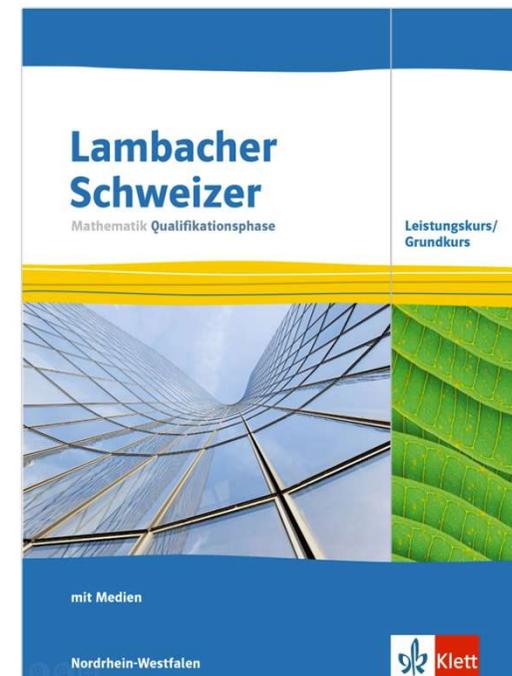


# Der neue Lambacher Schweizer Qualifikationsphase für Nordrhein-Westfalen

- Bericht aus der Autorenwerkstatt
- Prognose- und Konfidenzintervalle





D. Brandt

T. Jürgensen-Engl

T. Jörgens

R. Sonntag

I. Giersemehl

H. Spielmans

S. Aslanidis

J. Lohmann

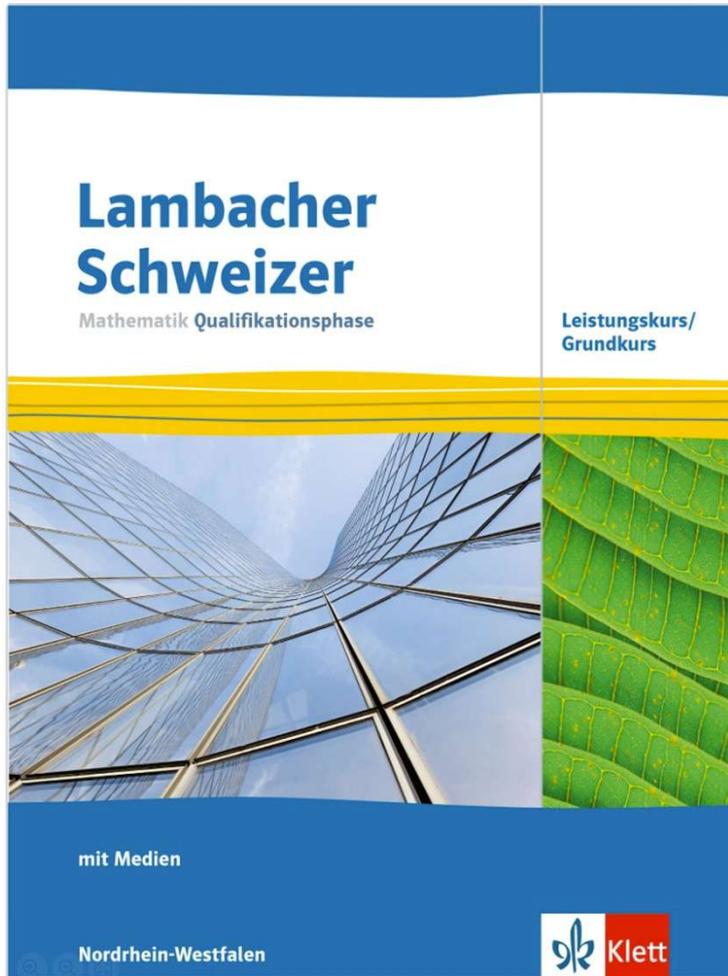
W. Riemer

W. Bucholzki

A. Vasicsek

A. Strunk

## Überblick über das Online-Seminar



### Teil I: Blick ins Buch

Warum (wie) das Buch mehr sein kann als eine differenzierende intelligente Aufgabensammlung

- 1) Struktureller Aufbau
- 2) Umgang mit MMS bzw. WTR
- 3) Prüfungsvorbereitung

### Teil II: Schätzen statt Testen

- 1) Stolpersteine bei Signifikanztests
- 2) Prognoseintervalle für relative Häufigkeiten  
 $1/\sqrt{n}$  Gesetz
- 3) Konfidenzintervalle

Sie dürfen „nebenbei“ in unseren Assets stöbern:  
[www.riemer-koeln.de/assets](http://www.riemer-koeln.de/assets)

# Willkommen in Teil II (Schätzen statt Testen)

## Spekulieren über Häufigkeiten

Werfen Sie **in Gedanken**  
**25-mal** eine L-Münze.  
 Notieren Sie 4 plausible  
 Trefferzahlen. Ebenso  
**100-mal** und für **400-mal**

	25	100	400
bernd 1	15	17	75
bernd 2	13	31	200
bernd 3	19	68	150
bernd 4	9	48	260

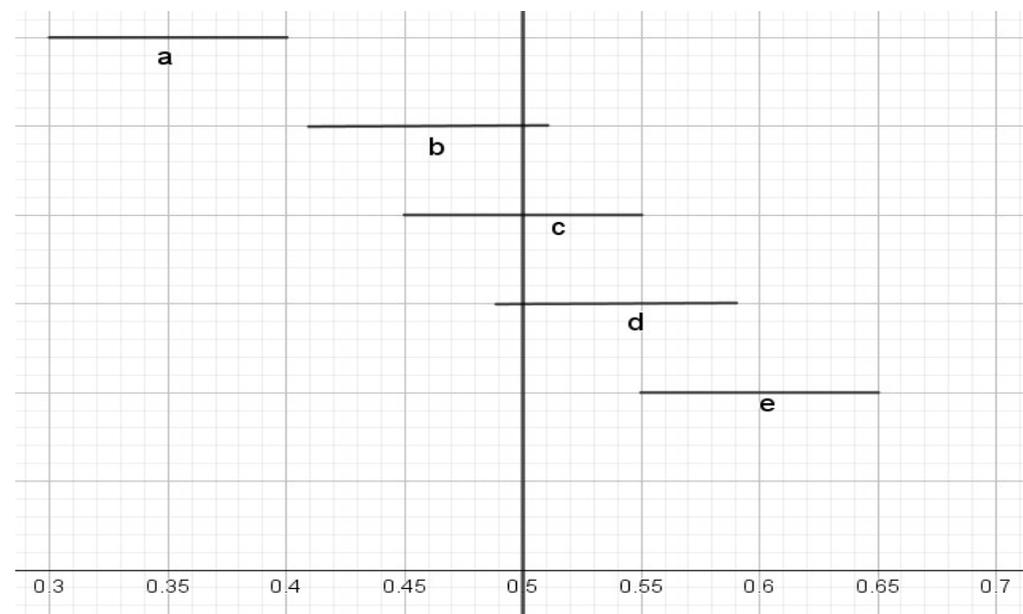
# LEGO – Münzen

## Spekulieren über Wahrscheinlichkeiten

Wo vermuten Sie „die“  
Wahrscheinlichkeit für  
Fußlage



- a) deutlich unter 50%
- b) eher unter 50%
- c) ziemlich genau 50%
- d) eher über 50%
- e) deutlich über 50%



# Schätzen statt Testen

## Wahrscheinlichkeit als Modell – Konfidenzintervalle

### Stolpersteine bei Signifikanztests

- Wahrscheinlichkeit als Modell

### Prognoseintervalle

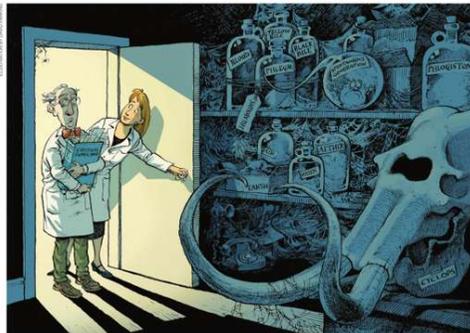
### Konfidenzintervalle

- Begriffsbildung digital
- klassisch-analog

### Resümee

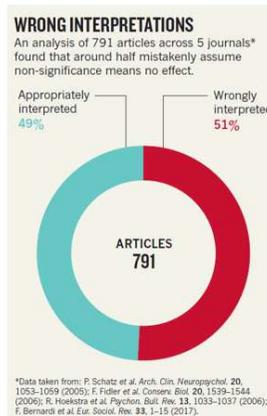
# Das steckt dahinter

MU 4/2020 Schätzen statt Testen



## Retire statistical significance

Valentin Amrhein, Sander Greenland, Blake McShane and more than 800 signatories call for an end to hyped claims and the dismissal of possibly crucial effects.



