

Statistik





Per Bruun Brockhoff
Professor i Statistik
DTU Informatik
19 Marts 2010
<http://www.imm.dtu.dk/~pbb>



DTU Informatics
 Department of Informatics and Mathematical Modelling UNF Bornholm, 19 Marts 2010, © Per Bruun Brockhoff

WHOAMI



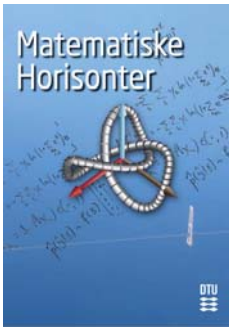


64-80: Født og opvokset i Vamdrup, Jylland
80-83: Munkensdam Gymnasium, Kolding
83-84: Værnepligt, Flyvestation Skrydstrup
84-91: Studerende Århus Universitet + diverse
91-94: Ph.d. studerende ved Landbohøjskolen (KVL)
94-04: Adjunkt og lektor i statistik på KVL
04- : Professor i Statistik, DTU Informatik
08- : Sektionsleder for 30 statistikere, DTU Informatik

1991: 5 mdr. University of Washington, Seattle, USA
1993: 6 mdr. University of California, Davis, USA
2001: 6 mdr. CSIRO, Sydney, Australia

DTU Informatics
 Department of Informatics and Mathematical Modelling UNF Bornholm, 19 Marts 2010, © Per Bruun Brockhoff

Bog for jer: (jeres lærere)

Matematiske Horisonter

DTU Informatics
 Department of Informatics and Mathematical Modelling UNF Bornholm, 19 Marts 2010, © Per Bruun Brockhoff



Øl og fladskærme
 - Statistik i aktion

Statistik



Hvad er statistik?




Mennesker som måleinstrumenter



DTU Informatics
 Department of Informatics and Mathematical Modelling UNF Bornholm, 19 Marts 2010, © Per Bruun Brockhoff

Hvad er Statistik



1. Et fantastisk spændende emne, som jeg altid har drømt om at bruge 2 timer en fredag eftermiddag på at høre mere om?
2. Noget kedeligt med tal, der kun overgås af revisorernes landsforening?
3. Jeg er f... ligeglad – hold dog op med at stille den slags latterlige spørgsmål!

DTU Informatics
 Department of Informatics and Mathematical Modelling UNF Bornholm, 19 Marts 2010, © Per Bruun Brockhoff



Matematik i medicinudvikling

AF PhD-studenten Anne Helge Jensen, PhD-studenten Søren Eick, PhD-studenten Sidsi Nilsen og Professor Henrik Holten, DTU Lyngby

Matematiske Horisonter

DTU

New York Times
5/8 2009:



Enlarge This Image

“People think of field archaeology as Indiana Jones, but much of what you really do is data analysis,” she said.

Now Ms. Grimes does a different kind of digging. She works at **Google**, where she uses statistical analysis of mounds of data to come up with ways to improve its search engine.

Ms. Grimes is an Internet-age statistician, one of many who are changing the image of the profession as a place for drowsy number nerds. They are finding themselves increasingly in demand — and even cool.

“I keep saying that the sexy job in the next 10 years will be statisticians,” said Hal Varian, chief economist at Google. “And I’m not kidding.”

DTU

Hvad er Statistik?

DTU

Egentlig ikke:

- Danmarks Statistik: <http://www.dst.dk/>
- Præsident Zuma og statistik: ISI meeting 17/8 2009. http://www.statssa.gov.za/isi2009/ISlbane_Newsletter_1.pdf
- Danske skolekarakterer: <http://www.karakter.dk/>

Men mere selve ANALYSEN og FORTOLKNINGEN:

- Statistik: Et af de 11 mest vigtige bidrag til medicin i 1000 år!!! <http://content.nejm.org/cgi/content/extract/342/1/142>
- Statistikere uundværlige for Google: New York Times 5/8 2009 <http://www.nytimes.com/2009/08/06/technology/06stats.html?scp=1&sq=statistics&st=Search>

DTU Informatics
Department of Informatics and Mathematical Modelling

UNF Bornholm, 19 Marts 2010, © Per Bruun Brockhoff

Øl og fladskærme

Statistik i aktion




DTU



Bang & Olufsen

DTU

- Interfaces such as touch screens are becoming more and more common



DTU Informatics
Department of Informatics and Mathematical Modelling

UNF Bornholm, 19 Marts 2010, © Per Bruun Brockhoff

Switch investigation



- *The relative importance of visual, auditory and haptic information for the user's experience of mechanical switches*
- Ditte Hvas Mortensen, Department of Psychology, Aarhus University, Denmark and Bang & Olufsen a/s, Struer, Denmark
- Søren Bech, Bang & Olufsen a/s, Struer, Denmark
- Ranking of switches



The sensory panel:



- **Trained assessors**
- **A sensory measurement instrument**
- **Evaluates products under controlled and designed conditions**
- **NOT preference evaluations**
- **NOT consumer studies**

What is sensometrics?

