

NGF 2008

Ekstreme hendelser
sett med en statistikers øyne

Magne Aarset

Handelshøyskolen BI

Høyskolen i Ålesund

FN's klimapanel

IPCC

The Intergovernmental Panel on Climate Change

Forecasts by Scientists vs Scientific Forecasts

- Beregning av prognoser som fagområde
- Ekstremal-prosesser
- Beslutningsfeller

Granli-utvalget (1976)

Utvalget skulle være «bredt sammensatt» av både «eksperter og samfunnsengasjerte ikke-eksperter», ifølge Stortinget.

Hva dette egentlig skulle innebære, ble det i praksis opp til Industridepartementet å bestemme.

Det ble et «ja» til kjernekraft, men et «nei» til kjernekraft på 1980-tallet.

NVE anbefalte 12 kjernekraftverk.

Prognosemodellering

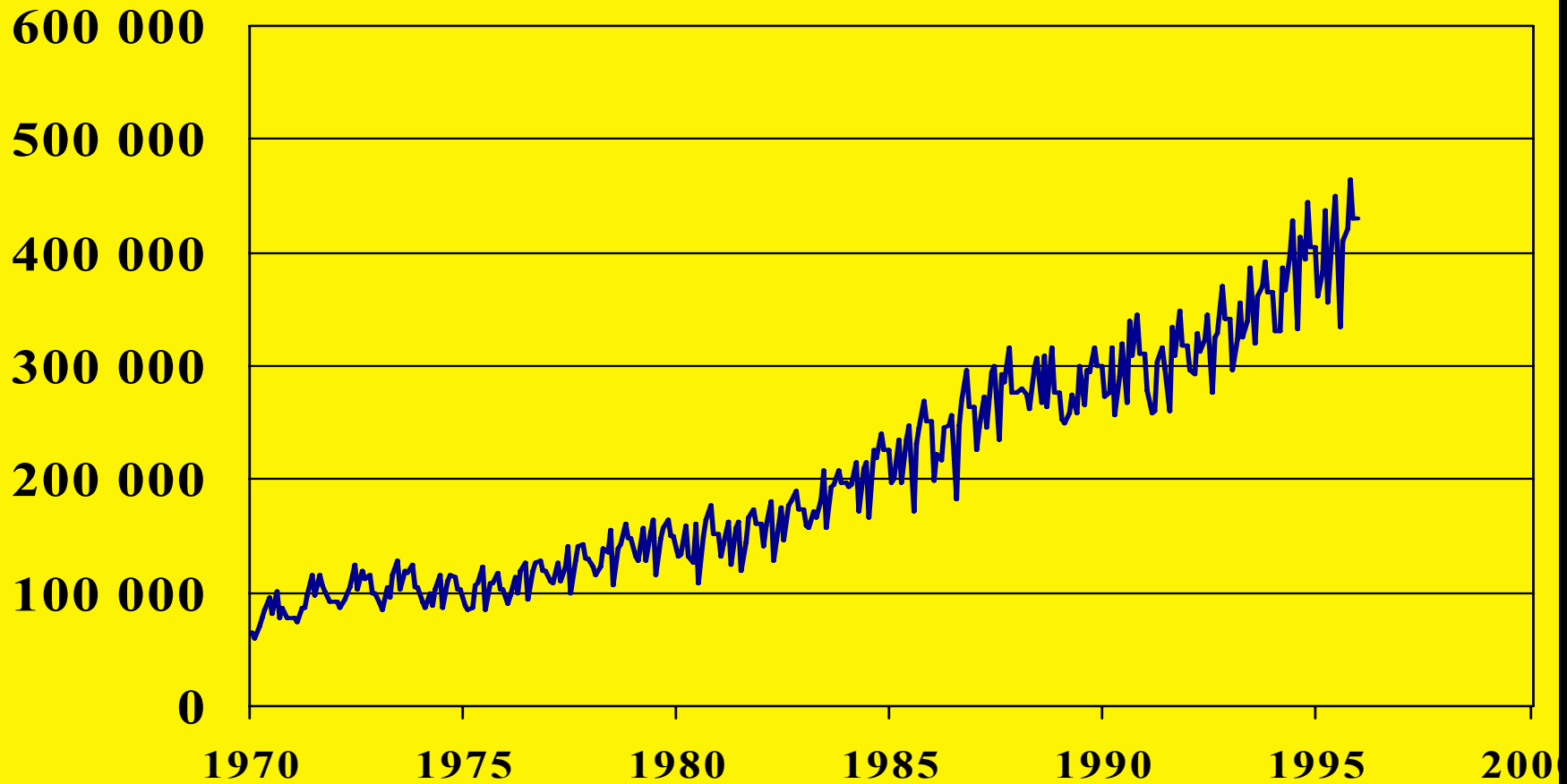
Forskjellige fremgangsmåter

- K.E.Trenberth: Klimapanelet har 20.000 - 30.000 variable i sine modeller.
- CO₂ inkludert i 1997
- Påvirkning frå rommet?
- Usikkerhet knyttet til prognosene.
- Statistikere går (vanligvis) ut fra 140 prognoseprinsipper. 89 kunne testes i IPCC WP1. 71 var ikke fulgt!

