

0: EM1 / Syys 2008 / F

1. TILASTOTIETEEN ROOLI
UUDEN TIETEELLISEN
TIEDON LUOMISESSA

EMPIIRISEN TUTKIMUSPROSESSIN VAIHEET

1. SUUNNITTELUVAIHE
2. AINEISTON KERUU
3. AINEISTON TARKISTUS JA KOODAUS
4. ANALYYSI
5. TULOSTEN RAPORTOINTI

LÄHDE

ALKULA, PÖNTTINEN & YLÖSTALO:

SOSIAALITUTKIMUKSEN KVANTITATIIVISET
MENETELMÄT

HUOM. Aloita ongelman ratkaisu aina

yksinkertaisia tilastollisia menetelmiä käyttäen.

ESIM. Graafiset havainnollistukset.

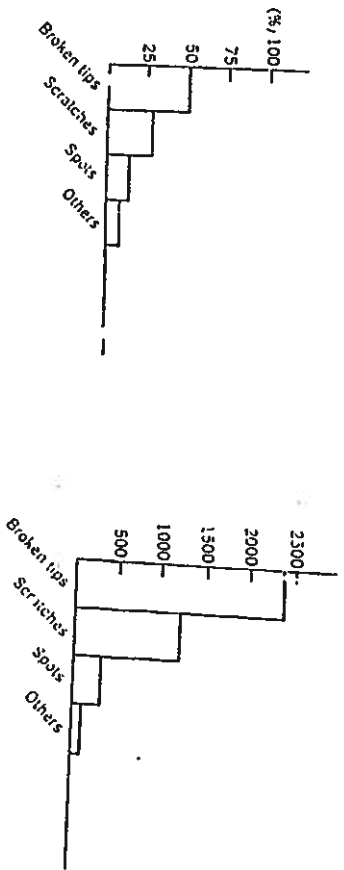
ESIM. Pareto diagramma

"The principle of vital few, trivial many."

Parannettu versio tukkimiehen kirjainpidosta !!!

Luokiteltu vihetekijän

- (a) suhteellisen esiintymisfrekvenssin tai
- (b) kustannusvaikutuksen suuruuden mukaan.



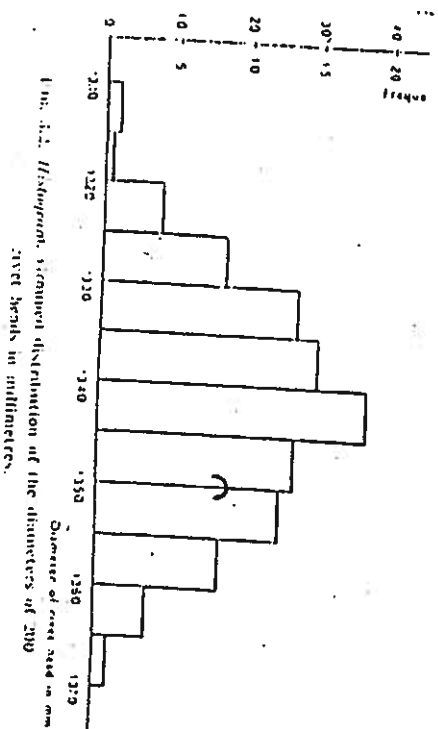
LÄHDE:
 Ryan, T.P.: Statistical Methods for Quality Improvement, Wiley (1989).

AVIEMI / STEIN 2008 / 22

Primary data.