21 MARCH 2019 | VOL 567 | NATURE | 305

# Süddeutsche Zeitung

SZ.de Zeitung Magazin

20. März 2019, 19:06 Uhr Statistik

## Signifikanter Unfug

Die statistische Signifikanz, gemessen mit dem sogenannten **p-Wert**, hat in der Wissenschaft eine geradezu götzenhafte Bedeutung erlangt. **800 Forscher beklagen Fehler und fordern ein Umdenken.**

Bestell-Nr. 524214

**MU**  
DER  
MATHEMATIK-  
UNTERRICHT

Blabla blabla blabla blabla **p=0,4**  
blabla blabla blabla blabla bla-  
bla blabla blabla blabla blabla bla-  
bla blabla blabla **800** blabla blabla  
blabla blabla blabla blabla blabla  
blabla blabla blabla blabla blabla  
blabla blabla blabla **unter 40%**  
blabla blabla blabla blabla bla-  
bla blabla blabla blabla blabla bla-  
bla blabla **höchstens 4%** blabla  
blabla blabla blabla blabla blabla  
blabla blabla blabla blabla blabla

**FRIEDRICH**  
**TY**

Schickt die statistische Signifikanz in den Ruhestand!

Jahrgang 66 · Heft 4 · Juli 2020

## Bildungsserver

Der Hypothesentest ist ein rechnerisches Verfahren, das **eine Entscheidungshilfe** ob ein Würfel **ideal ist oder nicht**

[https://lehrerfortbildung-bw.de/u\\_matnatech/mathematik/gym/bp2004/fb2/modul4/4\\_unterricht/](https://lehrerfortbildung-bw.de/u_matnatech/mathematik/gym/bp2004/fb2/modul4/4_unterricht/)



## Max

Aber wir wissen doch, dass es keine idealen Würfel gibt - genauso wenig wie echte Punkte in der Geometrie.

# Stolperstein II

**Nina:**

Ich konnte  $H_0$  mit der Irrtums-wahrscheinlichkeit 5% verwerfen.

Aber was bedeutet das jetzt?

**Max:**

„Das bedeutet:  $H_1$  gilt mit 95%. Denn  $H_0$  und  $H_1$  sind Gegenereignisse.

Die Wahrscheinlichkeiten von Gegenereignissen ergänzen sich zu 1.“

# Stolperstein II

**Nina:**

Ich konnte  $H_0$  mit der Irrtums-wahrscheinlichkeit 5% verwerfen.

Aber was bedeutet das jetzt?

**Max:**

„Das bedeutet:  $H_1$  gilt mit 95%. Denn  $H_0$  und  $H_1$  sind Gegenereignisse.

Die Wahrscheinlichkeiten von Gegenereignissen ergänzen sich zu 1.“

Annehmen der Hypothese bedeutet nicht, dass die Hypothese richtig ist. Verwerfen der Hypothese bedeutet nicht, dass die Hypothese falsch ist. :

Nina:

Und wie ist das mit signifikant und relevant?  
Gibt's da einen Unterschied?

Max:

Das bedeutet inhaltlich das Gleiche.  
Nur signifikante Ergebnisse werden  
veröffentlicht.  
Was veröffentlicht wird ist relevant.

# Stolperstein IV

Ich konnte  $H_0: p=0,5$   
bei  $n= 25 / 100 / 400$  verwerfen.  
Was ist der Unterschied?

	$n = 25$	$n = 100$	$n = 400$
$\alpha = 5 \%$			
$\alpha = 1 \%$			

Ich habe 0,5m bestellt und kontrolliert.  
Ich bin mir sicher/sehr sicher: Das Brett taugt nicht



Bloß nicht weitersagen!

Man kann jede Hypothese auf jedem Signifikanzniveau verwerfen, wenn man den Versuchsumfang nur groß genug macht! Irrelevant kleine (völlig unbedeutende) Unterschiede werden dann statistisch signifikant.

- Deute Wahrscheinlichkeit als Modell ...
- als „unsichere Festlegung“, die dadurch entstehen, dass aus Erfahrung Erwartung wird.
- Sammle die weniger unsicheren Festlegungen im Konfidenzintervall!



# VIII Daten und Wahrscheinlichkeit

Wahrscheinlichkeiten sind unsichere Festlegungen. Sie entstehen in unserer Vorstellung, wenn aus Erfahrungen (Daten) Erwartungen (Modelle) werden.

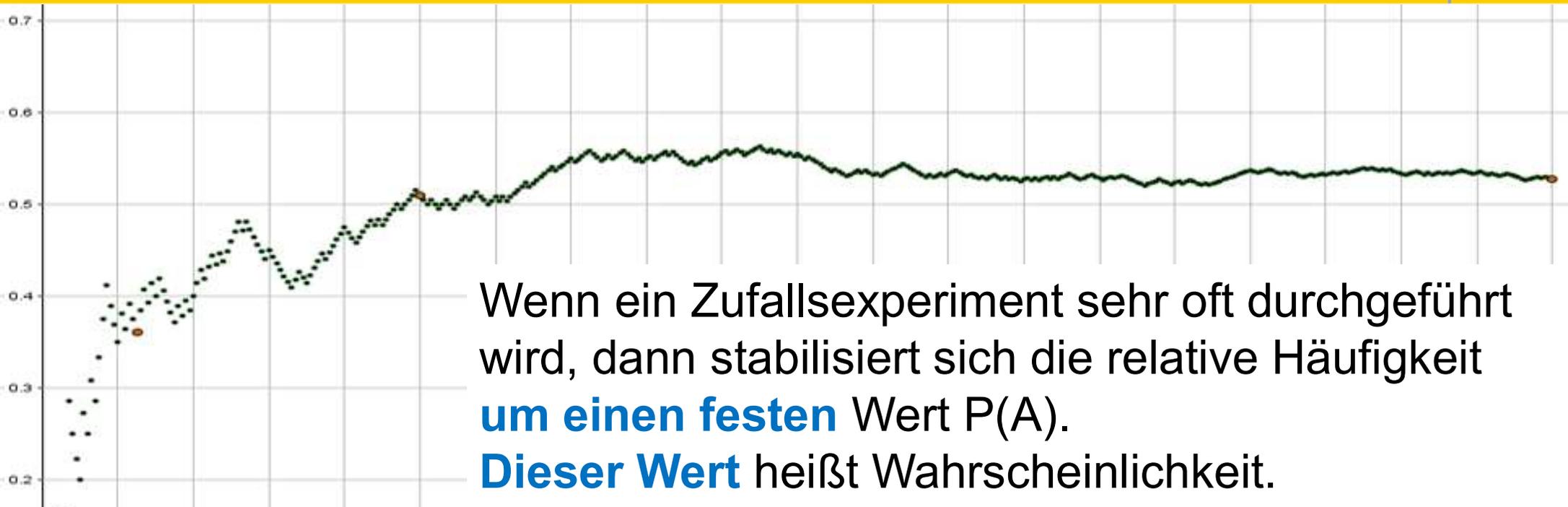
JEAN DUBUFFET (\* 1901 in Le Havre; † 1985 in Paris)  
Unsichere Festlegungen (1965) Kunstmuseum Basel



## Wahrscheinlichkeit als Modell – Konfidenzintervalle

- Stolpersteine bei Signifikanztests
- **Prognoseintervalle** (und das Konzept des Bezweifelns)
- **Konfidenzintervalle**
  - - Begriffsbildung digital
  - - klassisch-analog
- **Resümee**

