universität innsbruck

Stochastische Regressionsanalyse: Intro

Grundlagen der Ökonometrie

herbert.stocker@uibk.ac.at

www.hsto.info/econometrics

Stochastische Regressionsanalyse: Intro

Was bisher geschah:

- Wir uns bisher für existierende Daten interessiert,
  v.a. Kennzahlen berechnet, z.B. *bedingte Mittelwerte* (OLS).

Was kommt jetzt?

- Wir interessieren uns nicht länger dafür, was passiert ist (Beobachtungen), sondern *was passieren könnte!*
- Eine einfache Möglichkeit sich dies vorzustellen:
  - wir beobachten nur eine Stichprobe aus einer unbekannten Grundgesamtheit,
  - aber wir interessieren uns für die Grundgesamtheit!
- Wiederum versuchen wir, die unbekannte Grundgesamtheit durch Kennzahlen zu beschreiben, z.B. durch *bedingte Erwartungswerte*.
- Wir benötigen zusätzliche Kennzahlen um die Unsicherheit zu charakterisieren, z.B. *Standardfehler*.

Stochastische Regressionsanalyse: Intro

Vermuteter Zusammenhang in unbekannter Grundgesamtheit:

‘Population Regression Function’ PRF

$$y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$$

Autobeispiel (Grundgesamtheit, PRF):

$$\text{PRF:} \qquad \text{Preis}_i = 23\,081 - 2\,630\,\text{Alter}_i + \varepsilon_i$$

PRF ist unbeobachtbar! (Forscher kann nur eine *Stichprobe* beobachten!)

Stochastische Regressionsanalyse: Intro

Vermuteter Zusammenhang in unbekannter Grundgesamtheit:

'Population Regression Function' PRF

$$y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$$

Autobeispiel (Grundgesamtheit, PRF):

$$\text{PRF:} \qquad \text{Preis}_i = 23\,081 - 2\,630\,\text{Alter}_i + \varepsilon_i$$

PRF ist unbeobachtbar! (Forscher kann nur eine *Stichprobe* beobachten!)

Stochastische Regressionsanalyse: Intro

Vermuteter Zusammenhang in unbekannter Grundgesamtheit:

‘Population Regression Function’ PRF

$$y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$$

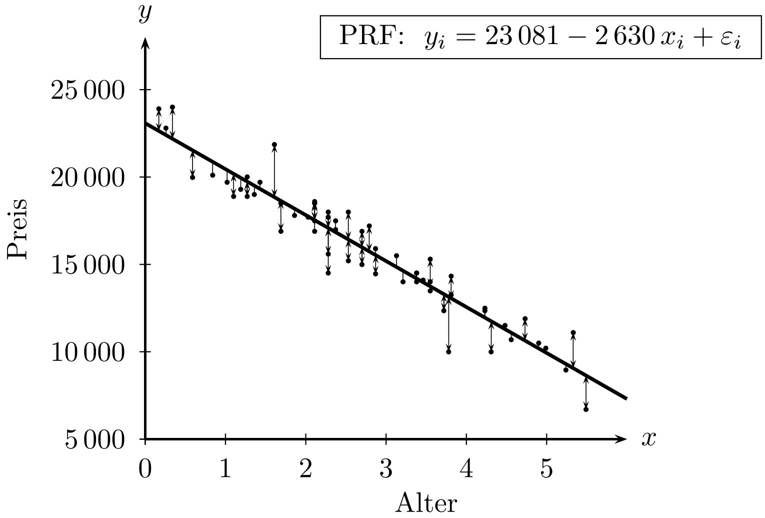
Autobeispiel (Grundgesamtheit, PRF):

$$\text{PRF:} \qquad \text{Preis}_i = 23\,081 - 2\,630\,\text{Alter}_i + \varepsilon_i$$

PRF ist unbeobachtbar! (Forscher kann nur eine *Stichprobe* beobachten!)

Stochastische Regressionsanalyse: Intro

Wenn sie beobachtbar wäre ...



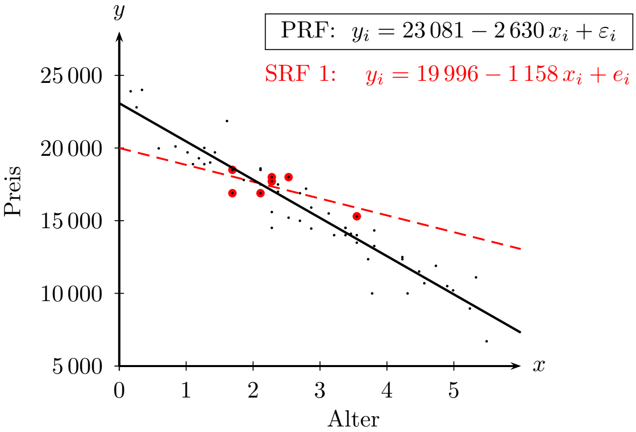
Stochastische Regressionsanalyse: Intro

Ein Forscher kann nur eine Stichprobe mit $n = 7$ beobachten:

