script (`sbatch script.sh`), which immediately started 100 jobs for 24 hours (checking with: `squeue -u username`), potentially costing $100 \cdot 24 \cdot 0.007 = 16.8$ CHF (if team goes above 1000 CHF during year)

sciCORE (HPC)

- ▶ already had a sciCORE account (otherwise to be created: see `support.scicore.unibas.ch`)
- ▶ installed and started VPN (see `its.unibas.ch/en/manuals/network-access/instructions-vpn`)
- ▶ added sciCORE storage as network drive in Nautilus (Other Locations / Connect to Server / `ssh://username@login.scicore.unibas.ch`)
- ▶ logged in (in terminal: `ssh username@login.scicore.unibas.ch`)
- ▶ loaded R (`m1 load R`), started R (`R`), and installed required packages (alternative: use `rstudio.scicore.unibas.ch`)
- ▶ run test code (`Rscript -vanilla testcode.R > testcode.Rout &`), but very slow (to estimate required time better use own computer)
- ▶ wrote SLURM script (see `wiki.biozentrum.unibas.ch/display/scicore/SLURM+user+guide`), replacing parallel computing in R with array job (100 times 1-core R code, changing only seed and output)
- ▶ run SLURM script (`sbatch script.sh`), which immediately started 100 jobs for 24 hours (checking with: `squeue -u username`), potentially costing $100 * 24 * 0.007 = 16.8$ CHF (if team goes above 1000 CHF during year)
- ▶ out of 100 jobs, 7 were killed when reaching the memory limit, while the others were killed as planned after 24 hours (`m1 purge`; `m1 load jobstats`; `jobstats -S mm/dd/yy`)

simulation results

- ▶ logistic model: 7085 simulations of dependent variable, 500 data points, 28 covariates (14 with an effect), of which 14 binary (7 with an effect), of which 2 fixed (1 with an effect), correlation of covariates ranging from -0.25 to 0.77 with median 0.09, absolute standardized coefficients (for the covariates with an effect) ranging from 0.02 to 7.81 with median 0.28

simulation results

- ▶ logistic model: 7085 simulations of dependent variable, 500 data points, 28 covariates (14 with an effect), of which 14 binary (7 with an effect), of which 2 fixed (1 with an effect), correlation of covariates ranging from -0.25 to 0.77 with median 0.09, absolute standardized coefficients (for the covariates with an effect) ranging from 0.02 to 7.81 with median 0.28
- ▶ 95% confidence intervals (CI): 1 - probability of non-coverage (i.e. probability of no error), in % (should be 95)

simulation results

- ▶ logistic model: 7085 simulations of dependent variable, 500 data points, 28 covariates (14 with an effect), of which 14 binary (7 with an effect), of which 2 fixed (1 with an effect), correlation of covariates ranging from -0.25 to 0.77 with median 0.09, absolute standardized coefficients (for the covariates with an effect) ranging from 0.02 to 7.81 with median 0.28
- ▶ 95% confidence intervals (CI): 1 - probability of non-coverage (i.e. probability of no error), in % (should be 95)
- ▶ level: probability of significance at 5% level for variables without effect (i.e. probability of type I error), in % (should be 5)

simulation results

- ▶ logistic model: 7085 simulations of dependent variable, 500 data points, 28 covariates (14 with an effect), of which 14 binary (7 with an effect), of which 2 fixed (1 with an effect), correlation of covariates ranging from -0.25 to 0.77 with median 0.09, absolute standardized coefficients (for the covariates with an effect) ranging from 0.02 to 7.81 with median 0.28
- ▶ 95% confidence intervals (CI): 1 - probability of non-coverage (i.e. probability of no error), in % (should be 95)
- ▶ level: probability of significance at 5% level for variables without effect (i.e. probability of type I error), in % (should be 5)
- ▶ power: probability of significance at 5% level for variables with an effect (i.e. probability of no type II error), in %