Diameters of 200 rivet heads in millimetres

13.39	13.43	13.54	13.64	13.40	13.55	13.40	13.26
13.42	13.50	13.32	13.31	13.28	13.52	13.46	13.63
13.36	13.44	13.52	13.53	13.37	13.33	13.24	13.13
13.53	13.59	13.57	13.51	13.34	13.39	13.47	13.17
13.51	13.48	13.62	13.58	13.57	13.53	13.51	13.40
13.30	13.48	13.40	13.57	13.51	13.40	13.52	13.55
13.40	13.34	13.23	13.37	13.48	13.48	13.48	13.35
13.40	13.36	13.45	13.48	13.29	13.58	13.44	13.56
13.28	13.59	13.47	13.46	13.82	13.54	13.20	13.38
13.43	13.35	13.56	13.51	13.47	13.40	13.29	13.20
13.46	13.44	13.42	13.29	13.41	13.39	13.50	13.48
13.53	13.34	13.45	13.42	13.29	13.32	13.45	13.50
13.55	13.33	13.32	13.45	13.29	13.32	13.32	13.48
13.29	13.25	13.44	13.60	13.46	13.32	13.48	13.58
13.24	13.28	13.58	13.31	13.31	13.51	13.43	13.38
13.34	13.49	13.50	13.38	13.48	13.43	13.43	13.44
13.54	13.33	13.36	13.46	13.25	13.43	13.43	13.29
13.46	13.26	13.40	13.52	13.59	13.48	13.36	13.27
13.43	13.36	13.50	13.38	13.43	13.34	13.41	13.40
13.42	13.55	13.37	13.41	13.38	13.14	13.42	13.24
13.38	13.54	13.30	13.18	13.32	13.46	13.39	13.52
13.37	13.37	13.50	13.61	13.42	13.32	13.35	13.35
13.57	13.31	13.48	13.36	13.28	13.58	13.58	13.40
13.26	13.37	13.28	13.39	13.32	13.20	13.43	13.38
13.33	13.33	13.31	13.45	13.39	13.45	13.41	13.45



2012/1/5/14/1/2

SOSIAALITUTKIMUKSEN KAKSI SUUNTAUSTA:

KVALITATIIVINEN

ESIM.: Elämänkerta-analyysi, tyyppiä

J.P. Roos: "Suomalainen mies"

KVANTITATIIVINEN

"Tilastollinen lähestymistapa"

Suuntauksel / lähestymistavat eivät ole toisensa poissulkevia, vaan parhaimmillaan toisiaan täydentäviä.

Alkula, Pöntinen, Ylöstalo: Sosiaalitutkimuksen
kvantitatiiviset menetelmät, WSOY (1994)

"Jos kvantitatiivinen mitaus pystyy antamaan tutkittavasta ilmiöstä tutkimusongelman kannalta relevanttia tietoa, voidaan analyseissä käyttää tilastollisia menetelmiä, jolloin aukeaa mahdollisuus sattuman ja systemaattisen vaihtelun erottamiseen johtopäätösten tueksi. Määrällisistä asioista saadaan tarkkaa tietoa, ja voidaan kuvata asioiden suuruusluokkia ja eri asioiden välisiä riippuvuuksia. Samoin on mahdollisuus vertailla ilmiötä eri tilanteissa tai yhteisöissä tai seurata sen kehitystä ajallisesti, vaikka kvantitatiiviset mittaukset eivät sellaisenaan takaa vertailukelpoisuutta. Pidemmälle viedyssä menetelmässä voidaan soveltaa tilastollisia malleja, niin että empiirisen aineiston (ja siis tutkimuskohteen) rakennetta ja toimintatapaa kuvataan mallia koskevien parametrien estimaattien avulla. ..."

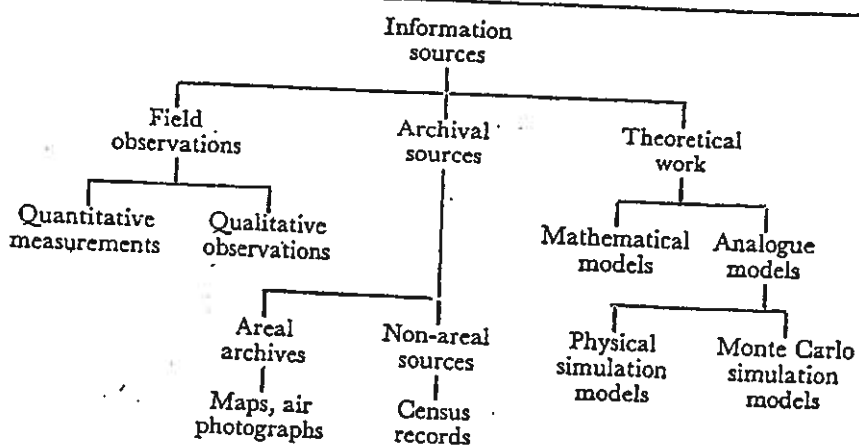
THE PROCESS OF QUANTITATIVE RESEARCH

1. THEORY
2. HYPOTHESIS
3. RESEARCH DESIGN
4. DEVISE MEASURES OF CONCEPTS
5. SELECT RESEARCH SITE(S)
6. SELECT RESEARCH SUBJECTS / RESPONDENTS
7. ADMINISTER RESEARCH INSTRUMENTS / COLLECT DATA
8. PROCESS DATA
9. ANALYSE DATA
10. FINDINGS / CONCLUSIONS
11. WRITE UP FINDINGS / CONCLUSIONS

LÄHDE:

ALAN BRYMAN: SOCIAL RESEARCH METHODS (2008)

Table 7.1. Sources of locational information in human geography



Source: Adapted from Chorley. In Dury, 1965, p. 276.