'Sample Regression Function' **SRF**

Stichprobe zu SRF 1

Obs.	Preis	Alter
3	18000	2.28
7	18000	2.53
14	17700	2.28
21	16900	1.69
31	15300	3.55
56	16900	2.11
59	18500	1.69



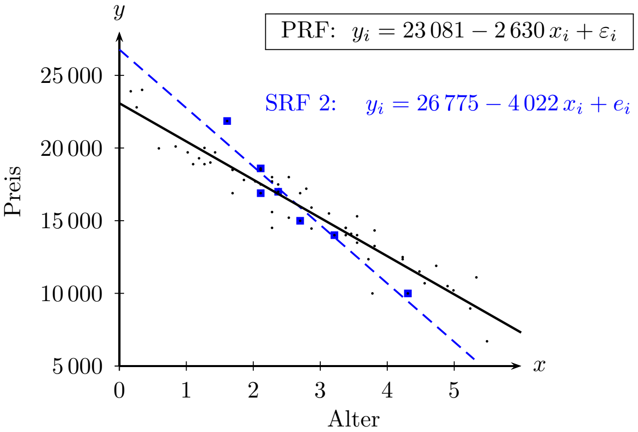
Stochastische Regressionsanalyse: Intro

Ein zweite Forscherin beobachtet andere Stichprobe ($n = 7$):

Stichprobe zu SRF 2

Obs.	Preis	Alter
1	16990	2.37
16	15000	2.70
25	10000	4.31
29	21850	1.61
35	18600	2.11
56	16900	2.11
60	14000	3.21

'Sample Regression Function 2' **SRF 2**



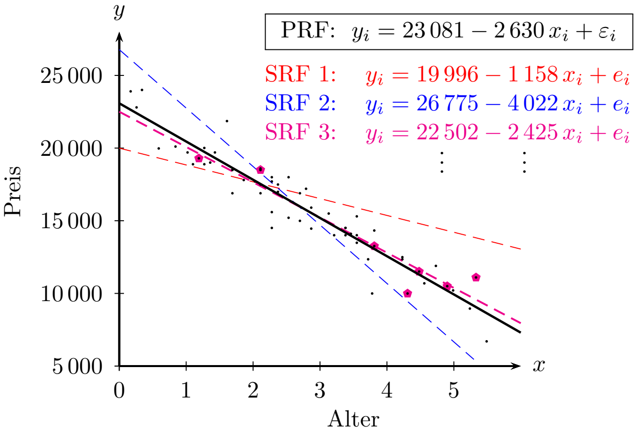
Stochastische Regressionsanalyse: Intro

Ein dritter Forscher zieht nochmals eine andere Stichprobe ($n = 7$):

Stichprobe zu SRF 2

Obs.	Preis	Alter
11	11100	5.33
25	10000	4.31
30	13250	3.81
40	18500	2.11
47	11500	4.48
50	10500	4.90
51	19300	1.19

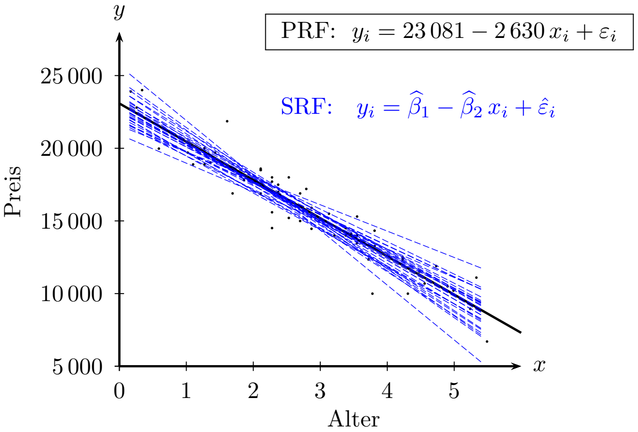
'Sample Regression Function 3' **SRF 3**



Stochastische Regressionsanalyse: Intro

Monte Carlo Simulation: Computer zieht z.B. 1000 Stichproben und berechnet 1000 SRF's:

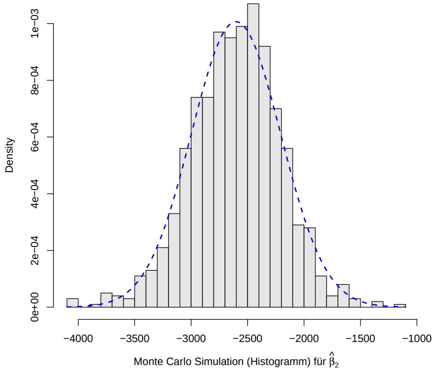
Stichprobe	b_1	b_2
1	22476	-2513
2	23525	-2658
3	22502	-2425
⋮	⋮	⋮
33	19996	-1158
⋮	⋮	⋮
369	26775	-4022
⋮	⋮	⋮
999	23327	-2714
1000	23598	-2875
Mittelwert:	22976	-2603



