

# Att undervisa i matematisk statistik

## Timo Koski, KTH

Cramér 4/12/2015

December 7, 2015

En acklimatisering: grundkursen, sf1901 sannolikhet och statistik, heter **sannstatkursen** bland teknologerna vid KTH.



## *Att undervisa i sannstat*

# En anslagstavla någonstans vid Teknikringen

KTH Matematik

Cramér  
4/12/2015



Cramér  
4/12/2015

# ÅNGEST INFÖR SANSTAT-TENTAN?

SANSTAT.SE

*ludu*

För studenter, av studenter

## **"Trodde du sanstat var obegripligt? Think again."**

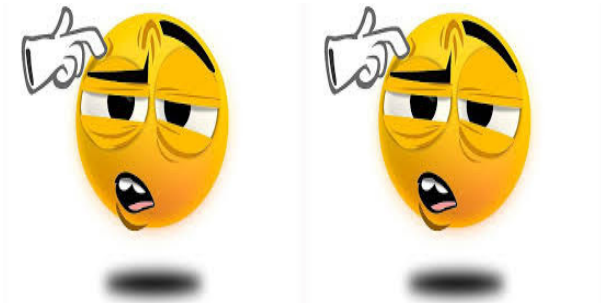
Vi alla har varit där. Allt är frid och fröjd tills det kommer en hattrick med obegripliga föreläsningar som kommer som sabbar det. Snabbspola några veckor till tentaperioden - man ska precis försöka förklara till kompisen hur man löste uppgiften och märker att man tittat i facit en gång för mycket. Med denna kurs vill jag underlätta för dig som läser sanstat och sudda ut ett flertal av dina frustrationsmoment. Förklarat från en student till en annan.

*Ett utfall är det direkta resultat vi kan observera från en sannolikhetsbaserad situation. Ett utfallsrum är mängden av alla möjliga utfall som kan uppstå. Utfallsrummet betecknas med  $\Omega$  medan individuella utfall betecknas med  $\omega_i$ . Notera att utfallsrummet inte har något att göra med vad den faktiska sannolikheten för varje värde att uppstå, den snarare beskriver alla värden som*

- 1 "en sannolikhetsbaserad situation" (?)
- 2 Bortsett från frågan ovan, är webbkursen faktiskt så annorlunda än vad som anförs av de, som anmodas, obegripliga föreläsarna?



*"En stokastisk variabel är egentligen en stokastisk funktionsvariabel som är definierad på ett utfallsrum."*



A random variable is neither random nor variable.  
— Gian-Carlo Rota



## ***How to Convince a Student that an Estimator is a Random Variable***

---

KERSTIN VÄNNMAN

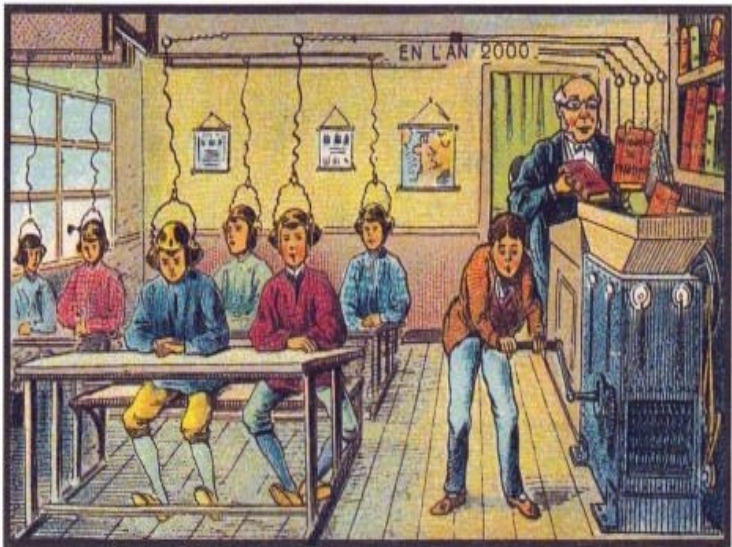
At the University of Luleå, all students of engineering subjects take a compulsory course in statistics during their second semester (out of eight). The course contains

Antoine Gombaud, Chevalier de Méré, hade visserligen lärt sig matematik, men tillämpade den felaktigt i en



sannolikhetsbaserad situation.