Niemi / Sources 2008 / 3

TILASTOTIEDE, MITÄ SE ON ???

Jokaisella tieteenalalla, kuten

- matematiikka
 - biologia
- on tunnusomaiset

- tieteelliset ongelmat, sekä
- tieteelliset menetelmät näiden ratkaisemiseksi.

Entä tilastotiede - mikä tekee siitä itsenäisen tieteenalan ???

Onko olemassa

puhtaasti tilastotieteellisiä (tieteen sisäisiä) ongelmia, joita tilastotiede ("tieteenä") pyrkii ratkaisemaan

vai onko tilastotiede vain

kokoelma menetelmiä / kaavoja muiden tieteenalojen ongelmien ratkaisemista varten ???

Tai päteekö Mark Twainin karakterisaatio:

"Lies, damned lies and statistics" ???

21.02.2005 / 2005 / 4

TAI VAIKKAPA NÄIN:

TILASTOTIEDE

= HAVAINTOAINESTOIHIN PERUSTUVA

ONGELMANRATKAISUMENETELMÄ

KUITENKIN:

Myös teoreettinen viitekehikko on tarpeen.

Tilastotieteen (tieteenä !!!) sisäiset ongelmat

= metodinen kehitys (erityisesti uusilla

sovellusaloilla, esim. biostatistiikka).

Tilastotieteen menetelmät tieteenalan sisäisten ongelmien ratkaisun:

Todennäköisyyslaskenta !!!

C. R. Rao (1989):

"Is statistics, as studied and practiced today, a science, technology or art? Perhaps it is a combination of all these."

YLEISESTI OTTAEN

"... statistics is the study of how information should be employed to reflect on, and give guidance for action in practical situation involving uncertainty."

Vic Barnett, Brittilastotieteilijä.

"Data Analysis

= Answering Specific Questions
+ Providing Information for
New Lines of Research"

C. R. Rao (1989).

1834 Royal Statistical Society (Englanti)

perustettiin

Tällöin esitetty karakterisaatio:
statistics

= facts relating to men, which are capable of being expressed in numbers, which sufficiently multiplied, to indicate general laws.

Lopuksi poiminto modernista kirjallisuudesta:

Niiniluoto, I. (1983):

"Tieteellinen päättely ja selittäminen"

"Tilastotieteessä, jonka tehtävänä on empiirisessä tiedonmuodostuksessa esiintyvän epävarmuuden systemaattinen käsittely ... "

Review 2008/5

EMPIIRISEN TUTKIMUKSEN PERUSASETELMA

VASTAKKAINASETTELU:

TEORIA <-----> EVIDENSSI

"A Beautiful theory, killed by a nasty, ugly little fact."

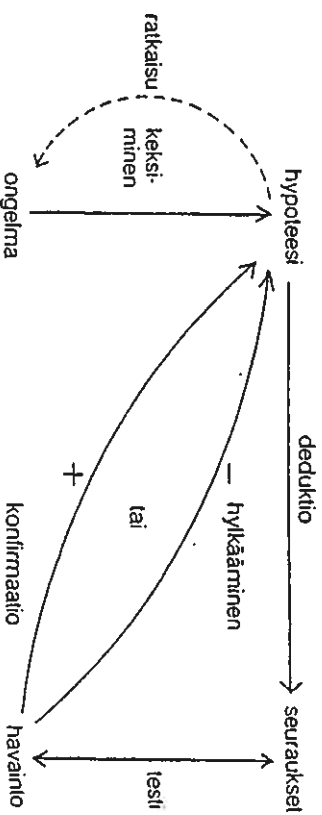
Thomas H. Huxley (1825 - 1895), Brittibioogi

"Scientific laws are not advanced by the principle of authority or justified by faith or medieval philosophy, statistics is the only court of appeal to new knowledge"