Sensometrics is the scientific area that applies mathematical and statistical methods to model data from sensory and consumer science.

Brockhoff, P.B. (2010). Sensometrics. In: *International Encyclopedia of Statistical Science*, Lovric, Miodrag (Ed.), Springer.

NEW Book for NON statisticians! (to appear)

DTU Informatics
Department of Informatics and Mathematical Modelling
UNF Bornholm, 19 Mars 2010, © Per Bruun Brockhoff

Sensometrics

(Another NEW Book – to appear)

STATISTICS FOR SENSORY AND CONSUMER SCIENCE
TORMOD NÆS, PER B. BROCKHOFF AND OLIVER TOMIC

WILEY

DTU Informatics
Department of Informatics and Mathematical Modelling

What is sensory?

Sensory Science

- Neurophysiology Anatomy
- Psychophysics
- Statistics Data analysis Chemometrics
- Psychology Cognitive science
- Marketing Consumer Science
- Chemistry Instrumental measurements
- Quality control
- Process Technology

DTU Informatics
Department of Informatics and Mathematical Modelling
UNF Bornholm, 19 Mars 2010, © Per Bruun Brockhoff

History: Since 1992!

Conferences:

Sensometrics (100-200)	Pangborn Sensory (500-700-900)
1992 Leiden, NL	1992 Järvenpää, Finland
1994 Edinburgh, Scotland	1995 Davis, USA
1996 Nantes, France	1998 Ålesund, Norge
1998 Copenhagen, Denmark	2001 Dijon, Frankrig
2000 Columbia, MS, USA	2003 Boston, USA
2002 Dortmund, Germany	2005 Harrogate, UK
2004 Davis, CA, USA	2007 Minneapolis, USA
2006 Ås, Norway	2009 Florence, Italy
2008 Ontario, Canada	2011 Bangkok, Thailand
2010 Rotterdam, Holland	
2012 Rennes, France	

The Sensometric Society
www.sensometric.org

DTU Informatics
Department of Informatics and Mathematical Modelling
UNF Bornholm, 19 Mars 2010, © Per Bruun Brockhoff



Triangle test:

A subject is presented with 3 samples, two of which are equal:

● ● ●

Question: Which is different?

Binomial data: answer is correct or incorrect

DTU Informatics
Department of Informatics and Mathematical Modelling
UNF Bornholm, 19 Mars 2010, © Per Bruun Brockhoff

Binomial models:



- X = Number of correct answers

X follows a binomial distribution with some unknown p

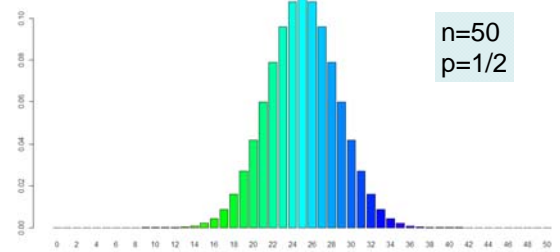
$$P(X = k) = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

”Plat-og-krone fordelingen”

Plat og krone p=1/2



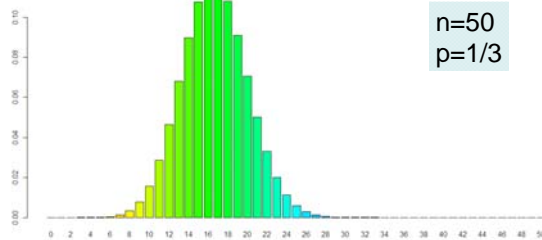
Plat og krone fordeling, p=1/2



Triangle test: (på ens produkter)



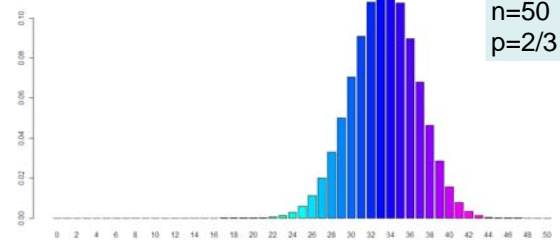
Forventet fordeling hvis rent gæsteri



Triangle test: (hvor halvdelen af alle folk erkender forskellen)



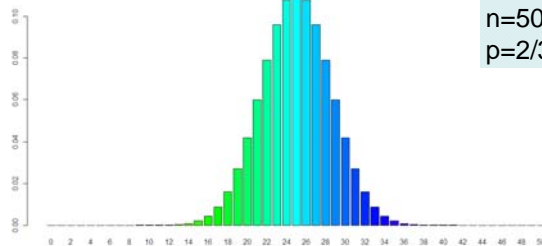
Forventet fordeling hvis 50% kan detektere forskellen