Stochastische Regressionsanalyse: Intro

Monte Carlo Simulation: Histogramm für Steigungskoeffizienten b_2 der SRF's:

Stichprobe	b_1	b_2
1	22476	-2513
2	23525	-2658
3	22502	-2425
⋮	⋮	⋮
33	19996	-1158
⋮	⋮	⋮
369	26775	-4022
⋮	⋮	⋮
999	23327	-2714
1000	23598	-2875
Mittelwert:	22976	-2603

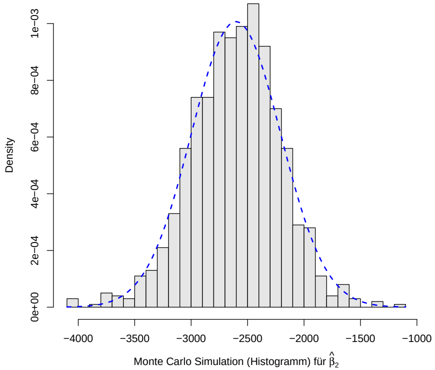


$\Rightarrow$ **1) Gesetz der Großen Zahl** **2) Zentraler Grenzwertsatz**

Stochastische Regressionsanalyse: Intro

Monte Carlo Simulation: Histogramm für Steigungskoeffizienten b_2 der SRF's:

Stichprobe	b_1	b_2
1	22476	-2513
2	23525	-2658
3	22502	-2425
$\vdots$	$\vdots$	$\vdots$
33	19996	-1158
$\vdots$	$\vdots$	$\vdots$
369	26775	-4022
$\vdots$	$\vdots$	$\vdots$
999	23327	-2714
1000	23598	-2875
Mittelwert:	22976	-2603

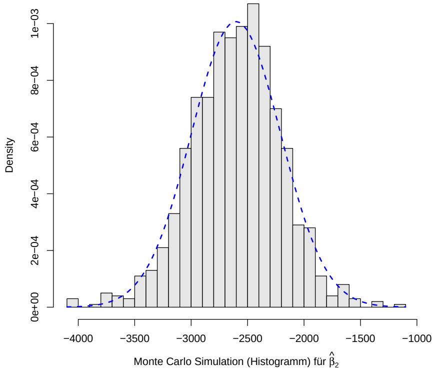


$\Rightarrow$ **1) Gesetz der Großen Zahl** **2) Zentraler Grenzwertsatz**

Stochastische Regressionsanalyse: Intro

Monte Carlo Simulation: Histogramm für Steigungskoeffizienten b_2 der SRF's:

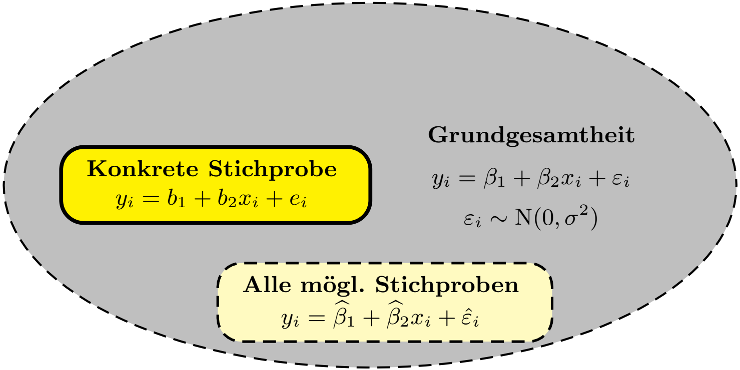
Stichprobe	b_1	b_2
1	22476	-2513
2	23525	-2658
3	22502	-2425
$\vdots$	$\vdots$	$\vdots$
33	19996	-1158
$\vdots$	$\vdots$	$\vdots$
369	26775	-4022
$\vdots$	$\vdots$	$\vdots$
999	23327	-2714
1000	23598	-2875
Mittelwert:	22976	-2603