simulation results

- ▶ logistic model: 7085 simulations of dependent variable, 500 data points, 28 covariates (14 with an effect), of which 14 binary (7 with an effect), of which 2 fixed (1 with an effect), correlation of covariates ranging from -0.25 to 0.77 with median 0.09, absolute standardized coefficients (for the covariates with an effect) ranging from 0.02 to 7.81 with median 0.28
- ▶ 95% confidence intervals (CI): 1 - probability of non-coverage (i.e. probability of no error), in % (should be 95)
- ▶ level: probability of significance at 5% level for variables without effect (i.e. probability of type I error), in % (should be 5)
- ▶ power: probability of significance at 5% level for variables with an effect (i.e. probability of no type II error), in %
- ▶ time: using 1 core of a standard laptop (1.80GHz)

simulation results

- ▶ logistic model: 7085 simulations of dependent variable, 500 data points, 28 covariates (14 with an effect), of which 14 binary (7 with an effect), of which 2 fixed (1 with an effect), correlation of covariates ranging from -0.25 to 0.77 with median 0.09, absolute standardized coefficients (for the covariates with an effect) ranging from 0.02 to 7.81 with median 0.28
- ▶ 95% confidence intervals (CI): 1 - probability of non-coverage (i.e. probability of no error), in % (should be 95)
- ▶ level: probability of significance at 5% level for variables without effect (i.e. probability of type I error), in % (should be 5)
- ▶ power: probability of significance at 5% level for variables with an effect (i.e. probability of no type II error), in %
- ▶ time: using 1 core of a standard laptop (1.80GHz)

	CI			level			power			time
	min	median	max	min	median	max	min	median	max	median
full (glm+confint)	71.2	92.1	92.8	6.5	7.0	7.4	6.9	59.0	100	0.7 sec

simulation results

- ▶ logistic model: 7085 simulations of dependent variable, 500 data points, 28 covariates (14 with an effect), of which 14 binary (7 with an effect), of which 2 fixed (1 with an effect), correlation of covariates ranging from -0.25 to 0.77 with median 0.09, absolute standardized coefficients (for the covariates with an effect) ranging from 0.02 to 7.81 with median 0.28
- ▶ 95% confidence intervals (CI): 1 - probability of non-coverage (i.e. probability of no error), in % (should be 95)
- ▶ level: probability of significance at 5% level for variables without effect (i.e. probability of type I error), in % (should be 5)
- ▶ power: probability of significance at 5% level for variables with an effect (i.e. probability of no type II error), in %
- ▶ time: using 1 core of a standard laptop (1.80GHz)

	CI			level			power			time
	min	median	max	min	median	max	min	median	max	median
full (glm+confint)	71.2	92.1	92.8	6.5	7.0	7.4	6.9	59.0	100	0.7 sec
naive (step+confint)	75.9	88.0	93.0	8.8	10.6	18.0	8.6	73.7	100	2.0 sec

simulation results

- ▶ logistic model: 7085 simulations of dependent variable, 500 data points, 28 covariates (14 with an effect), of which 14 binary (7 with an effect), of which 2 fixed (1 with an effect), correlation of covariates ranging from -0.25 to 0.77 with median 0.09, absolute standardized coefficients (for the covariates with an effect) ranging from 0.02 to 7.81 with median 0.28
- ▶ 95% confidence intervals (CI): 1 - probability of non-coverage (i.e. probability of no error), in % (should be 95)
- ▶ level: probability of significance at 5% level for variables without effect (i.e. probability of type I error), in % (should be 5)
- ▶ power: probability of significance at 5% level for variables with an effect (i.e. probability of no type II error), in %
- ▶ time: using 1 core of a standard laptop (1.80GHz)