# Die frequentistische Wahrscheinlichkeit



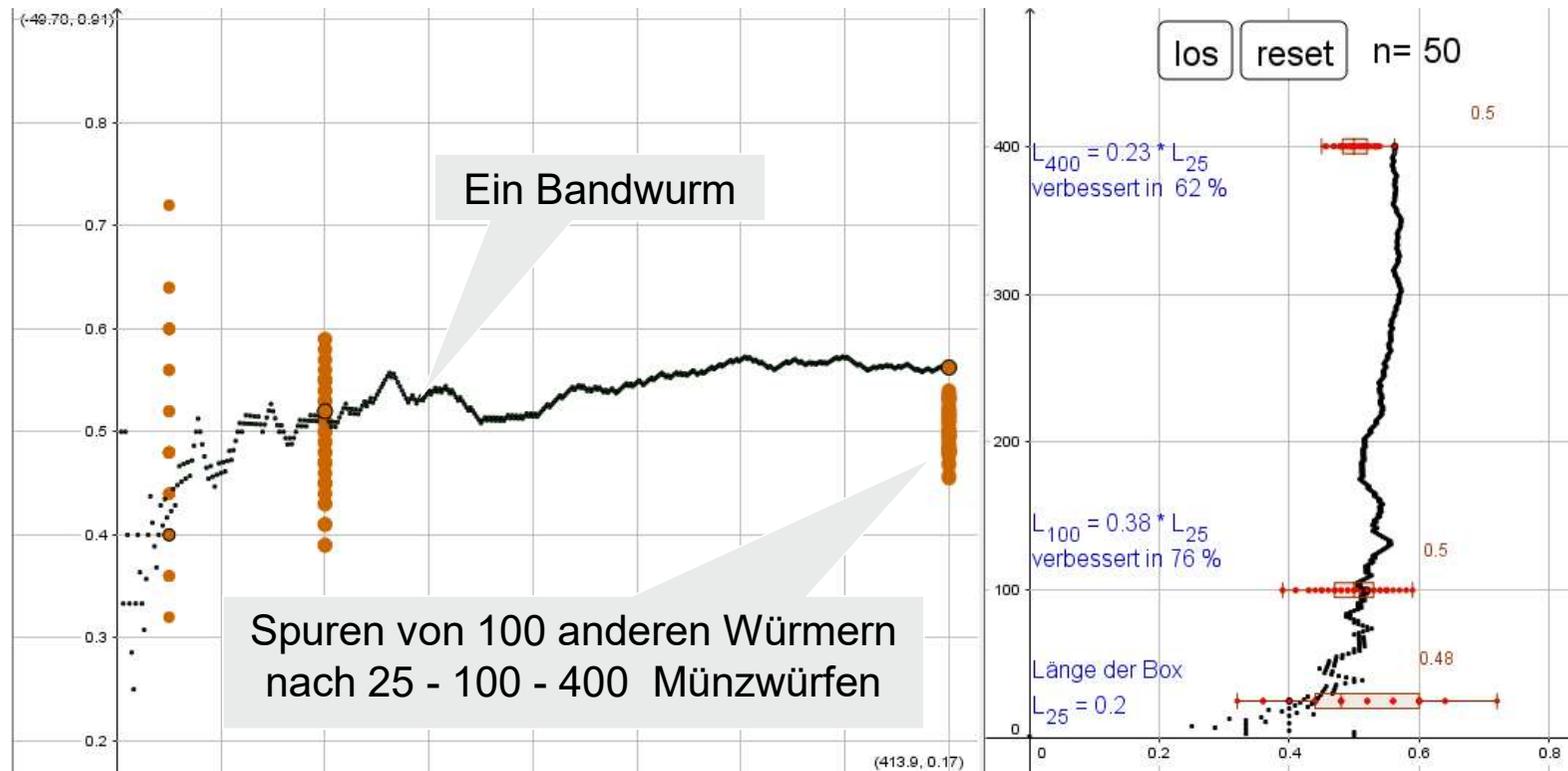
**Lord Voldemord:**

Der, dessen Name nicht genannt werden darf / Du weißt schon wer.

**Frequentistische Wahrscheinlichkeit:**

Die, deren Wert man nicht bestimmen kann / Du weißt schon welche.

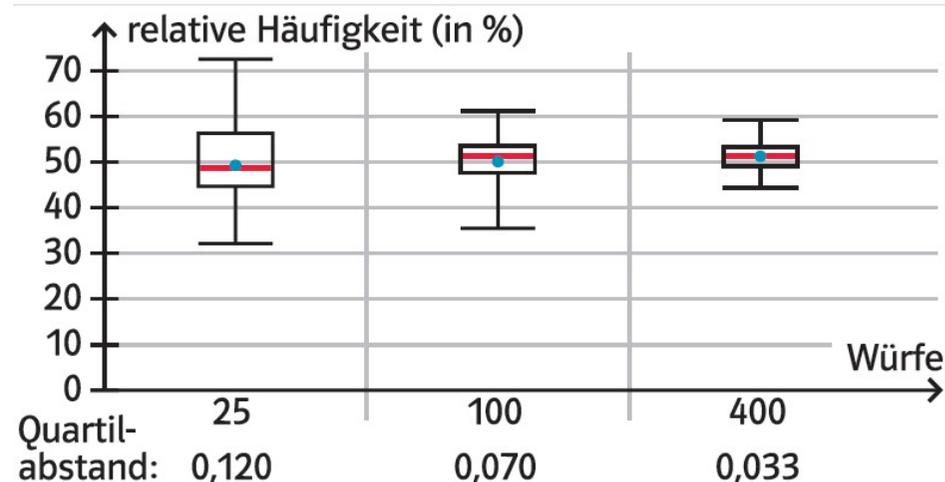
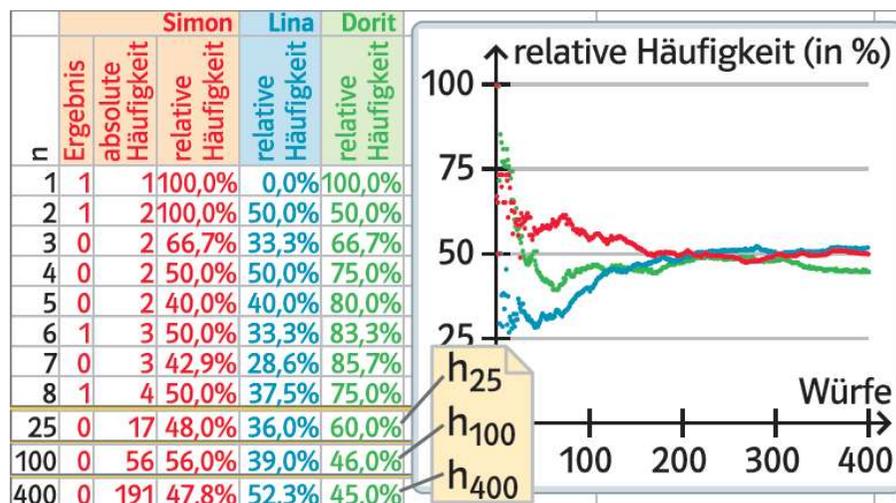
# Viele kurze Serien sind besser als eine lange



**Fundamentalsatz** - mit Boxplots ab Klasse 7: Bei **vierfachem** Versuchsumfang streut die r. H. nur **halb** so stark um die **vorgegebene** Wk.

# Lambacher-Schweizer Klasse 7

Das große Vertrauen in Schätzungen, die auf vielen Versuchen beruhen, lässt sich durch das **Gesetz der großen Zahlen** rechtfertigen. Es besagt: Die Abweichungen der relativen Häufigkeiten von den Wahrscheinlichkeiten werden mit steigender Versuchszahl tendenziell kleiner.



knackig: viermal so viel – doppelt genau

# $1/\sqrt{n}$ Gesetz: Prognoseintervall für rel Hfgk. h

Fast immer ( $\sim 95,4\%$ ) liegt h

- beim fairen Münzwurf in  $\left[0,5 - \frac{1}{\sqrt{n}}; 0,5 + \frac{1}{\sqrt{n}}\right]$