$\Rightarrow$ **1) Gesetz der Großen Zahl** **2) Zentraler Grenzwertsatz**

Stochastische Regressionsanalyse: Intro

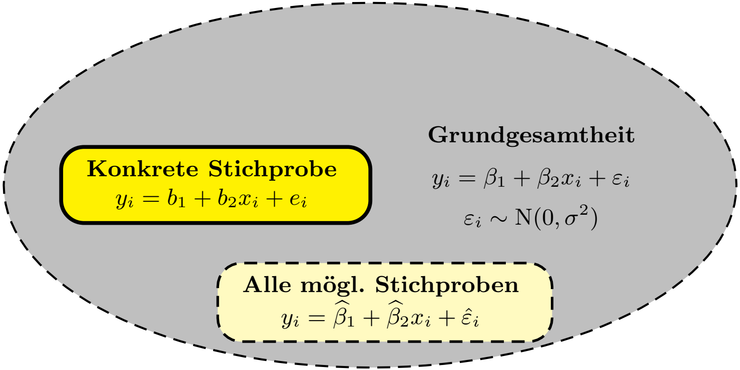
Notation:



β . . . Parameter; $\widehat{\beta}$ . . . Zufallsvariable; b . . . Realisation

Stochastische Regressionsanalyse: Intro

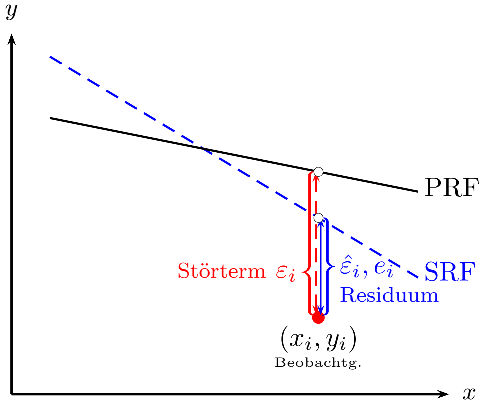
Notation:



β ... Parameter; $\widehat{\beta}$ ... Zufallsvariable; b ... Realisation

Notation

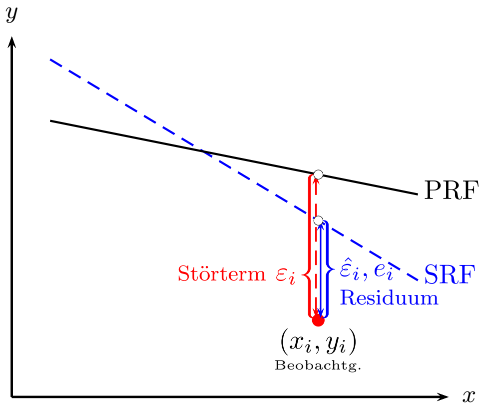
Störterme und Residuen:



Achtung: 'Residuum' wird für Zufallsvariable und Realisation verwendet!

Notation

Störterme und Residuen:



Achtung: ‘Residuum’ wird für Zufallsvariable und Realisation verwendet!

Notation

- **DGP (Grundgesamtheit):**
  - $\beta_1, \ldots, \beta_k$: unbeobachtbare Parameter der PRF, deterministisch!
  - ε_i: unbeobachtbare Störterme (*errors*), stochastisch!

- **Stichprobe:**
  - *Zufallsvariablen:*
    - $\widehat{\beta}_1, \ldots, \widehat{\beta}_k$: Zufallsvariablen (stochastisch!), Ergebnisse für jede mögliche Stichprobe
    - $\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_n$: Residuen, Ergebnisse für jede mögliche Stichprobe
  - *Realisationen:*
    - $b_1, \ldots, b_k$: d.h. Ergebnis für die *eine* beobachtete Stichprobe, deterministisch!
    - $e_i, \ldots, e_n)$: Realisationen der Residuen, deterministisch!

Achtung: In der Literatur wird für die Zufallsvariablen $\widehat{\beta}_h$ und für die deterministischen Realisationen b_h häufig das selbe Symbol $\widehat{\beta}_h$ verwendet!

Stochastische Regressionsanalyse: Intro

OLS: die Mathematik funktioniert gleich

$$\min_{\widehat{\beta}_1,\widehat{\beta}_2} \sum_{i=1}^{n} \hat{\varepsilon}_i^2 = \min_{\widehat{\beta}_1,\widehat{\beta}_2} \sum_{i=1}^{n} (y_i - \widehat{\beta}_1 - \widehat{\beta}_2 x_i)^2$$

FOC:

$$\frac{\partial \sum_i \hat{\varepsilon}_i^2}{\partial \widehat{\beta}_1} = 2\sum_i \underbrace{\left(y_i - \widehat{\beta}_1 - \widehat{\beta}_2 x_i\right)}_{\hat{\varepsilon}_i}(-1) = 0 \quad \Rightarrow \quad \sum_i \hat{\varepsilon}_i = 0$$

$$\frac{\partial \sum_i \hat{\varepsilon}_i^2}{\partial \widehat{\beta}_2} = 2\sum_i \underbrace{\left(y_i - \widehat{\beta}_1 - \widehat{\beta}_2 x_i\right)}_{\hat{\varepsilon}_i}(-x_i) = 0 \quad \Rightarrow \quad \sum_i x_i \hat{\varepsilon}_i = 0$$

$$\widehat{\beta}_2 = \frac{\widehat{\mathrm{cov}}(x,y)}{\widehat{\mathrm{var}}(x)}, \qquad \widehat{\beta}_1 = \bar{y} - \widehat{\beta}_2 \bar{x}$$

