

Effizientes Bootstrapping mit SAS

Wien, 04.5.2009

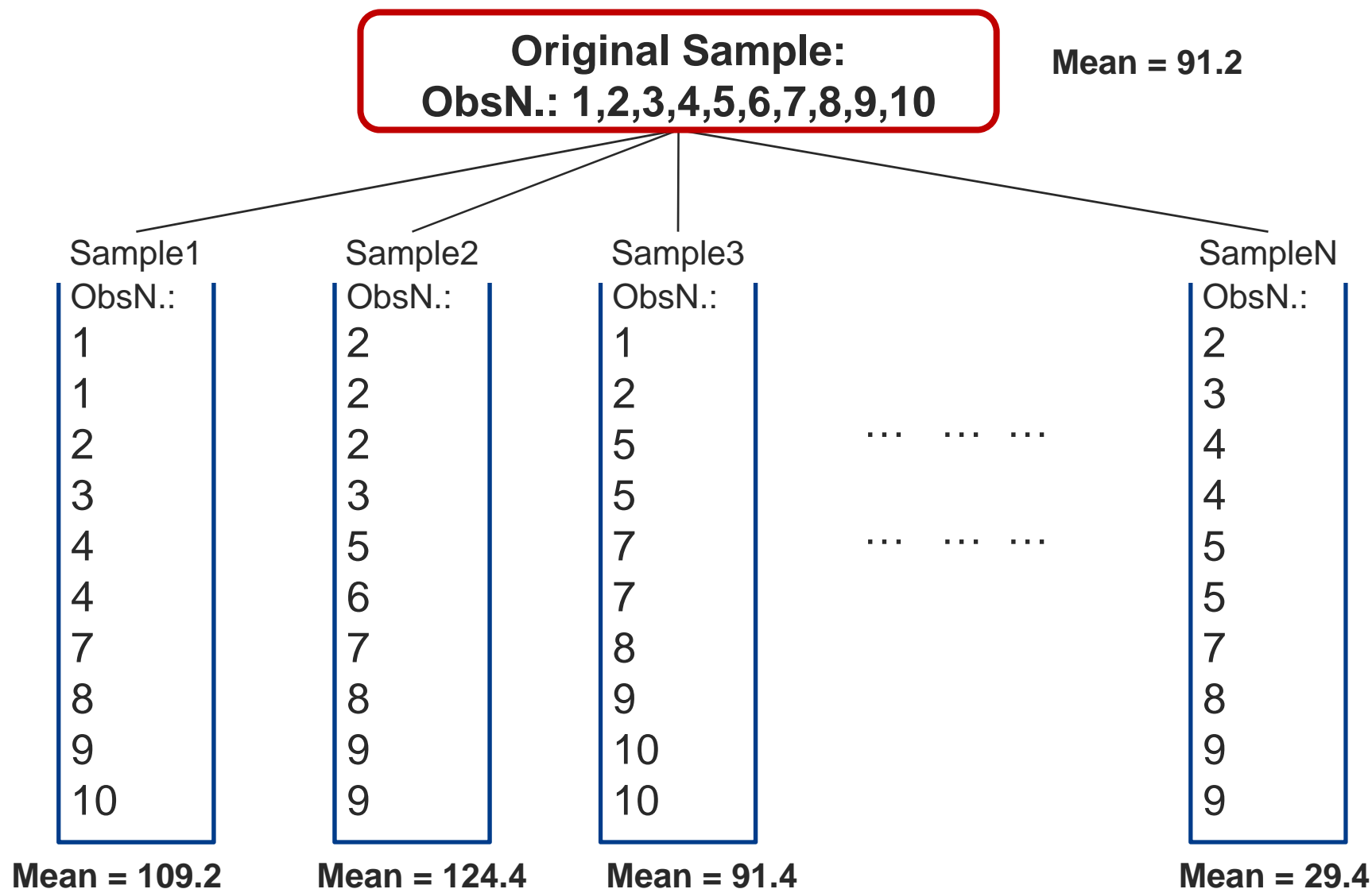
Dr. Mihai Paunescu

**THE
POWER
TO KNOW.**

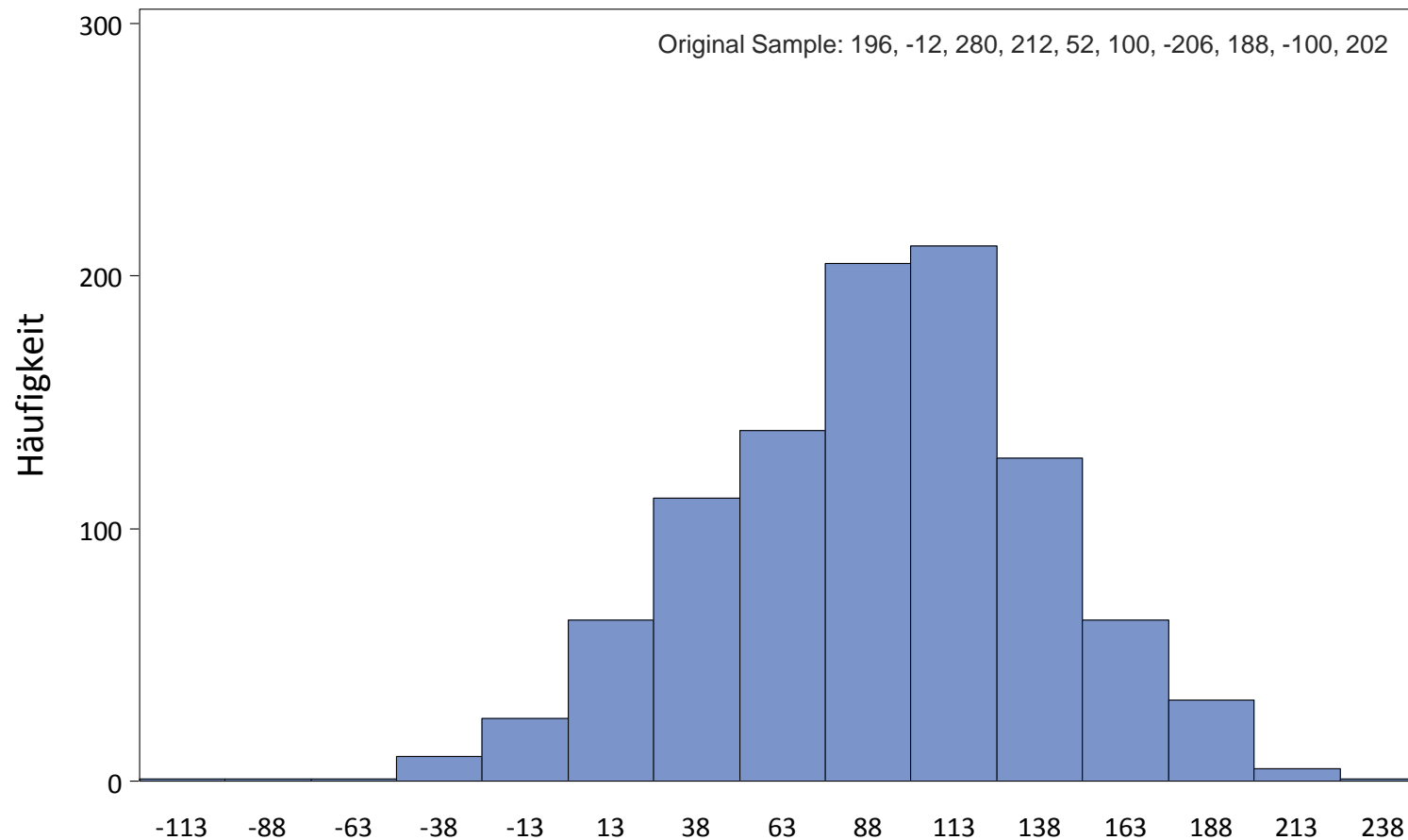
Agenda

- Was ist Bootstrapping?
- Effizientes Programmieren von Bootstrapping?
- Anwendungsbeispiel für Ordered Logit

Was ist Bootstrapping?



Mittelwert und Konfidenzintervall mit Bootstrapping



Mean = 90.1

95% Konfidenzintervall [-10.1, 183.5]

Wann braucht man Bootstrapping?