Forte, James A: Teaching statistics without sadistics, *Journal of Social Work Education*, 31, 2, pp. 204–218, 1995.



University of Sweden vore inte bara en tillgång för kunskapstörstiga utanför landet. Det ger även svenska studenter tillgång till de bästa kurserna landet förmår prestera med de bästa lärarna.

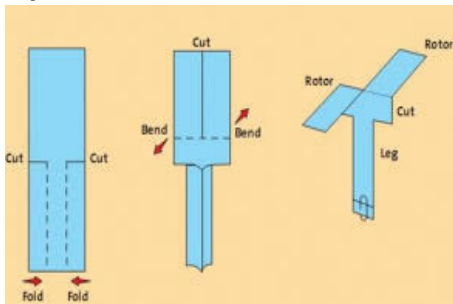
Emil Källström, Fredrik Lagergren, Per Ödling i DN-Debatt  
2013-11-03

# George Box helikopter (1)

KTH Matematik

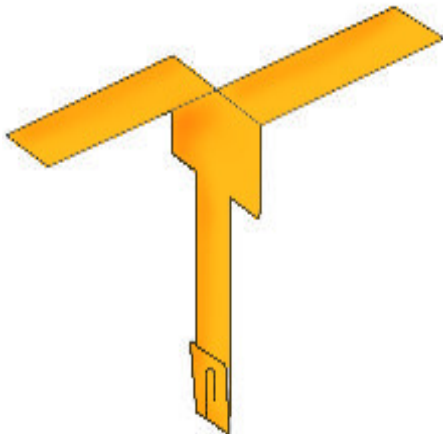
Cramér  
4/12/2015

<https://www.minitab.com/en-us/Published-Articles/Teaching-DoE-with-Paper-Helicopters-and-Minitab>



## George Box helicopter (2)

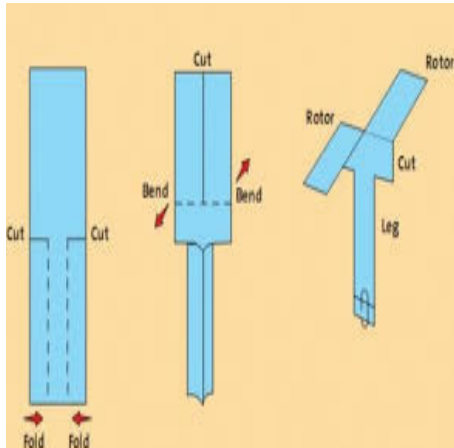
Box, George EP: Teaching engineers experimental design with a paper helicopter, *Quality Engineering*, 4, 3, 1992.



# George Box helikopter (3)

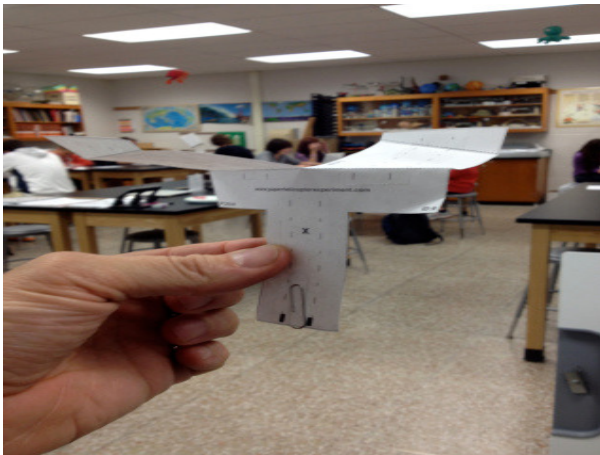
Bisgaard, Søren: Teaching statistics to engineers, *The American Statistician*, 45, 4, pp. 274–283, 1991.

S.B. beskriver sin kurs (utvecklad tillsammans med G.E.P. Box) för teknologer vid Univ. of Wisconsin, Madison.