Triangle test: (hvor ¼ af alle folk erkender forskellen)



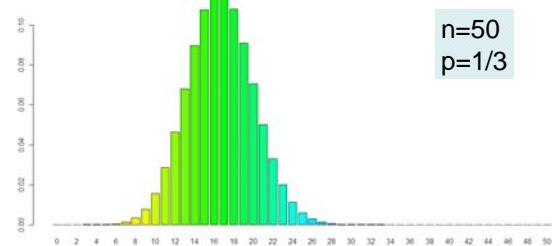
Forventet fordeling hvis 25% kan detektere forskellen



Triangle test: (på ens produkter)



Forventet fordeling hvis rent gæsteri



Hvad nu hvis??



Vi får 50 korrekte svar ud af 50?

Hvad er svaret på spørgsmålet:

1. Kan vi påvise en (detekterbar) forskel på øllene?
2. Kan I smage forskel på Carlsberg og Tuborg?

DTU Informatics

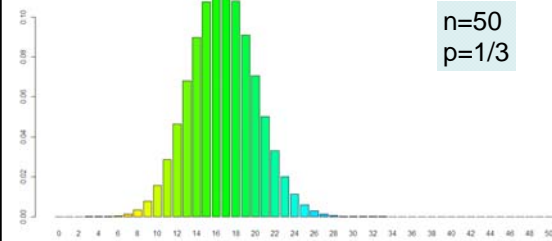
Department of Informatics and Mathematical Modelling

UNF Bornholm, 19 Marts 2010, © Per Bruun Brockhoff

Triangle test: (på ens produkter)



Forventet fordeling hvis rent gæteri



DTU Informatics

Department of Informatics and Mathematical Modelling

UNF Bornholm, 19 Marts 2010, © Per Bruun Brockhoff

Hvad nu hvis??



Vi får 17 korrekte svar ud af 50?

Hvad er svaret på spørgsmålet:

1. Kan vi påvise en (detekterbar) forskel på øllene?
2. Kan I smage forskel på Carlsberg og Tuborg?

DTU Informatics

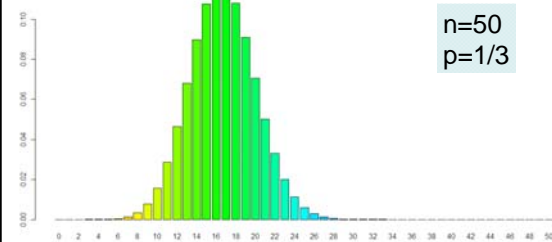
Department of Informatics and Mathematical Modelling

UNF Bornholm, 19 Marts 2010, © Per Bruun Brockhoff

Triangle test: (på ens produkter)



Forventet fordeling hvis rent gæteri



DTU Informatics

Department of Informatics and Mathematical Modelling

UNF Bornholm, 19 Marts 2010, © Per Bruun Brockhoff

Hvad nu hvis??



Vi får 20 korrekte svar ud af 50?

Hvad er svaret på spørgsmålet:

1. Kan vi påvise en (detekterbar) forskel på øllene?
2. Kan I smage forskel på Carlsberg og Tuborg?

DTU Informatics

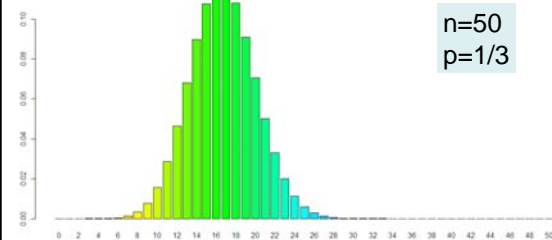
Department of Informatics and Mathematical Modelling

UNF Bornholm, 19 Marts 2010, © Per Bruun Brockhoff

Triangle test: (på ens produkter)



Forventet fordeling hvis rent gæteri



DTU Informatics

Department of Informatics and Mathematical Modelling

UNF Bornholm, 19 Marts 2010, © Per Bruun Brockhoff

Hvad nu hvis??



Vi får 26 korrekte svar ud af 50?