P. C. Mahalanobis (1893 - 1972)

Maailmankuulu intialainen tilastotieteilijä

Indian Statistical Institute'n perustaja



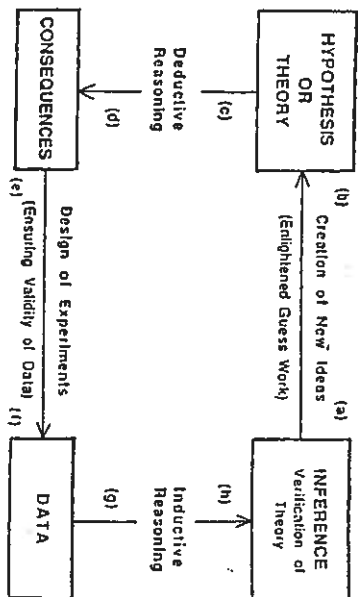
Kuva 10. Hypoteettis-deduktiivinen menetelmä.

LÄHDE: NIINILUOTO

Tieteellinen päättely

ja selittäminen (Otava, 1983)

NIINILUOTO / 5.11.2008 / 6



Every hypothesis is possibly rejected with the accumulation of more data, a situation bluntly described by Karl Popper :

"Supporting evidence for a scientific hypothesis is merely an attempt at falsification which failed."

The scientific method as shown in the above diagram involves two logical processes—deductive reasoning and inductive reasoning*. Deductive reasoning is deriving consequences from given premises, a thought process perfected by the Greeks in the third century B.C. and systematized through the study of mathematics over the last 2000 years. Inductive reasoning, which involves arguing from insufficient premises and taking decisions, must have been practised from the very beginning of mankind. However, it remained uncodified, perhaps due to uncertain nature of inference and the natural inhibition of the human mind to prescribe rules which might lead to wrong results. The general philosophy was : *Do not count the chickens before they are hatched*. The major breakthrough in our thinking came in the beginning of the present century when it was realized that inductive inference, though uncertain, can be made precise by specifying the amount of uncertainty in the conclusions drawn. The first fruits of this thinking were the development of tests of significance by which hypotheses could be judged (accepted or rejected) subject to a specified amount of uncertainty.

Lähde:

C.R. Rao *Statistics and Truth* (1989)