	CI			level			power			time
	min	median	max	min	median	max	min	median	max	median
full (glm+confint)	71.2	92.1	92.8	6.5	7.0	7.4	6.9	59.0	100	0.7 sec
naive (step+confint)	75.9	88.0	93.0	8.8	10.6	18.0	8.6	73.7	100	2.0 sec
selcorr	79.4	92.6	94.3	6.8	7.4	9.5	8.3	66.6	100	7.7 sec

simulation results

- ▶ logistic model: 7085 simulations of dependent variable, 500 data points, 28 covariates (14 with an effect), of which 14 binary (7 with an effect), of which 2 fixed (1 with an effect), correlation of covariates ranging from -0.25 to 0.77 with median 0.09, absolute standardized coefficients (for the covariates with an effect) ranging from 0.02 to 7.81 with median 0.28
- ▶ 95% confidence intervals (CI): 1 - probability of non-coverage (i.e. probability of no error), in % (should be 95)
- ▶ level: probability of significance at 5% level for variables without effect (i.e. probability of type I error), in % (should be 5)
- ▶ power: probability of significance at 5% level for variables with an effect (i.e. probability of no type II error), in %
- ▶ time: using 1 core of a standard laptop (1.80GHz)

	CI			level			power			time
	min	median	max	min	median	max	min	median	max	median
full (glm+confint)	71.2	92.1	92.8	6.5	7.0	7.4	6.9	59.0	100	0.7 sec
naive (step+confint)	75.9	88.0	93.0	8.8	10.6	18.0	8.6	73.7	100	2.0 sec
selcorr	79.4	92.6	94.3	6.8	7.4	9.5	8.3	66.6	100	7.7 sec
selcorr(boot.repl=10)	79.7	95.2	96.0	4.0	4.4	5.8	5.1	52.2	100	1.7 min

simulation results

- ▶ logistic model: 7085 simulations of dependent variable, 500 data points, 28 covariates (14 with an effect), of which 14 binary (7 with an effect), of which 2 fixed (1 with an effect), correlation of covariates ranging from -0.25 to 0.77 with median 0.09, absolute standardized coefficients (for the covariates with an effect) ranging from 0.02 to 7.81 with median 0.28
- ▶ 95% confidence intervals (CI): 1 - probability of non-coverage (i.e. probability of no error), in % (should be 95)
- ▶ level: probability of significance at 5% level for variables without effect (i.e. probability of type I error), in % (should be 5)
- ▶ power: probability of significance at 5% level for variables with an effect (i.e. probability of no type II error), in %
- ▶ time: using 1 core of a standard laptop (1.80GHz)

	CI			level			power			time
	min	median	max	min	median	max	min	median	max	median
full (glm+confint)	71.2	92.1	92.8	6.5	7.0	7.4	6.9	59.0	100	0.7 sec
naive (step+confint)	75.9	88.0	93.0	8.8	10.6	18.0	8.6	73.7	100	2.0 sec
selcorr	79.4	92.6	94.3	6.8	7.4	9.5	8.3	66.6	100	7.7 sec
selcorr(boot.repl=10)	79.7	95.2	96.0	4.0	4.4	5.8	5.1	52.2	100	1.7 min
selcorr(boot.repl=100)	79.8	95.3	96.2	4.2	4.7	5.5	5.1	56.6	100	16.9 min

simulation results

- ▶ logistic model: 7085 simulations of dependent variable, 500 data points, 28 covariates (14 with an effect), of which 14 binary (7 with an effect), of which 2 fixed (1 with an effect), correlation of covariates ranging from -0.25 to 0.77 with median 0.09, absolute standardized coefficients (for the covariates with an effect) ranging from 0.02 to 7.81 with median 0.28
- ▶ 95% confidence intervals (CI): 1 - probability of non-coverage (i.e. probability of no error), in % (should be 95)
- ▶ level: probability of significance at 5% level for variables without effect (i.e. probability of type I error), in % (should be 5)
- ▶ power: probability of significance at 5% level for variables with an effect (i.e. probability of no type II error), in %
- ▶ time: using 1 core of a standard laptop (1.80GHz)