- Geringe Zahl von Beobachtungen
- Verteilungsannahmen nicht möglich

Agenda

- Was ist Bootstrapping?
- Effizientes Programmieren von Bootstrapping?
- Anwendungsbeispiel für Ordered Logit

```
%macro bootie ( input=, reps= );
%do i = 1 %to &REPS ;
  data gen;
    do i=1 to nobs;
      rec = ceil(nobs * ranuni(0));
      set &INPUT nobs=nobs point=rec;
      output;
    end;
  stop;
run;
proc univariate data=gen;
  var x;
  output out=outx mean=mean;
run;
%if &I = 1 %then %do;
  data outall;
  set outx;
  run;
%end;
%else %do;
  proc append base=outall data=outx; run;
%end;
%end; /* i=1 to &REPS loop */
proc univariate data=outall;
  var mean;
  output out=final pctlpts=2.5, 97.5 pctlpre=ci;
run;
%mend;
```

Ineffiziente Variante

- Generieren von Samples
- Errechnen des Mittelwertes für jedes Sample
- Zusammenfügen der Mittelwertdaten
- Ausgabe des Mittelwertes und CI

Effizientes Bootstrapping

```
proc surveyselect data=data_org out=outboot
  seed=3
  method=urs
  samprate=1
  outhits
  rep=1000;
run;
proc univariate data=outboot;
  var x;
  by Replicate;
  output out=outall mean=mean;
run;
proc univariate data=outall;
  var mean;
  histogram;
  output out=final
  pctlpts=2.5, 97.5 pctlpre=ci;
run;
```

- Generierung aller Samples
- Berechnung des Mittelwertes für jedes Sample mit BY Anweisung
- Ausgabe des Mittelwertes und CI

Agenda

- Was ist Bootstrapping?
- Effizientes Programmieren von Bootstrapping
- **Anwendungsbeispiel für Ordered Logit**

Bootstrapping Anwendung für Ordered Logit

```
proc surveyselect data=data_org out=data_boot
```

```
seed=3
```

```
method = urs
```

```
samprate = 1
```

```
outhits
```

```
rep = 100;
```

```
quit;
```

```
proc logistic data= data_boot
```

```
noprint outest=parm_est_boot;
```

```
by replicate;
```

```
freq numberhits;
```

```
model output = var1 var2 /
```

```
link = logit;
```

```
run;quit;
```

```
proc means data=parm_est_boot noprint nway;
```

```
var Intercept: var1 var2;
```

```
id type_ _name_;
```

```
output out=parm_boot (drop=_freq_) mean=;
```

```
run;
```

■ Generierung aller Samples

■ Berechnung der Schätzer für jedes Sample

■ Ausgabe Schätzer mit Bootstrapping

Stepwise Selection und Bootstrapping kombinieren

1. Bootstrapping mit Logit mit Stepwise Selection durchführen.
2. Alle nicht selektieren Parameter auf Null setzen.
3. T-Test, ob Mittelwert jedes Parameters von Null verschieden ist.
4. Bootstrapping mit allen Variablen durchführen, deren Parameter signifikant waren.

Details zum Nachlesen

- Don't be loopy: Re-Sampling and Simulation the SAS Way (Cassell 2007)
- A Practical Introduction to the Bootstrap Using the SAS System (Barker, 2005)
- Efron and Tibshirani (1993): An Introduction to the Bootstrap, Chapman & Hall/CRC (ISBN: 0-412-04231-2).

Agenda

- Integration von SAS in eine bestehende DWH-und BI-Landschaft
Andreas Krystovsky – Telekom Austria AG
- Neuerungen in SAS®9.2 zu Reporting und Datenintegration
Kashif Din – SAS Austria
- Neuerungen in SAS®9.2 zu Analytik und Data Mining
Gerhard Svolba – SAS Austria
- SAS Tipps&Tricks:
Effizientes Bootstrapping mit SAS
Mihai Paunescu – SAS Austria
- SAS News Corner

SAS-Kurse im Juli

- SAS 9.2 Data Integration: Neuerungen und Erweiterungen
- SAS® Data Integration Studio 2: komplexe ETL-Prozesse
- SAS® Data Integration Studio 3: Administration und Performance-optimierung
- Kompaktkurs: Platform Administration 9.2
- Erstellung analytischer Data Marts
- Warehouse-Architektur und Data Integration Design

SAS Events 2009



25. Juni, SAS Office Wien

Frühstücksseminar:

„Schluss mit Garbage-In-Garbage-Out.
Fünf Schritte für bessere Datenqualität.“



1.-2. Juli, Kopenhagen (Dänemark)

Erste Analytics Konferenz Europa's:

Vorträge zu Data Mining, Forecasting, Operational
Research und Statistical Analysis



1. Oktober,

Orangerie & Tagungszentrum Schönbrunn

„Innovate. Optimise. Transform“