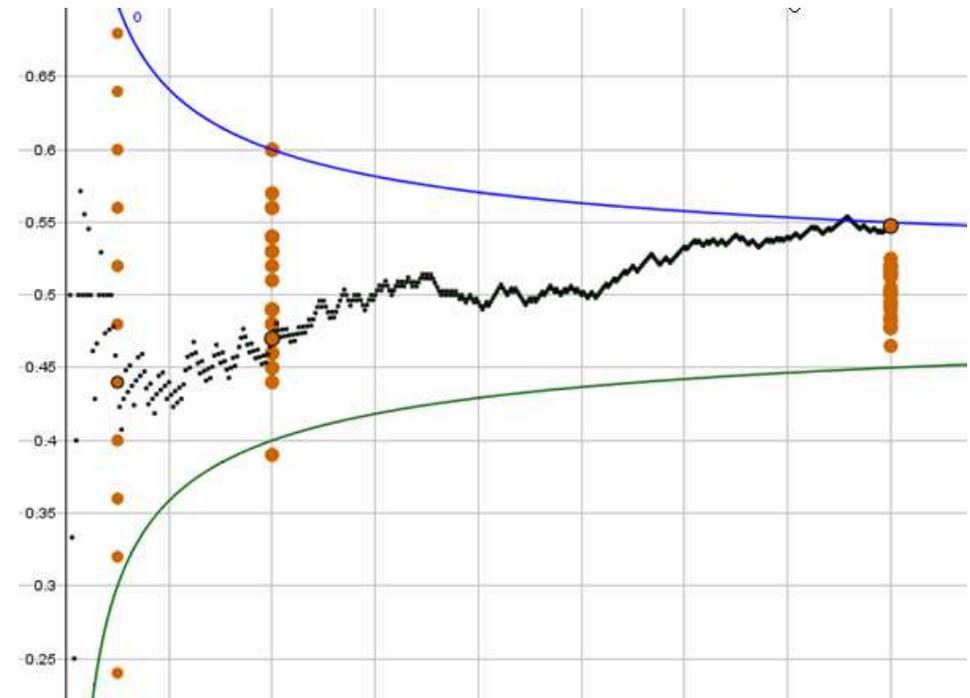
- allgemein  $\left[p - 2 \frac{\sqrt{p(1-p)}}{\sqrt{n}}; p + 2 \frac{\sqrt{p(1-p)}}{\sqrt{n}}\right]$

**NUR ! wenn Du unsicher bist: Bezweifle**

dass  $p=0,5$  ein gutes Modell ist,

wenn  $h \notin \left[0,5 - \frac{1}{\sqrt{n}}; 0,5 + \frac{1}{\sqrt{n}}\right] \dots$

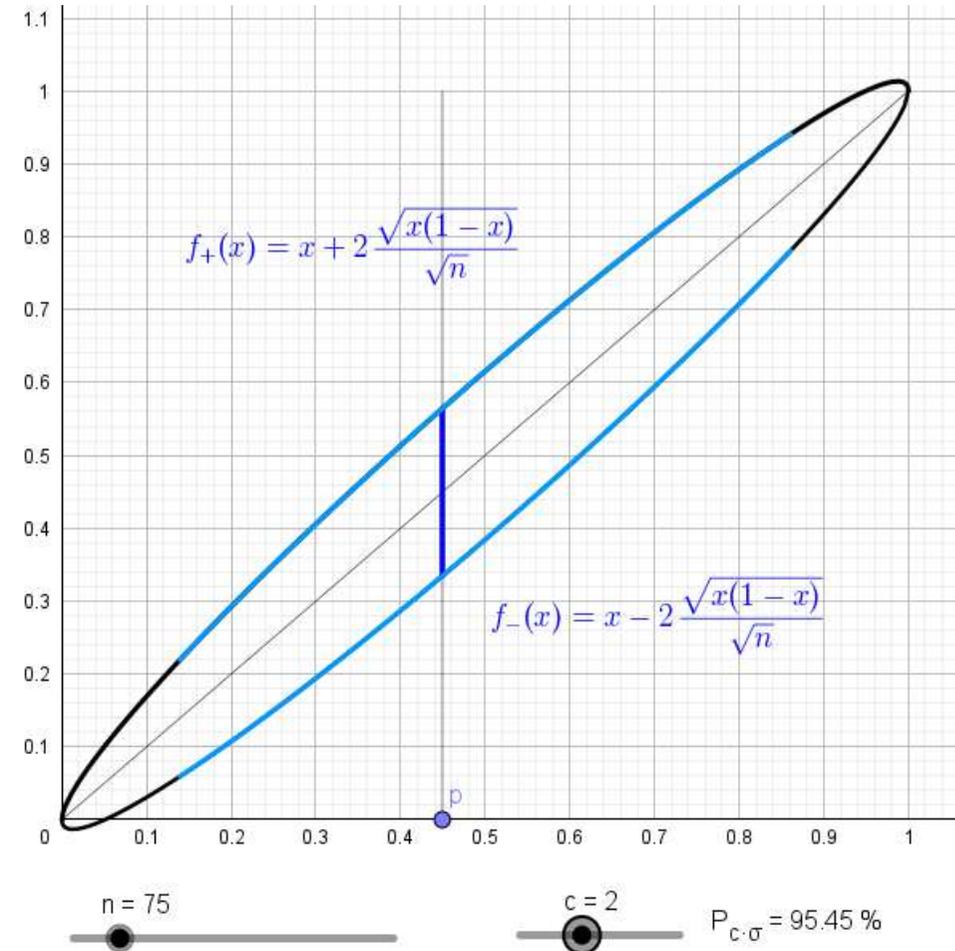
wenn eine **signifikante Abweichung**“ der Wirklichkeit vom Modell  $p=0,5$  vorliegt!



# So hängt das Prognoseintervall von p ab: Ellipsendiagramm

Prognoseintervalle werden umso kleiner

- je weiter p sich von 0.5 entfernt
- je größer n



# Umfrage I

Bernd lag 7 von 12mal außerhalb des Prognoseintervalls...

$$\left[0.5 - \frac{1}{\sqrt{n}}; 0.5 + \frac{1}{\sqrt{n}}\right]$$

	25	100	400
bernd 1	15	17	73
bernd 2	13	31	200
bernd 3	19	68	150
bernd 4	9	48	260

n = 25

n = 100

n = 400

[0.3; 0.7]

[0,4; 0,6]

[0.45; 0.55]

relativ

[7.5; 17.5]

[40 ; 60]

[180; 220]