# George Box helikopter (4)

*As a key point, we do not really teach statistics in the traditional sense. Instead, we teach engineering problem solving that incidentally requires statistics*



*After this introduction we go directly to a problem related to the development of a new product. (No coin flipping or red and blue balls first.) To make it real, we have a small paper helicopter (our product) that we bring to the classroom. We tell the students that we are in the process of developing a better helicopter and, therefore, need to test this prototype. Specifically, we tell them that an important characteristic of the helicopter is flight time.*

*We then climb up on a ladder and drop the helicopter four times from the ceiling, measure the time it takes to hit the floor, and, of course, get different numbers. Next we plot the data as a dot diagram showing that the first step in any analysis is to plot the data. From the plot, we develop the ideas of location and dispersion and show how to quantify these notions in terms of the average and the standard deviation.*



*is about conceptual change, not just the acquisition of information*

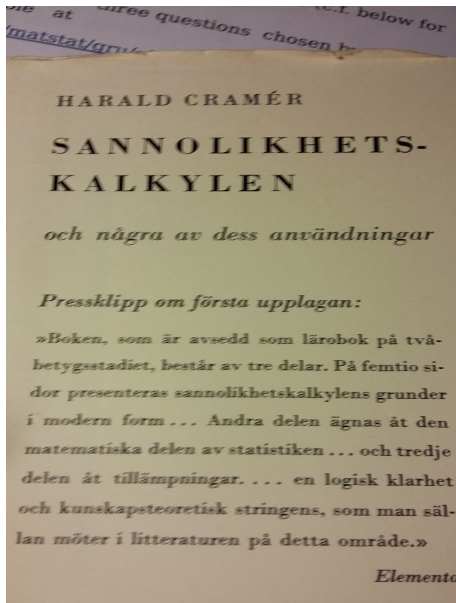
—Biggs, J. & Tang, C.: **Teaching for Quality Learning at University**, McGraw-Hill and Open University Press, Maidenhead, 2007.

# Undervisningen i sanstat har rötterna i kursböcker författade av Harald Cramér

KTH Matematik

Cramér  
4/12/2015





lan möter i litteraturen på detta område.»

*Elementa*

»... språket (är) lättflytande och läsningen ett  
nöje. Boken innehåller talrika övningsexempel  
med svar.»

*Expressen*

»... saknar inte ens ett stänk av humor.»

*Aftonbladet*

# Harald Cramér : syntes av Kolmogorov, Neyman-Pearson & Fisher

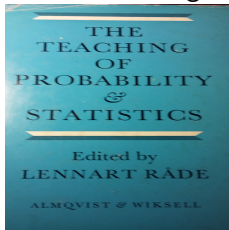
- 1) En punktskattning  $\theta_{\text{obs}}^*$  av  $\theta$  är *väntevärdesriktig* om  $E(\theta^*(X_1, \dots, X_n)) = \theta$ .
- 2) En punktskattning  $\theta_{\text{obs}}^*$  av  $\theta$  är *konsistent* om  $P(|\theta^*(X_1, \dots, X_n) - \theta| > \varepsilon) \rightarrow 0$  då  $n \rightarrow \infty$ .
- 3) Om  $\theta_{\text{obs}}^*$  och  $\theta_{\text{obs}}^{**}$  är väntevärdesriktiga skattningar av  $\theta$  så säger man att  $\theta_{\text{obs}}^*$  är *effektivare* än  $\theta_{\text{obs}}^{**}$  om  $V(\theta^*(X_1, \dots, X_n)) < V(\theta^{**}(X_1, \dots, X_n))$ .

# Sir R.A. Fisher (1938)

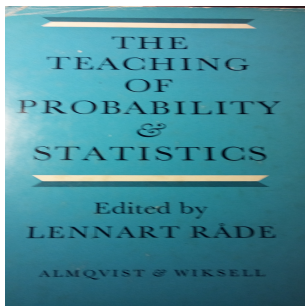
*... the responsibility for the teaching of statistical methods in our universities must be entrusted, certainly to highly trained mathemati(cal statisti)cians, but only to such mathemati(cal statisti)cians as have had sufficient prolonged experience of practical research, and of responsibility for drawing conclusions from actual data, upon which practical action is to be taken. Mathematical acuteness alone is not enough.*