Stochastische Regressionsanalyse: Intro

OLS: die Mathematik funktioniert gleich

$$\min_{\widehat{\beta}_1,\widehat{\beta}_2} \sum_{i=1}^{n} \hat{\varepsilon}_i^2 = \min_{\widehat{\beta}_1,\widehat{\beta}_2} \sum_{i=1}^{n} (y_i - \widehat{\beta}_1 - \widehat{\beta}_2 x_i)^2$$

FOC:

$$\frac{\partial \sum_i \hat{\varepsilon}_i^2}{\partial \widehat{\beta}_1} = 2 \sum_i \underbrace{\left(y_i - \widehat{\beta}_1 - \widehat{\beta}_2 x_i\right)}_{\hat{\varepsilon}_i}(-1) = 0 \quad \Rightarrow \quad \sum_i \hat{\varepsilon}_i = 0$$

$$\frac{\partial \sum_i \hat{\varepsilon}_i^2}{\partial \widehat{\beta}_2} = 2 \sum_i \underbrace{\left(y_i - \widehat{\beta}_1 - \widehat{\beta}_2 x_i\right)}_{\hat{\varepsilon}_i}(-x_i) = 0 \quad \Rightarrow \quad \sum_i x_i \hat{\varepsilon}_i = 0$$

$$\widehat{\beta}_2 = \frac{\widehat{\text{cov}}(x,y)}{\widehat{\text{var}}(x)}, \qquad \widehat{\beta}_1 = \bar{y} - \widehat{\beta}_2 \bar{x}$$

Stochastische Regressionsanalyse: Intro

OLS: die Mathematik funktioniert gleich

$$\min_{\widehat{\beta}_1,\widehat{\beta}_2} \sum_{i=1}^{n} \hat{\varepsilon}_i^2 = \min_{\widehat{\beta}_1,\widehat{\beta}_2} \sum_{i=1}^{n} (y_i - \widehat{\beta}_1 - \widehat{\beta}_2 x_i)^2$$

FOC:

$$\frac{\partial \sum_i \hat{\varepsilon}_i^2}{\partial \widehat{\beta}_1} = 2 \sum_i \underbrace{\left(y_i - \widehat{\beta}_1 - \widehat{\beta}_2 x_i\right)}_{\hat{\varepsilon}_i} (-1) = 0 \quad \Rightarrow \quad \sum_i \hat{\varepsilon}_i = 0$$

$$\frac{\partial \sum_i \hat{\varepsilon}_i^2}{\partial \widehat{\beta}_2} = 2 \sum_i \underbrace{\left(y_i - \widehat{\beta}_1 - \widehat{\beta}_2 x_i\right)}_{\hat{\varepsilon}_i} (-x_i) = 0 \quad \Rightarrow \quad \sum_i x_i \hat{\varepsilon}_i = 0$$

$$\widehat{\beta}_2 = \frac{\widehat{\text{cov}}(x,y)}{\widehat{\text{var}}(x)}, \qquad \widehat{\beta}_1 = \bar{y} - \widehat{\beta}_2 \bar{x}$$

Stochastische Regressionsanalyse: Intro

- Das einfache 'Grundgesamtheit – Stichproben' Modell ist nützlich für erste Einsichten und um die Notation zu verstehen . . .
  . . . aber es eignet sich weniger gut für tiefere Einsichten.

Stochastische Regressionsanalyse: Intro

- Das einfache ‘Grundgesamtheit – Stichproben’ Modell ist nützlich für erste Einsichten und um die Notation zu verstehen . . .
  . . . aber es eignet sich weniger gut für tiefere Einsichten.

- Modell von R.A. Fisher ist allgemeiner und zeigt die prinzipiellen Ideen.

Stochastische Regressionsanalyse: Intro

- Das einfache ‘Grundgesamtheit – Stichproben’ Modell ist nützlich für erste Einsichten und um die Notation zu verstehen . . .
  . . . aber es eignet sich weniger gut für tiefere Einsichten.

- Modell von R.A. Fisher ist allgemeiner und zeigt die prinzipiellen Ideen.
- Im Folgenden zeige ich die einfachst mögliche Anwendung, die Schätzung eines Anteils.

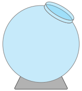
Stochastische Regressionsanalyse: Intro

- Das einfache ‘Grundgesamtheit – Stichproben’ Modell ist nützlich für erste Einsichten und um die Notation zu verstehen . . .
  . . . aber es eignet sich weniger gut für tiefere Einsichten.