absolut

1mal

3mal

3mal

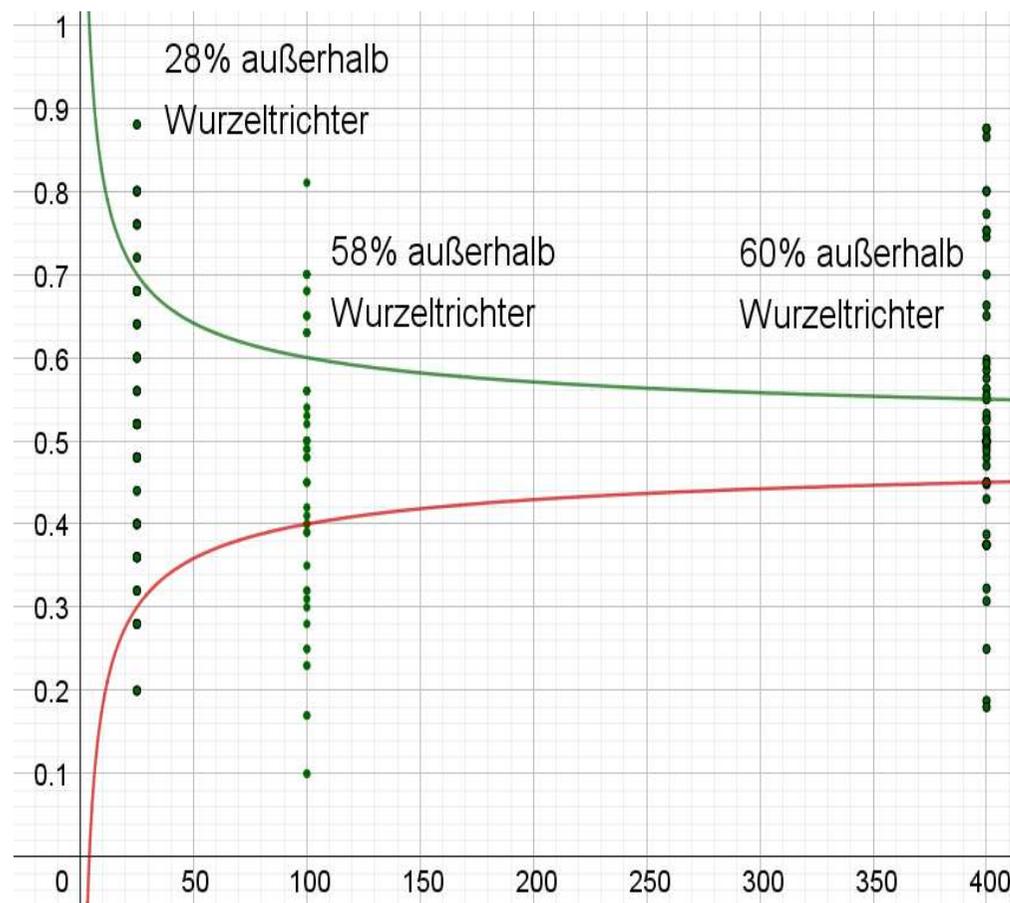
->

7mal

Auch ich lag mehr als einmal außerhalb

ja     nein

				28%	58%	60%
	25	100	400	25	100	400
bernd 1	15	17	75	0.6	0.17	0.1875
bernd 2	13	31	200	0.52	0.31	0.5
bernd 3	19	68	150	0.76	0.68	0.375
bernd 4	9	48	260	0.36	0.48	0.65
denise 1	14	56	220	0.56	0.56	0.55
denise 2	8	39	239	0.32	0.39	0.5975
denise 3	9	41	179	0.36	0.41	0.4475
denise 4	13	63	211	0.52	0.63	0.5275
gregor 1	17	32	150	0.68	0.32	0.375
gregor 2	7	53	280	0.28	0.53	0.7
gregor 3	10	70	350	0.4	0.7	0.875
gregor 4	19	17	220	0.76	0.17	0.55
silke 1	10	39	225	0.4	0.39	0.5625
silke 2	9	45	213	0.36	0.45	0.5325
silke 3	13	52	180	0.52	0.52	0.45
silke 4	14	63	172	0.56	0.63	0.43
simone 1	17	48	129	0.68	0.48	0.3225
simone 2	9	23	298	0.36	0.23	0.745
simone 3	12	32	346	0.48	0.32	0.865
simone 4	18	65	309	0.72	0.65	0.7725
steffi 1	10	25	150	0.4	0.25	0.375
steffi 2	15	10	200	0.6	0.1	0.5
steffi 3	5	50	350	0.2	0.5	0.875
steffi 4	20	70	100	0.8	0.7	0.25



# Prognoseintervalle und der kreiselnde Penny



Am 25.04.2023 um 21:27 schrieb J. Konietzko:

Ich habe keine Info darüber, wie sie sich bei einem klassischen Münzwurf verhält, aber wenn man sie rotieren lässt, bleibt sie mit 80-prozentiger Wahrscheinlichkeit mit dem Kopf nach unten liegen, da dieser ungewöhnlich weit heraussteht und so den Schwerpunkt minimal zu dieser Seite hin verschiebt.

Film: [pennyexperiment-Klasse 8](#)

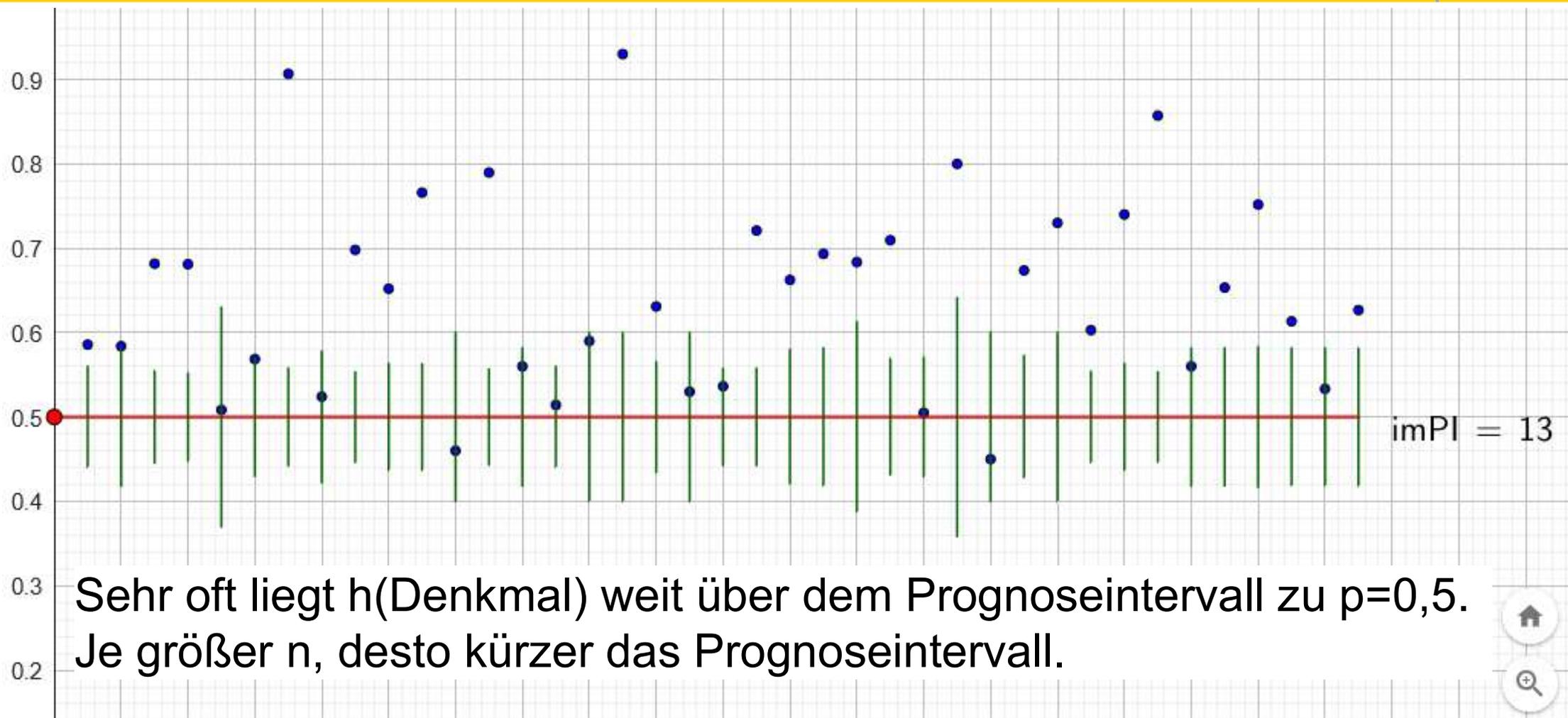
Film: [tom-scotts youtube-kanal](#)

Listen to full episodes 



**Abby and Bella can't decide where to eat. Abby takes out a US 1 cent coin and says, 'Let's use this to decide fairly'. Abby is able to give herself an 80 per cent chance of winning without Bella suspecting anything. How?**