# CSMP= Comprehensive School Mathematics Program

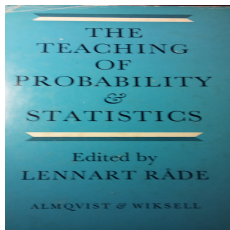
Proceedings of the first CSMP international conference co-sponsored by Southern Illinois University and Central Midwestern Regional Educational Laboratory. 969: Carbondale,



*... while all pedagogical aspects ... are of deep concern, priority is given to the **selection of content** that is sound, relevant and enjoyable*



Bidrag av bl.a. E.F. Beckenbach, C.B. Bell, W.Kruskal,  
D.V. Lindley, J. Neyman, C.R. Rao, A. Renyi, S. Holm.  
Konferensen handlar om 'the content of pre-college curriculum'.



*I would never recommend that we teach any technique of mathematical statistics to college freshmen or high school students.*

*Mathematical statistics, though invented to handle numerical data with critical mind is often used to substitute mechanics for criticism.*

## A non-frequentist view of probability and statistics

$\mathbf{x}$  = the outcomes of  $n$  flips of a coin.

$$P(X_{n+1} = \text{head} | \mathbf{x}) = \int_0^1 P(\text{head} | \Theta = \theta) \cdot \pi(\theta | \mathbf{x}) d\theta$$

- *prediktion av observabla är en mer fundamental uppgift än inferens om icke-observabla.*
- *bedömningar om (partiell) utbytbarhet utgör grunden för prediktiv modellering.*
- *Inferens baserar sig på observerade  $\mathbf{x}$ , ej på ett icke-observerat utfallsrum.*
- *$\pi(\theta | \mathbf{x})$  är den enda kvantitet som beräknas för inferens om  $\theta$  (substitute mechanics for criticism ?).*

Barry, MDJ and Steele, NC: A core curriculum in mathematics for the European engineer: an overview *International Journal Of Mathematical Education in Science and Technology*, 24, 223–229, 1993,



a.a.

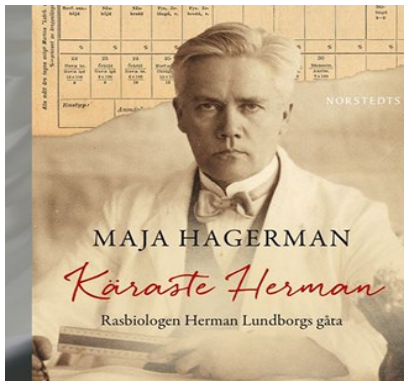
KTH Matematik

Cramér  
4/12/2015

What is the essence of statistical thinking?

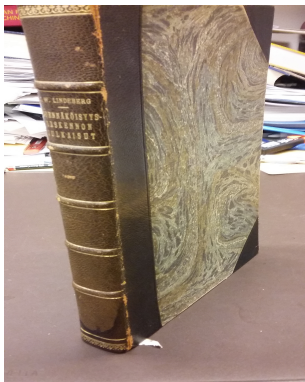
– ”samlade kranier, mätte, fyllde hålkort med oändliga data från de många tiotusentals individer han kartlade”.

Cramér  
4/12/2015



# J.W. Lindeberg: Todennäköisyysslasku, Otava, Helsinki 1927

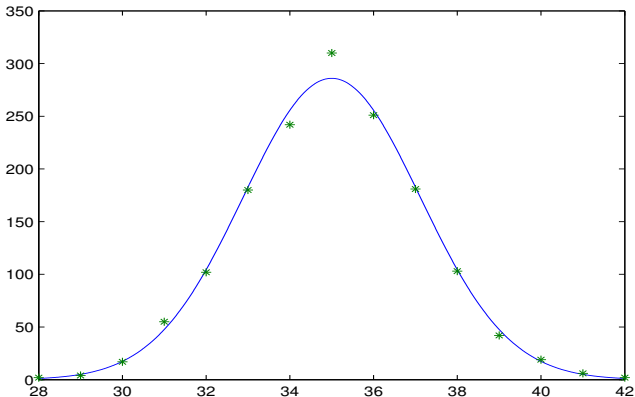
En kursbok för universitetsstudenter i lant- och skogsbruk från 1927. Lindeberg tycks ej känna till maximum likelihood och han finner ingen användning för stokastiska variabler eller axiomatisk sannolikhet.