NIEMI / 5.12.2005 / 2

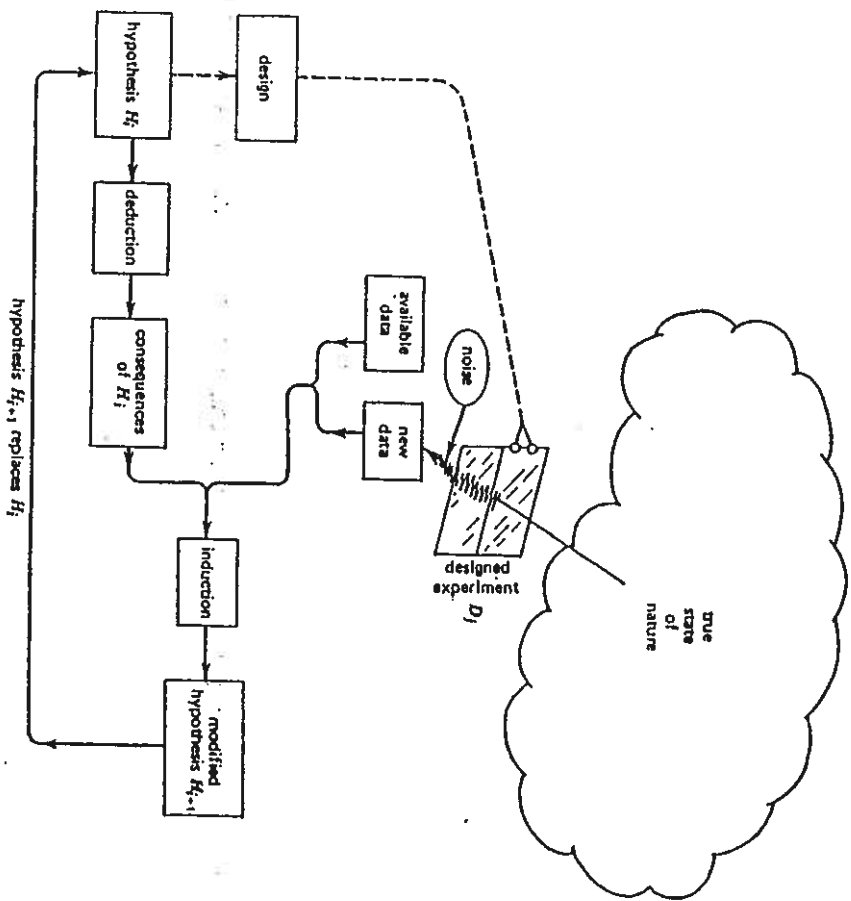


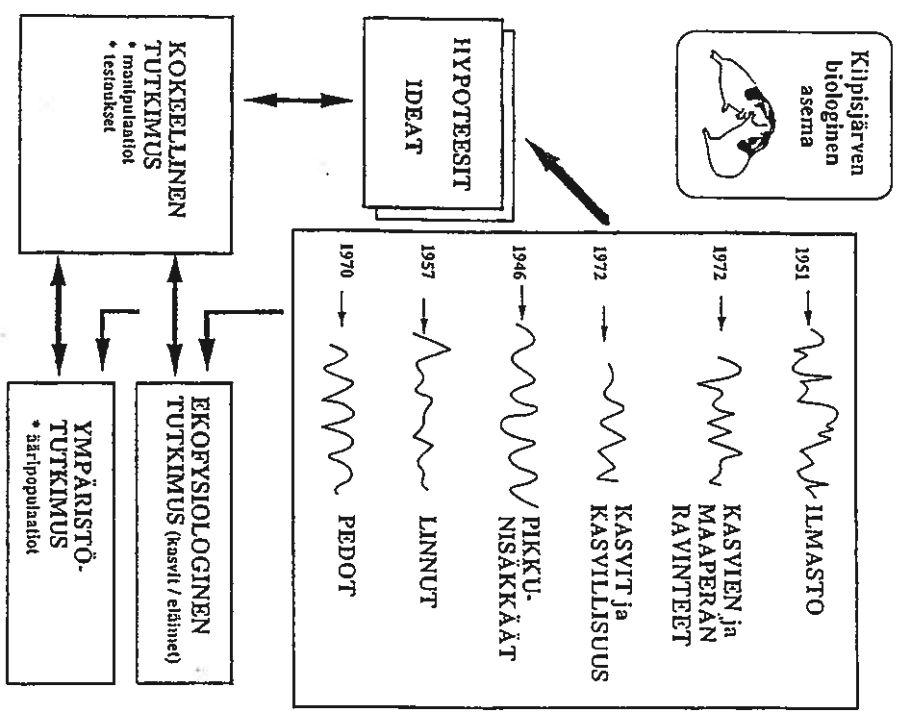
FIGURE 13. Data generation and data analysis in scientific investigation.

LÄHDE: BOX, G.E.P., HUNTER, W.G & HUNTER, J.S.
Statistics for experimenters, Wiley (1978)

G.1

21 EMU / 15.12.2008 / S

PITKÄAIKAISET SEURANTATUTKIMUKSET



TIILASTOTIEDE =

HAVAINTOAINEISTOIHIN
 PERUSTUVA ONGELMAN-
 RATKAISUMENETELMÄ

LÄHESTYMISSAVAT

(A) HAVAINNOIVA TIEDONKERUU

- tilastomatemattisen mallin
 verifiointi / konstruointi

Esim. Keplerin lait

(B) KOKEELLINEN TIEDONKERUU

- aktiivinen vaikuttaminen havaintojentekoon
 tai koeolosuhteisiin mahdollista

Esim. Optimaalisten säätöjen määrittäminen

(C) OTANTA / NÄYTTIENOTTO

- vain pieni osa kokonaisuainestosta tutkitaan
 yksityiskohteisesti

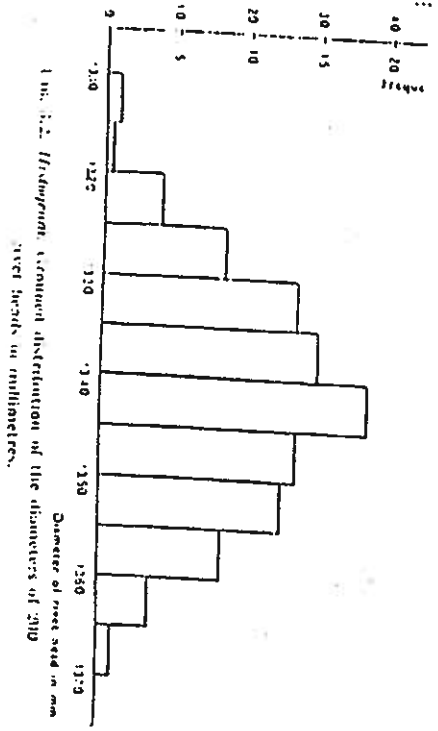
Esim. Tilastuorantoro / Tilastokeskus

51.1

NIEMI / Jyväskylä 2008 / 9

Primary data.