# h(Denkmal) bei 39 Pennys



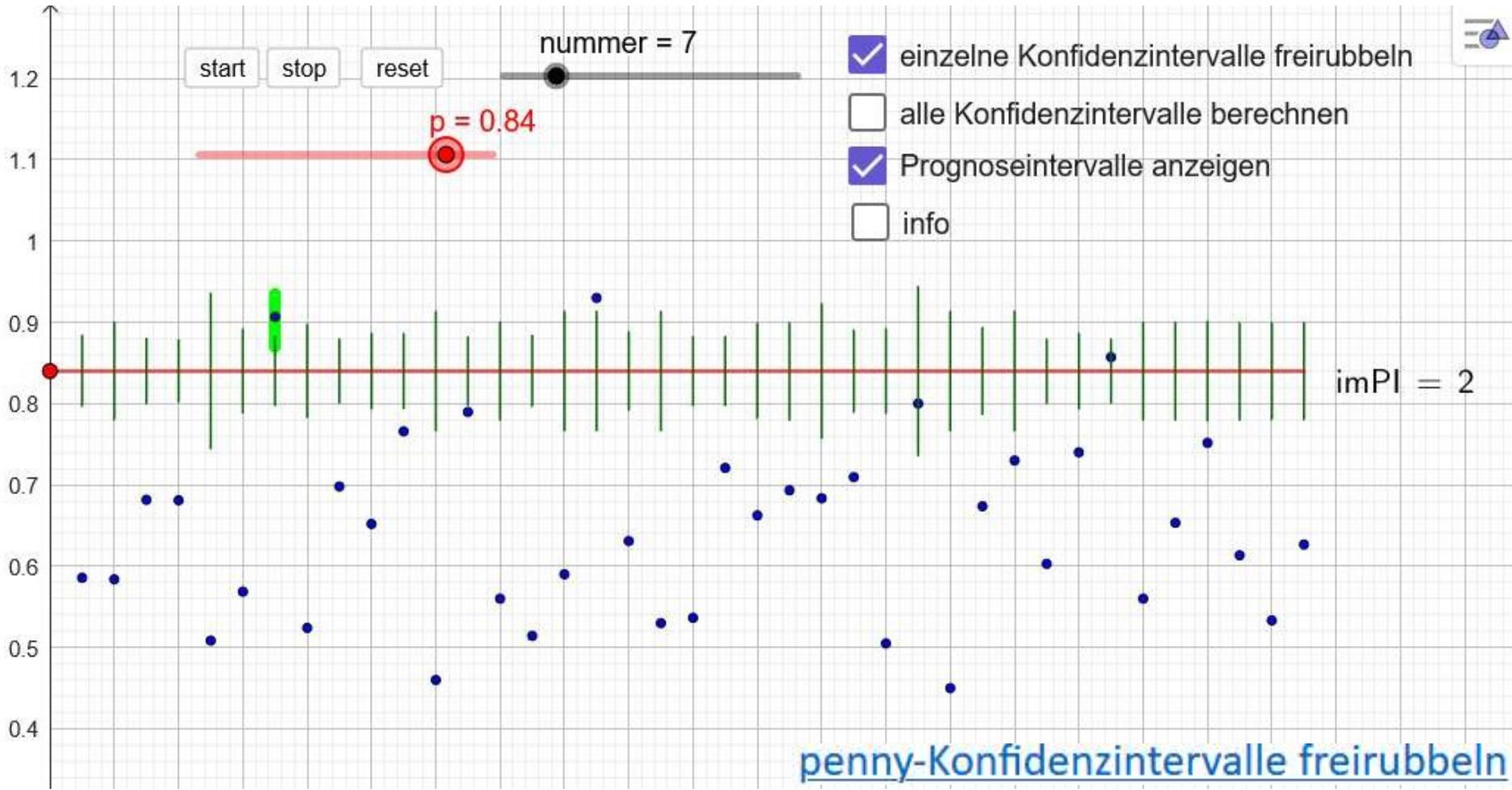
## Wahrscheinlichkeit als Modell – Konfidenzintervalle

- Stolpersteine bei Signifikanztests
- Prognoseintervalle (und das Konzept des Bezweifelns)
- Konfidenzintervalle
  - Begriffsbildung digital
  - klassisch-analog
- Resümee

## Mediendidaktik:

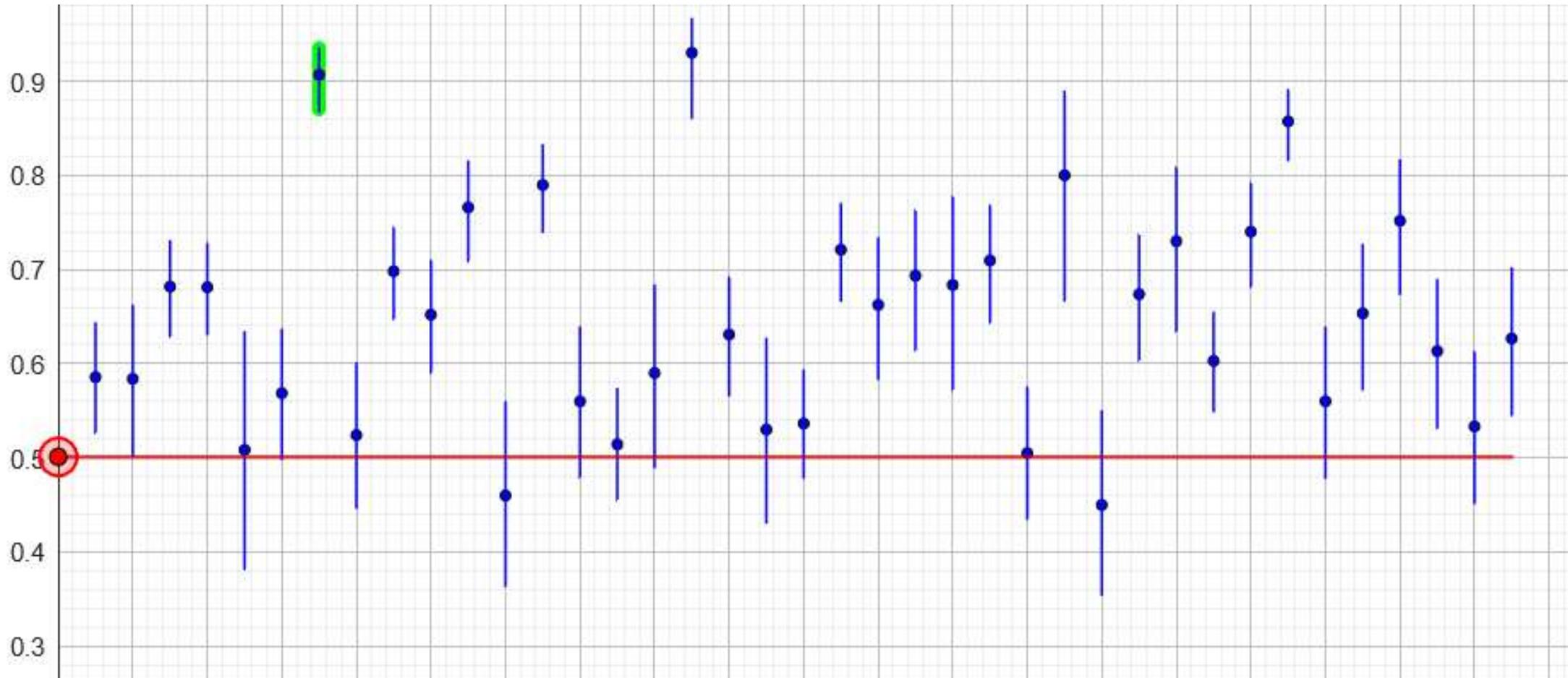
- Blackbox-Whitebox Prinzip
- SAMR Stufe IV (Reconstruction)





Die probierten Wahrscheinlichkeiten  $p$ , in deren Prognoseintervall die beobachtete rel. H. liegt, hinterlassen eine Schleifspur, die das Konfidenzintervall der vertrauenswürdigen Wahrscheinlichkeiten bildet.

$p = 0.5$  ist nur für wenige Pennys ein brauchbares Modell.



# Schätzen statt Testen

## Wahrscheinlichkeit als Modell – Konfidenzintervalle

- Stolpersteine bei Signifikanztests
- Prognoseintervalle (und das Konzept des Bezweifelns)
- Konfidenzintervalle
  - Begriffsbildung digital
  - klassisch-analoge Variante
- Resümee

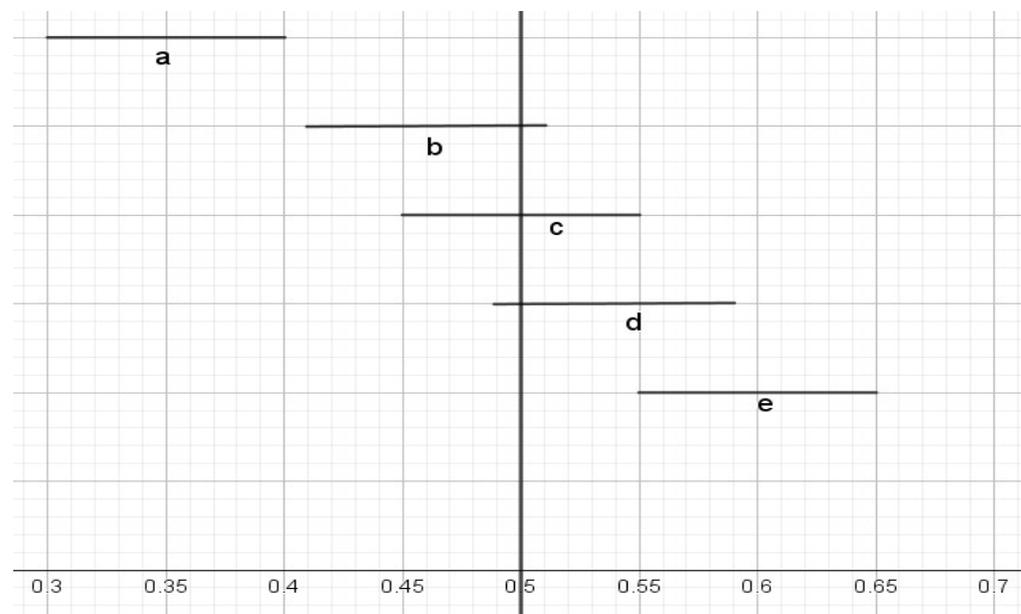
# Umfrage II

## Dreisatz: Spekulieren – Experimentieren – Reflektieren

Wo vermuten Sie „die“  
Wahrscheinlichkeit für  
Fußlage

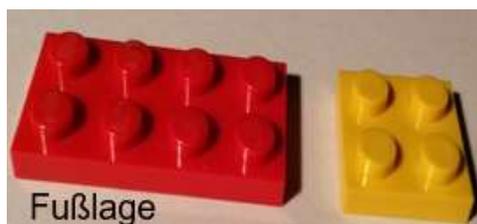


- a) deutlich unter 50%
- b) eher unter 50%
- c) ziemlich genau 50%
- d) eher über 50%
- e) deutlich über 50%