	CI			level			power			time
	min	median	max	min	median	max	min	median	max	median
full (glm+confint)	71.2	92.1	92.8	6.5	7.0	7.4	6.9	59.0	100	0.7 sec
naive (step+confint)	75.9	88.0	93.0	8.8	10.6	18.0	8.6	73.7	100	2.0 sec
selcorr	79.4	92.6	94.3	6.8	7.4	9.5	8.3	66.6	100	7.7 sec
selcorr(boot.repl=10)	79.7	95.2	96.0	4.0	4.4	5.8	5.1	52.2	100	1.7 min
selcorr(boot.repl=100)	79.8	95.3	96.2	4.2	4.7	5.5	5.1	56.6	100	16.9 min
selcorr(fixed.vars=all)	71.2	92.1	92.8	7.2	7.7	8.0	8.0	60.2	100	1.3 sec

simulation results

- ▶ logistic model: 7085 simulations of dependent variable, 500 data points, 28 covariates (14 with an effect), of which 14 binary (7 with an effect), of which 2 fixed (1 with an effect), correlation of covariates ranging from -0.25 to 0.77 with median 0.09, absolute standardized coefficients (for the covariates with an effect) ranging from 0.02 to 7.81 with median 0.28
- ▶ 95% confidence intervals (CI): 1 - probability of non-coverage (i.e. probability of no error), in % (should be 95)
- ▶ level: probability of significance at 5% level for variables without effect (i.e. probability of type I error), in % (should be 5)
- ▶ power: probability of significance at 5% level for variables with an effect (i.e. probability of no type II error), in %
- ▶ time: using 1 core of a standard laptop (1.80GHz)

	CI			level			power			time
	min	median	max	min	median	max	min	median	max	median
full (glm+confint)	71.2	92.1	92.8	6.5	7.0	7.4	6.9	59.0	100	0.7 sec
naive (step+confint)	75.9	88.0	93.0	8.8	10.6	18.0	8.6	73.7	100	2.0 sec
selcorr	79.4	92.6	94.3	6.8	7.4	9.5	8.3	66.6	100	7.7 sec
selcorr(boot.repl=10)	79.7	95.2	96.0	4.0	4.4	5.8	5.1	52.2	100	1.7 min
selcorr(boot.repl=100)	79.8	95.3	96.2	4.2	4.7	5.5	5.1	56.6	100	16.9 min
selcorr(fixed.vars=all)	71.2	92.1	92.8	7.2	7.7	8.0	8.0	60.2	100	1.3 sec
selcorr(fixed.vars=all,boot.repl=10)	71.7	95.5	96.2	3.8	4.2	4.6	4.4	44.7	100	13.8 sec

simulation results

- ▶ logistic model: 7085 simulations of dependent variable, 500 data points, 28 covariates (14 with an effect), of which 14 binary (7 with an effect), of which 2 fixed (1 with an effect), correlation of covariates ranging from -0.25 to 0.77 with median 0.09, absolute standardized coefficients (for the covariates with an effect) ranging from 0.02 to 7.81 with median 0.28
- ▶ 95% confidence intervals (CI): 1 - probability of non-coverage (i.e. probability of no error), in % (should be 95)
- ▶ level: probability of significance at 5% level for variables without effect (i.e. probability of type I error), in % (should be 5)
- ▶ power: probability of significance at 5% level for variables with an effect (i.e. probability of no type II error), in %
- ▶ time: using 1 core of a standard laptop (1.80GHz)

	CI			level			power			time
	min	median	max	min	median	max	min	median	max	median
full (glm+confint)	71.2	92.1	92.8	6.5	7.0	7.4	6.9	59.0	100	0.7 sec
naive (step+confint)	75.9	88.0	93.0	8.8	10.6	18.0	8.6	73.7	100	2.0 sec
selcorr	79.4	92.6	94.3	6.8	7.4	9.5	8.3	66.6	100	7.7 sec
selcorr(boot.repl=10)	79.7	95.2	96.0	4.0	4.4	5.8	5.1	52.2	100	1.7 min
selcorr(boot.repl=100)	79.8	95.3	96.2	4.2	4.7	5.5	5.1	56.6	100	16.9 min
selcorr(fixed.vars=all)	71.2	92.1	92.8	7.2	7.7	8.0	8.0	60.2	100	1.3 sec
selcorr(fixed.vars=all,boot.repl=10)	71.7	95.5	96.2	3.8	4.2	4.6	4.4	44.7	100	13.8 sec
selcorr(fixed.vars=all,boot.repl=100)	71.9	95.6	96.2	3.8	4.2	4.6	4.4	49.2	100	2.1 min