# J.W. Lindebergs introduktion till a.a.: omkretsen av bröstkorgen hos 1516 belgiska soldater (Quetelet)

| Class | Absolute frequency | The values of                              |
|-------|--------------------|--------------------------------------------|
| $i$   | $f_i$              | $g(x) = 286e^{-0.112(x-35)^2}$<br>$g(x_i)$ |
| 28    | 2                  | 1.2                                        |
| 29    | 4                  | 5.1                                        |
| 30    | 17                 | 17.4                                       |
| 31    | 55                 | 47.7                                       |
| 32    | 102                | 104.4                                      |
| 33    | 180                | 182.7                                      |
| 34    | 242                | 255.7                                      |
| 35    | 310                | 286.0                                      |
| 36    | 251                | 255.7                                      |
| 37    | 181                | 182.7                                      |
| 38    | 103                | 104.4                                      |
| 39    | 42                 | 47.7                                       |
| 40    | 19                 | 17.4                                       |
| 41    | 6                  | 5.1                                        |
| 42    | 2                  | 1.2                                        |

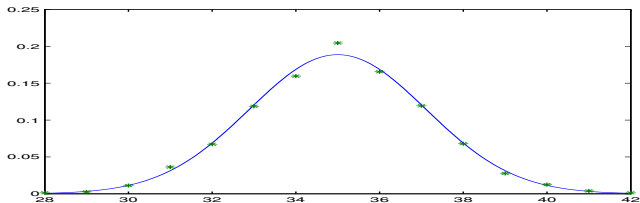
Plot av the frekvensdata och funktionen  $286e^{-0.112(x-35)^2}$



# 'stickprov ur en fördelning' !

Cramér  
4/12/2015

Plotten av de relativa frekvenserna av Quetelets data och funktionen  $\frac{286}{1516} e^{-0.112(x-35)^2}$ .



- S.M. Stigler: Darwin, Galton and the Statistical Enlightenment. *J.R. Statist. Soc. A.* (2010), 173, 3, pp 469–482

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \in N \left( \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \right).$$

Då gäller att

$$\frac{E[X_2 | X_1] - \mu_2}{\sigma_2} = \rho \frac{X_1 - \mu_1}{\sigma_1}. \quad (1)$$

$\rho$  är korrelationskoefficienten.

# Vad är en regression ?

Om  $|\rho| < 1$ , så är

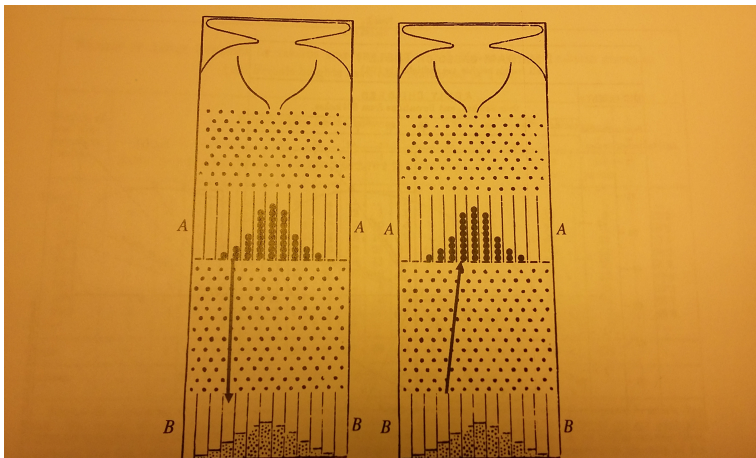
$$\frac{|E[X_2 | X_1] - \mu_2|}{\sigma_2} < \frac{|X_1 - \mu_1|}{\sigma_1}. \quad (2)$$

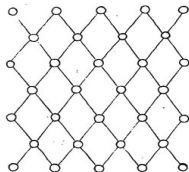
Ett statistiskt fenomen, **return to the mean** d.v.s. **regression** upptäckt av Sir Francis Galton, 1822 –1911.