## Vom Konzept des Bezweifelns zum Konfidenzintervall

- a) Würfeln Sie den flachen Klemmbaustein 100-mal (freier Fall aus Ellenbogenhöhe).  
Notieren Sie die relative Häufigkeit von Fußlage!



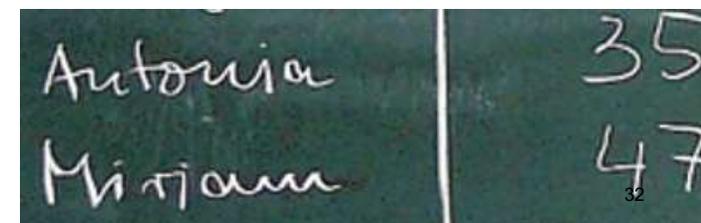
Autoria	35
Miriam	47

- b) Prüfen Sie, ob das Versuchsergebnis  
(i) nach Bauchgefühl (ii) nach  $1/\sqrt{n}$  – Gesetz  
Anlass bietet, zu bezweifeln, dass  $p(\text{Fußlage}) = 0,5$  ein  
gutes Modell für den flachen Legostein ist.
- c) Suchen Sie nach möglichst vielen Wahrscheinlichkeiten  
für  $p(\text{Fußlage})$ , die Sie nicht anzweifeln brauchen!

In welchem Intervall müsste bei 100 Würfeln die relative Häufigkeit für „Fuß“ liegen, wenn gilt

p	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
bis	0.000	0.160	0.280	0.392	0.498	0.600	0.698	0.792	0.880	0.960	1.000
von	0.000	0.040	0.120	0.208	0.302	0.400	0.502	0.608	0.720	0.840	1.000
100				A	A	M	M				

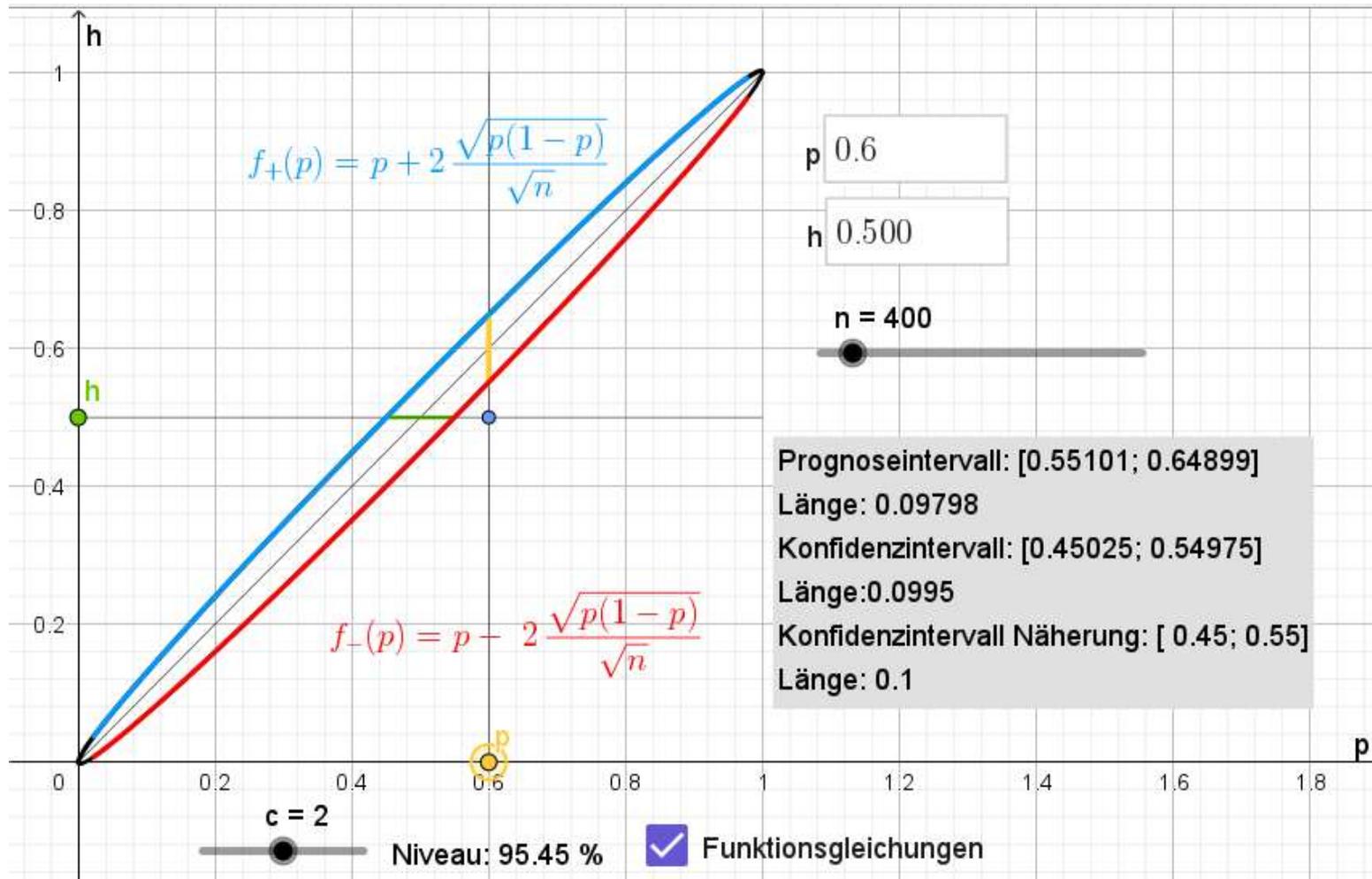
Suche alle diejenigen Wahrscheinlichkeiten, die von Antonia/Mirjam nicht angezweifelt werden bräuchten