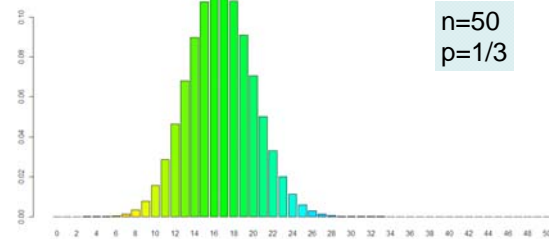
Hvad er svaret på spørgsmålet:

1. Kan vi påvise en (detekterbar) forskel på øllene?
2. Kan I smage forskel på Carlsberg og Tuborg?

Triangle test: (på ens produkter)



Forventet fordeling hvis rent gæteri



Hypotesetest



Hvad nu hvis: vi får x korrekte svar ud af n

Hvad er svaret på spørgsmålet:

1. Er der en (detekterbar) forskel på øllene?
2. Kan I smage forskel på Carlsberg og Tuborg?

$$H_0 : p = 1/3$$

$$H_1 : p > 1/3$$

Hypothesis test:

1. Find P-value and compare with the nominal level (alpha)
2. How odd does the data look? (IF the hypothesis is OK)

Hvad er statistik?



At tage intelligente beslutninger på basis af tal (empirisk information).

Type I and II risks



	Claiming Difference ("Rejection")	Claiming Similarity ("Acceptance")
Products are similar	TYPE 1 ERROR (alpha)	Correct decision
Products are different	Correct decision	TYPE 2 ERROR (beta)

Consequences of errors:



- Example: New and "better" recipe for an existing product is sought!
- Type I error: A perfect recipe is rejected (Production department is unhappy!)
- Type II error: A "wrong" product is accepted and sent on the market (Marketing department is VERY unhappy)

Confidence intervals



- Intervals are used instead of single numbers
- Good tool to summarize information
- Example:

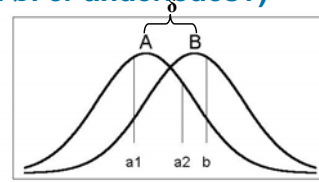
$$p \pm 1.96 \times \sqrt{p(1-p)/n}$$

DTU Informatics

Department of Informatics and Mathematical Modelling

UNF Bornholm, 19 Marts 2010, © Per Bruun Brockhoff

Triangle test: (Hvilken øl er anderledes?)



Sensory continuum

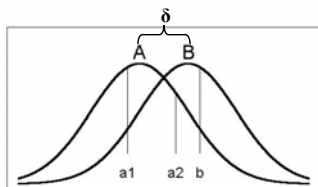
$$g(\delta) = 2 \int_0^{\infty} \left\{ \Phi \left[-z\sqrt{3} + \delta\sqrt{2/3} \right] + \Phi \left[-z\sqrt{3} - \delta\sqrt{2/3} \right] \right\} \frac{1}{\sqrt{2\pi}} \exp[-z^2/2] dz$$

DTU Informatics

Department of Informatics and Mathematical Modelling

UNF Bornholm, 19 Marts 2010, © Per Bruun Brockhoff

3-AFC test: Hvilken er mest bitter?



Sensory continuum

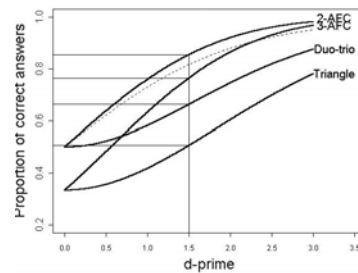
$$g(\delta) = \int_{-\infty}^{\infty} \phi(z - \delta) \Phi(z)^2 dz$$

DTU Informatics

Department of Informatics and Mathematical Modelling

UNF Bornholm, 19 Marts 2010, © Per Bruun Brockhoff

The “paradox” is resolved!



DTU Informatics

Department of Informatics and Mathematical Modelling

UNF Bornholm, 19 Marts 2010, © Per Bruun Brockhoff

Hvad er Statistik?



1. Historisk, samfundsmæssigt, naturvidenskabeligt
 2. Egentlig introduktion til statistik (link til matematik på gymnasiet)
- FORBAVSENDE kort vej fra GYM til Business/Videnskab!!**



Mennesker som måleinstrumenter

DTU Informatics

Department of Informatics and Mathematical Modelling

UNF Bornholm, 19 Marts 2010, © Per Bruun Brockhoff

Statistics DTU Profile



- **Examples:**
 - Climatology
 - Energy (Wind)
 - Pharmaceutics
 - Finance
 - Food science
 - Environmental research
 - Marine biology
 - Microbiology
- **Characteristics:**
 - Working with the use and development of statistics in several areas
 - **KORT vej fra kursusviden til forskning og/eller business**



DTU Informatics

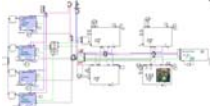
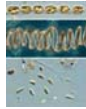
Department of Informatics and Mathematical Modelling

UNF Bornholm, 19 Marts 2010, © Per Bruun Brockhoff

Bachelor i Matematik og Teknologi:



http://www.dtu.dk/Uddannelse/Civilingenioer/Bacheloruddannelsen/Matematik_og_Teknologi.aspx



DTU Informatics
Department of Informatics and Mathematical Modelling

UNF Bornholm, 19 Mars 2010, © Per Bruun Brockhoff