Diameters of 200 rivet heads in millimetres											
13.39	13.43	13.54	13.64	13.40	13.55	13.40	13.26				
13.42	13.50	13.32	13.31	13.28	13.52	13.46	13.63				
13.38	13.44	13.52	13.53	13.37	13.37	13.24	13.13				
13.53	13.53	13.39	13.57	13.51	13.34	13.39	13.47				
13.51	13.48	13.62	13.58	13.57	13.33	13.51	13.40				
13.30	13.48	13.40	13.57	13.51	13.40	13.52	13.56				
13.40	13.34	13.23	13.37	13.48	13.48	13.62	13.35				
13.10	13.36	13.45	13.46	13.29	13.58	13.44	13.56				
13.28	13.59	13.47	13.46	13.62	13.54	13.20	13.38				
13.43	13.35	13.56	13.51	13.47	13.40	13.29	13.20				
13.46	13.44	13.42	13.29	13.41	13.39	13.50	13.48				
13.53	13.34	13.45	13.42	13.29	13.38	13.45	13.50				
13.53	13.33	13.32	13.69	13.46	13.32	13.32	13.48				
13.39	13.35	13.44	13.60	13.43	13.51	13.38	13.38				
13.24	13.28	13.58	13.31	13.31	13.45	13.43	13.44				
13.34	13.49	13.40	13.38	13.46	13.43	13.37	13.29				
13.34	13.33	13.36	13.46	13.33	13.44	13.38	13.27				
13.66	13.26	13.40	13.52	13.59	13.48	13.46	13.40				
13.43	13.26	13.50	13.38	13.43	13.34	13.41	13.24				
13.42	13.55	13.37	13.41	13.38	13.14	13.42	13.52				
13.38	13.54	13.30	13.18	13.42	13.46	13.35	13.40				
13.34	13.37	13.50	13.61	13.42	13.37	13.35	13.35				
13.57	13.31	13.40	13.36	13.28	13.58	13.58	13.36				
13.36	13.37	13.38	13.39	13.32	13.30	13.43	13.34				
13.33	13.33	13.31	13.45	13.39	13.43	13.41	13.45				



TIILASTOLLISELLA ANALYYSILLÄ HALUTAAN
SELVITTÄÄ SYY - SEURAUSSUHTEITA

VAROITUS:

Tilastollisten ohjelmistojen "mekanismien" käyttö saattaa johtaa virhetulkitoihin. Tulosten tulkinta tehtävä substanssialueen asiantuntijoiden kanssa.

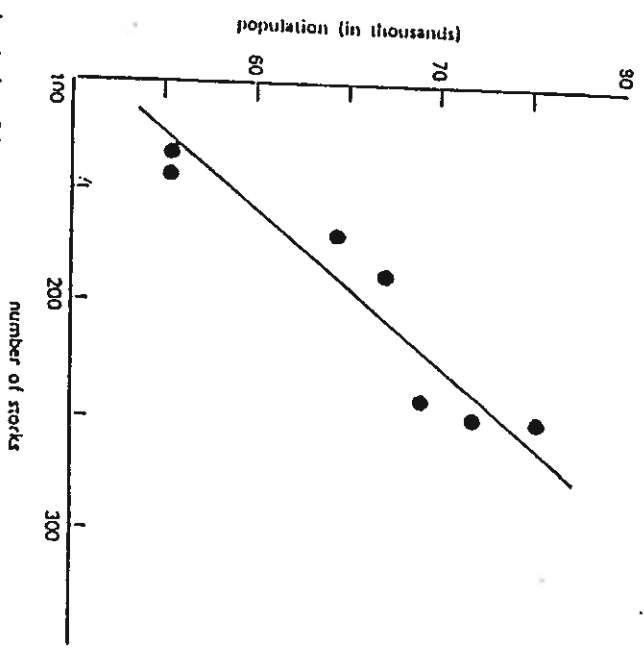


FIGURE 1.4. A plot of the population of Oldenburg at the end of each year against the number of storks observed in that year, 1930-1935.

Confusion of Correlation with Causation

LÄHDE: Box, G.E.P., Hunter, W.G. & Hunter, J.S.:
Statistics for Experimenters, Wiley (1978).