# Konfidenzintervall berechnen Asset als Rechenwerkzeug

prognose-konfidenzintervall-tool



## Konfidenzintervall

linke Grenze

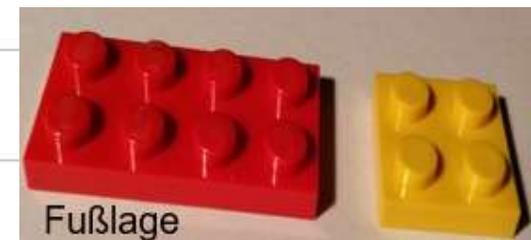
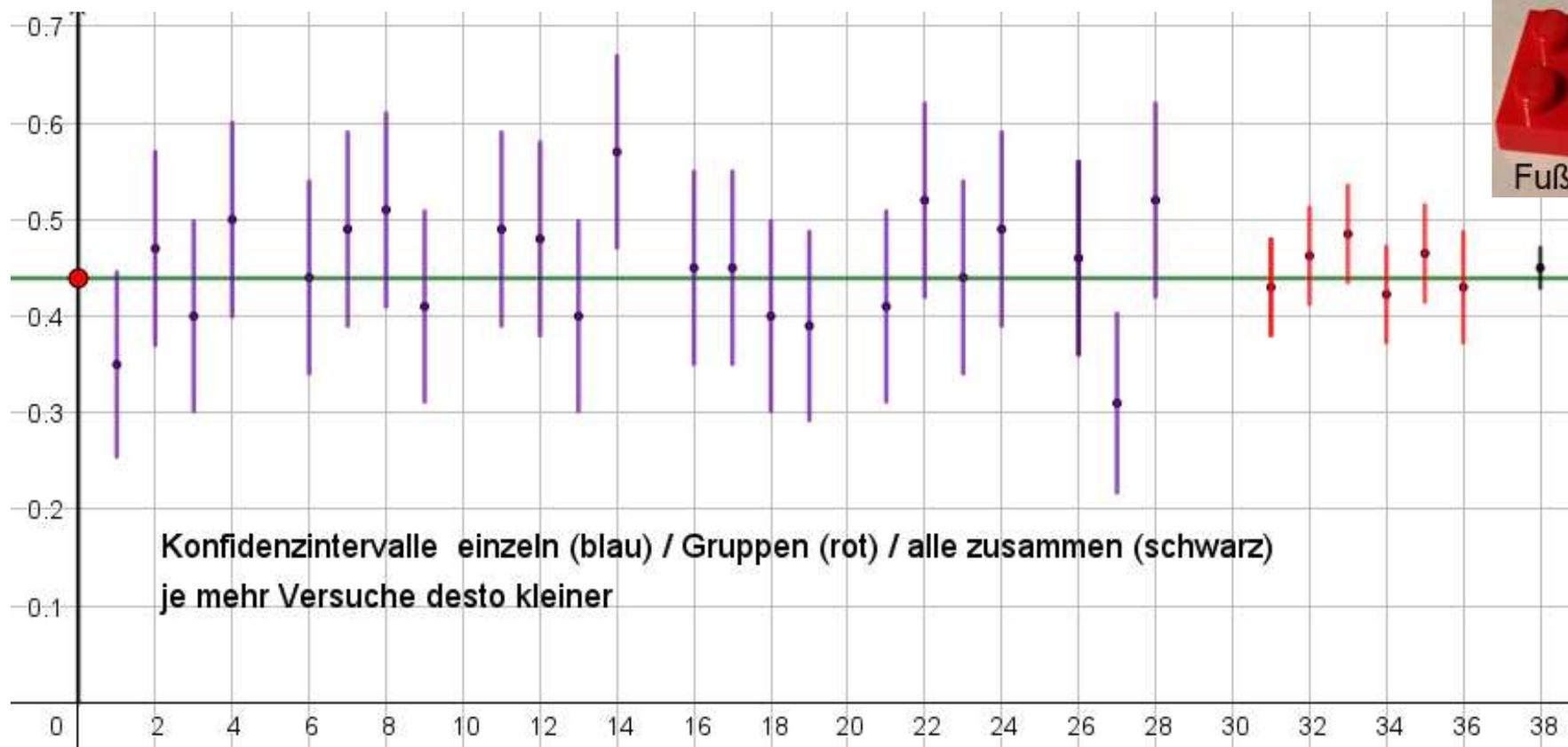
$$\text{solve } x + 2 \frac{\sqrt{x(1-x)}}{\sqrt{n}} = h$$

rechte Grenze

$$\text{solve } x - 2 \frac{\sqrt{x(1-x)}}{\sqrt{n}} = h$$

# Lego-Münzen in Bechern „würfeln“

$p(\text{Fußlage}) = 0.44$  ist ein sehr gutes Modell



Stolpersteine bei Signifikanztests

Prognoseintervalle (und das Konzept des Bezweifelns)

Konfidenzintervalle

- Begriffsbildung digital
- klassisch-analogue

Resümee

# • Kurz und knackig ... fürs Regelheft

- „**vertikale**“ Prognoseintervalle „sind **fest**“ sie hängen von  $p$  ab, sie enthalten relative Häufigkeiten  $h$ , sie sind **symmetrisch** zu  $p$
- „**horizontale**“ Konfidenzintervalle „sind **zufallsabhängig**“, sie hängen von  $h$  ab, sie enthalten Wahrscheinlichkeiten  $p$  sie sind **unsymmetrisch** zur rel. Hfgk.  $h$  (Ausnahme  $h=0.5$ )
- In Konfidenzintervallen sammelt man nach einer Beobachtung vertrauenswürdige Wahrscheinlichkeitsmodelle
- Dualität:  $p$  liegt ganz genau dann im Konfidenzintervall zu  $h$  (ist ganz genau dann vertrauenswürdig), wenn  $h$  im Prognoseintervall von  $p$  liegt (nichts Ungewöhnliches ist)
- In  $\sim 95\%$  aller Fälle überdeckt das zufallsabhängige Konfidenzintervall „die“ unbekannte“ Wahrscheinlichkeit ... falls man deren Existenz unterstellt
- Wir haben in alle „Modelle“, die im Konfidenzintervall liegen, ein **nicht quantifizierbares** Vertrauen, unter uns: **in die  $p$ 's vom Rand** ein bisschen **weniger** als in die  **$p$ 's aus der Mitte!**
- Bei hohem Stichprobenumfang wird der Bereich vertrauenswürdiger Modelle fuzzi-klein. Signifikante Unterschiede werden inhaltlich irrelevant...

- Kurz und knackig ... fürs Regelheft

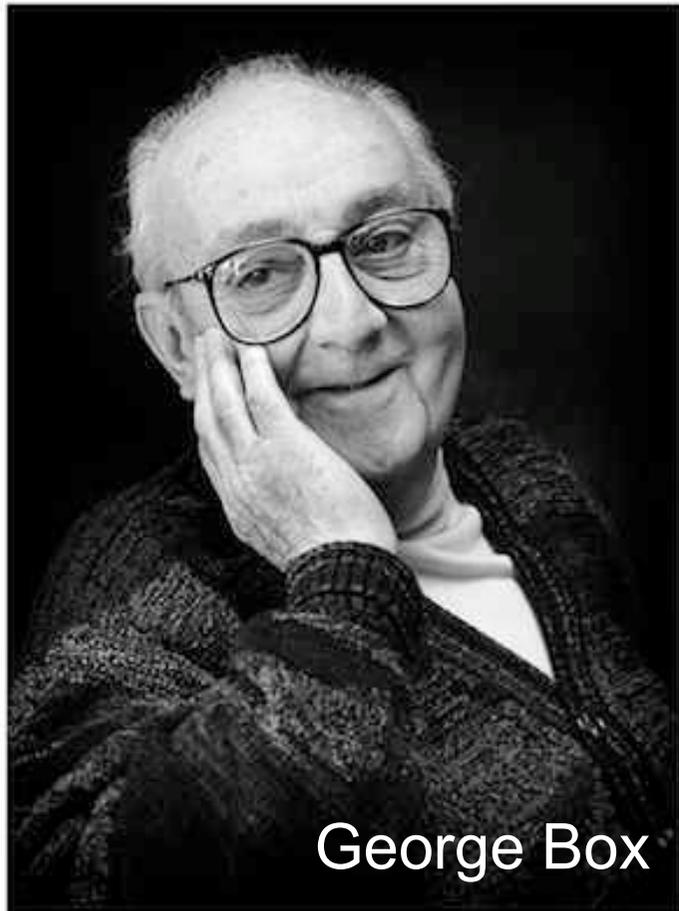
## Prognoseintervall

- Modell  $\rightarrow$  Wirklichkeit
- $p$  gegeben  $\rightarrow$   $h$  gesucht
- Sicherheitswahrscheinlichkeit  $P$
- enthält relative Häufigkeiten
- ist symmetrisch zu  $p$
- liegt in der Ellipse vertikal
- ist fest
- $p$  ist vertrauenswürdig, wenn  $h$  im  $2\sigma$ -Prognoseintervall zu  $p$  liegt

## Konfidenzintervall

- Wirklichkeit  $\rightarrow$  Modell
- $h$  gegeben  $\rightarrow$   $p$  gesucht
- Konfidenzniveau  $\beta$
- enthält Wahrscheinlichkeiten
- ist annähernd symmetrisch zu  $h$
- liegt in der Ellipse horizontal
- ist zufallsabhängig
- $p$  ist vertrauenswürdig, wenn es im Konfidenzintervall ( $\beta = 0,954$ ) zu  $h$  liegt

All models are wrong but some are useful



George Box

Sammeln Sie  
die nützlichen,  
die weniger unsicheren,  
die vertrauenswürdigen  
Modelle  
im Konfidenzintervall!



## Prognose/Konfidenzintervalle



Sonderdruck  
Hessen/Thüringen  
[www.klett.de](http://www.klett.de)  
Code: v6i59w