- Modell von R.A. Fisher ist allgemeiner und zeigt die prinzipiellen Ideen.
- Im Folgenden zeige ich die einfachst mögliche Anwendung, die Schätzung eines Anteils.
- Dieses Modell werden wir später für das Regressionsmodell erweitern . . .

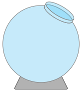
Schätzung eines Anteils nach R.A. Fisher

Was der Forscher
sieht (vermutet) ...

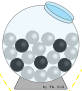


Schätzung eines Anteils nach R.A. Fisher

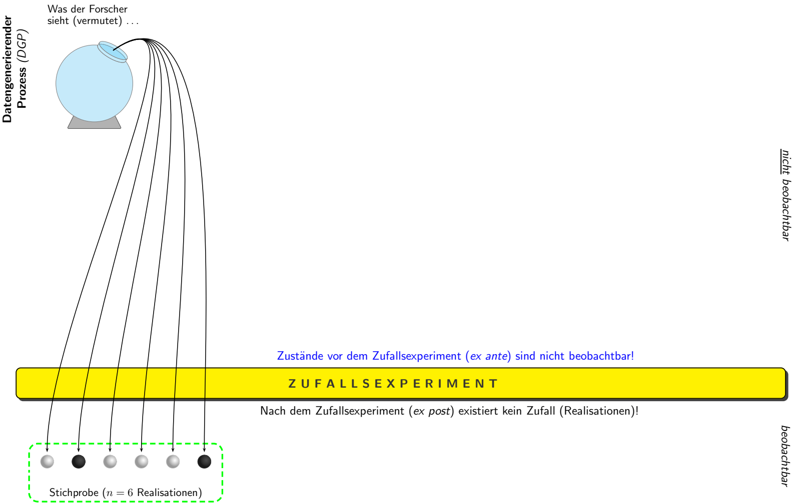
Was der Forscher
sieht (vermutet) ...



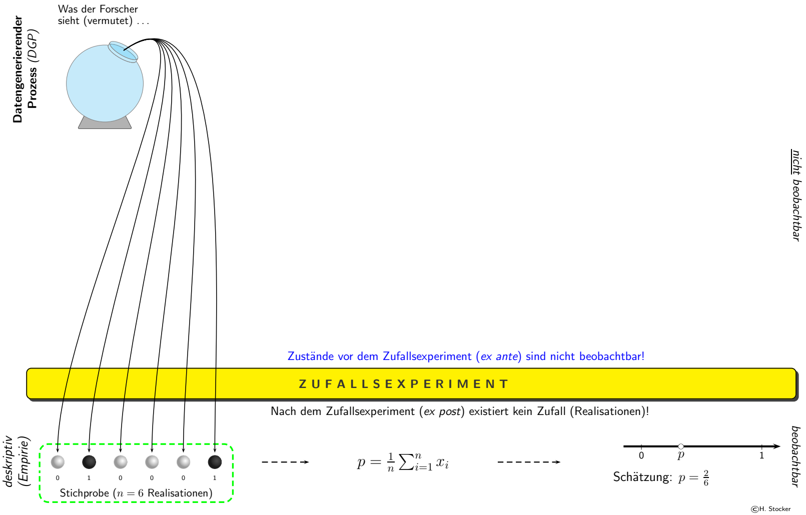
& was der liebe Gott weiß:
$\pi = 5/17$



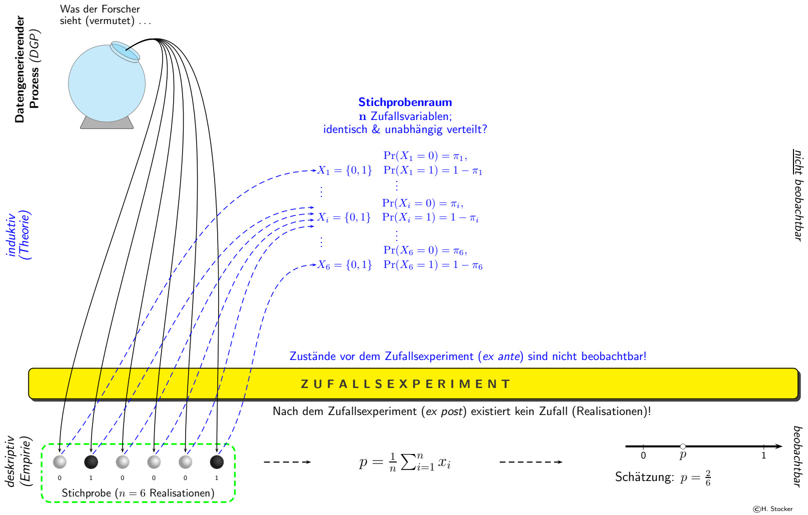
Schätzung eines Anteils nach R.A. Fisher