# Galton's Quincunx (Galtons bräde)

KTH Matematik

Cramér  
4/12/2015





*Quid Quincunx speciosius, qui, in  
quam cuncta partem spectaueris,  
rectus est: Quintilian:¶*

THE  
G A R D E N  
OF  
C Y R U S.

OR,  
The Quincunciall, Lozenge,  
or Net-work Plantations  
of the Ancients, Artificially  
Naturally, Mystically  
Considered.

BY  
Thomas Brown D. of Physick

Printed in the Year, 1658.

C.R. Rao: Statistics: reflections on the past and visions for the future.  
Communications in Statistics , Theory and Methods, 30, pp.  
2235–2257, 2001

Statistical methodology has for long been and is being employed to find useful and usable information in data. During the recent years computer scientists have harnessed the power of computer technology to find useful and usable patterns in massive data sets.

On the other hand, statistics as a discipline has grown by solving practical problems in other subjects and by developing suitable methods to a given situation. Later these methods have consolidated to unified mathematical theories.

McKinsey Global Institute *Big data: The next frontier for innovation, competition, and productivity*, May 2011

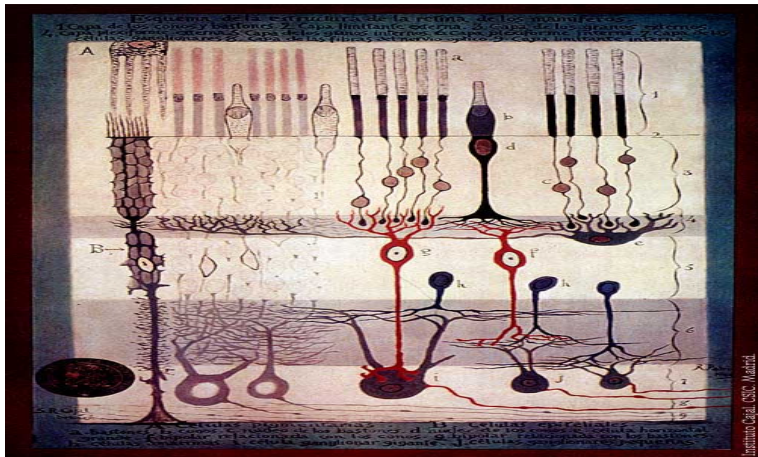
*p. 10: A significant constraint on realizing value from big data will be a shortage of talent, particularly of people with deep expertise in statistics . . .*

*The United States alone faces a shortage of 140,000 to 190,000 people with deep analytical skills as well as 1.5 million managers and analysts to analyze big data and make decisions based on their findings.*

# Surrealistiskt konstverk föreställande ångest och obegriplighet inför tentan i sannstat?

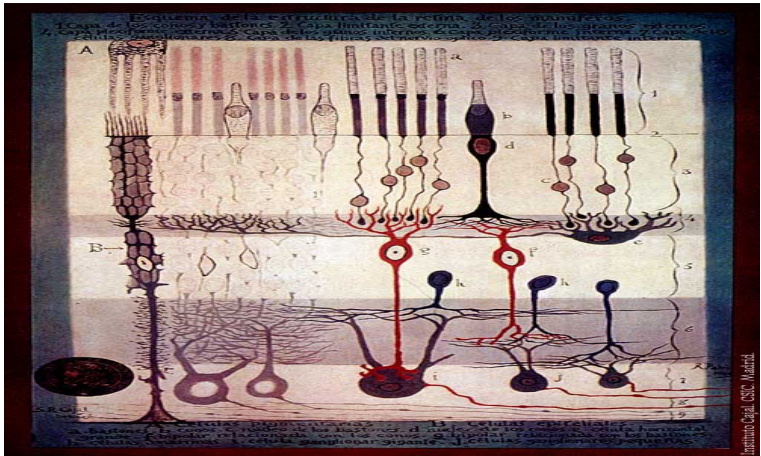
KTH Matematik

Cramér  
4/12/2015



Nej, Santiago Ramón y Cajals resultat om  
närhinnans cellstruktur, delade Nobelpriset i  
medicin år 1904 för detta.

Cramér  
4/12/2015



Instituto Cajal, CSIC, Madrid.