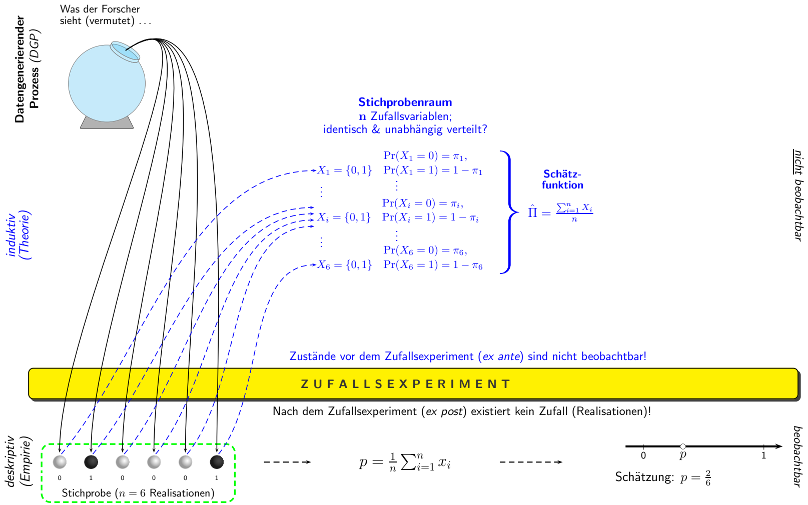
Schätzung eines Anteils nach R.A. Fisher



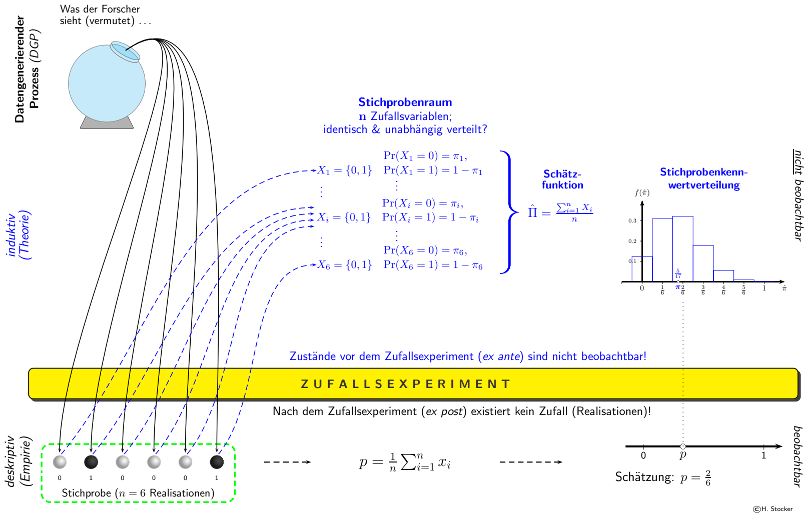
Schätzung eines Anteils nach R.A. Fisher



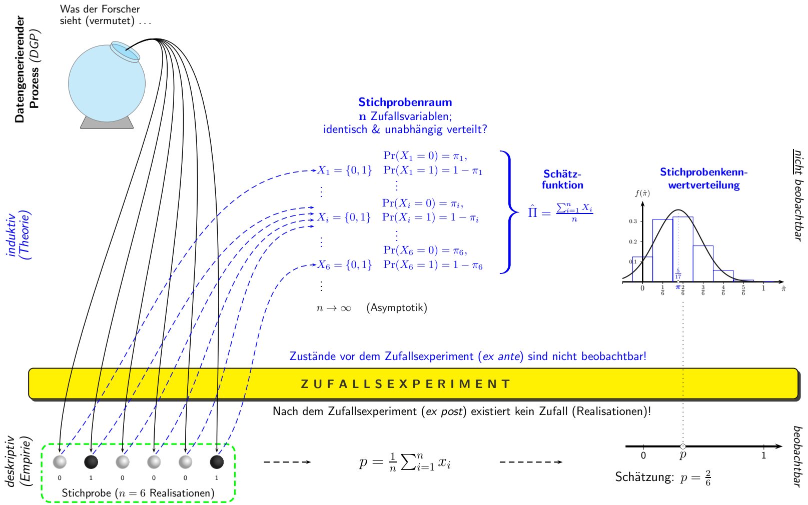
Schätzung eines Anteils nach R.A. Fisher



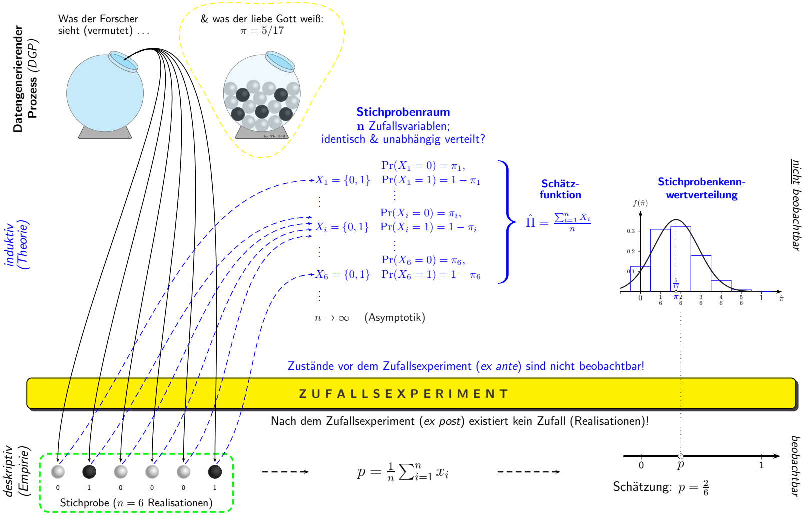
Schätzung eines Anteils nach R.A. Fisher



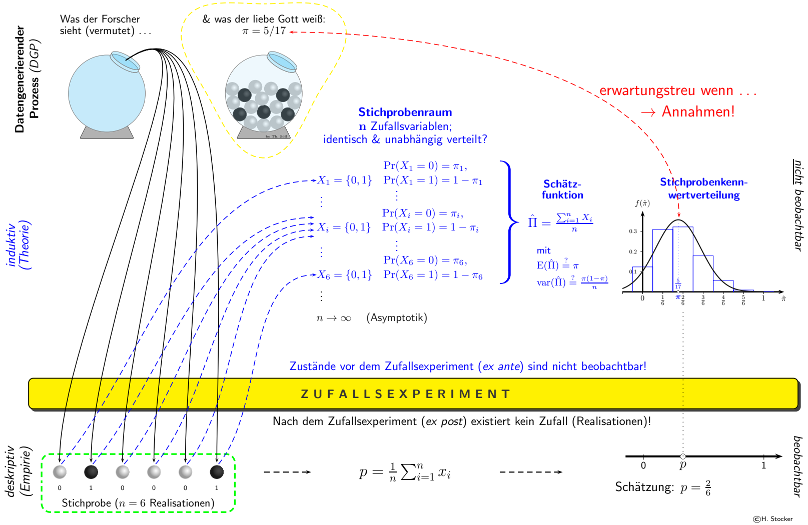
Schätzung eines Anteils nach R.A. Fisher



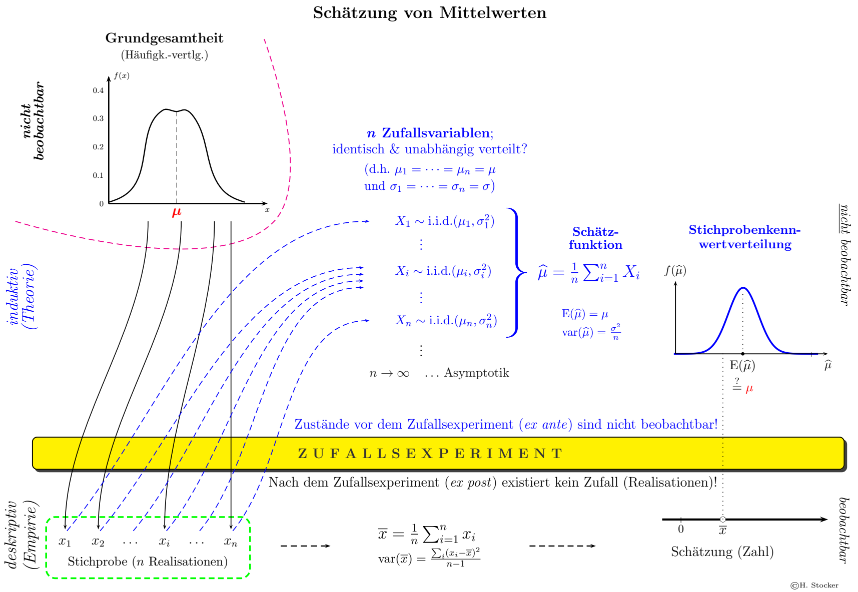
Schätzung eines Anteils nach R.A. Fisher



Schätzung eines Anteils nach R.A. Fisher



Das funktioniert auch für metrisch skalierte Variablen:



Stochastische Regressionsanalyse: Intro

Dieses Modell werden werden wir später für die lineare Regression erweitern ...

... aber vorher kommt noch die PRF und die bedingte Erwartungswertfunktion (*'conditional expectation function'*, **CEF**)

$$\textbf{PRF:} \qquad y_i = \underbrace{\beta_1 + \beta_2 x_i}_{\textbf{CEF}} + \varepsilon_i$$

Thanx ...