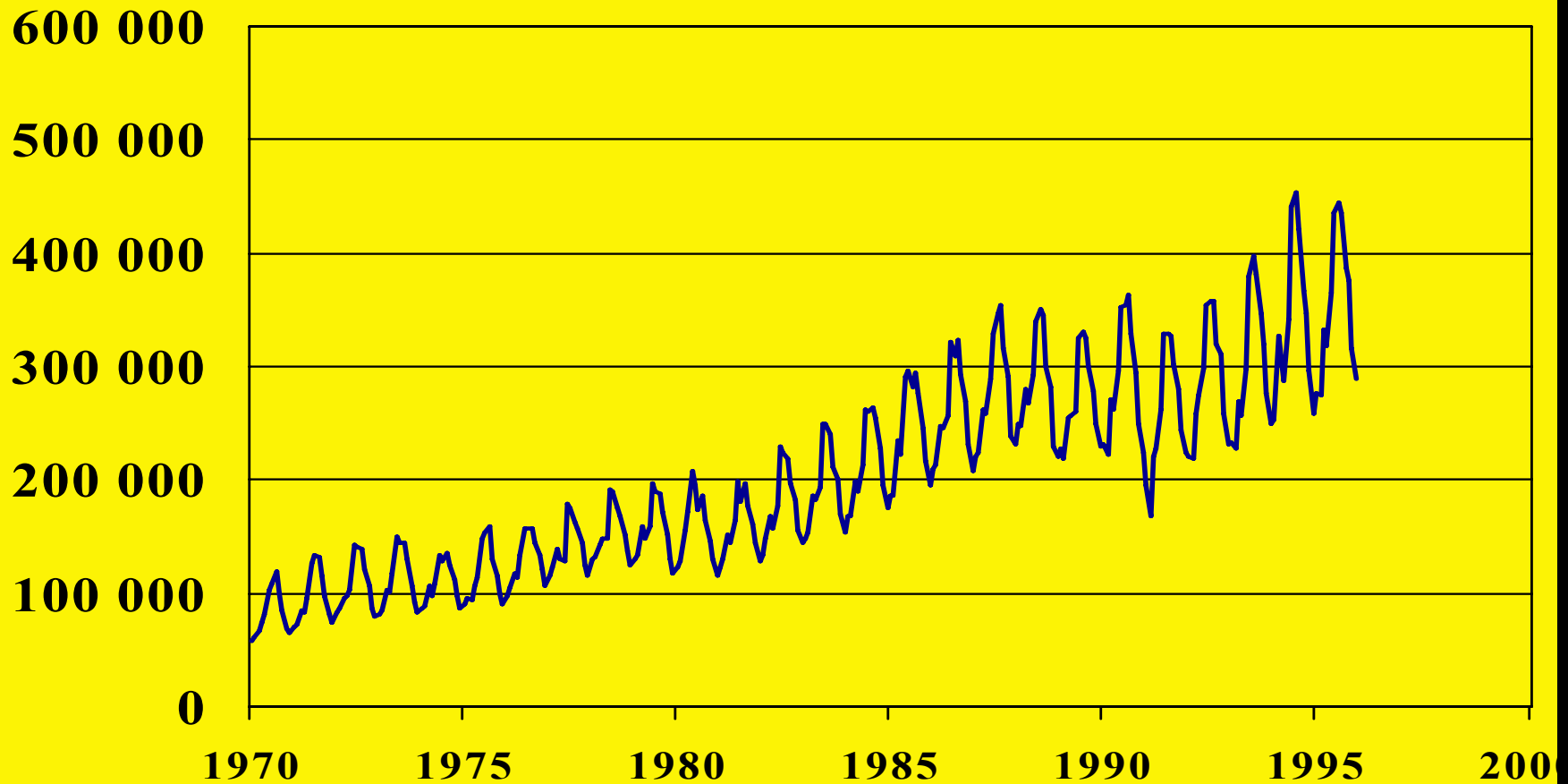
K.E. Trenberth

- In fact there are no predictions at all. And there never has been. The IPCC instead offers "what if" projections of future climate that correspond to certain emissions scenarios.
- ... None of the models used by the IPCC are initialized to the observed state and none of the climate states in the models correspond even remotely to the current observed climate. In particular, the state of the oceans, sea ice, and soil moisture has no relationship to the observed state at any recent time in any of the IPCC models.

Domestic passengers

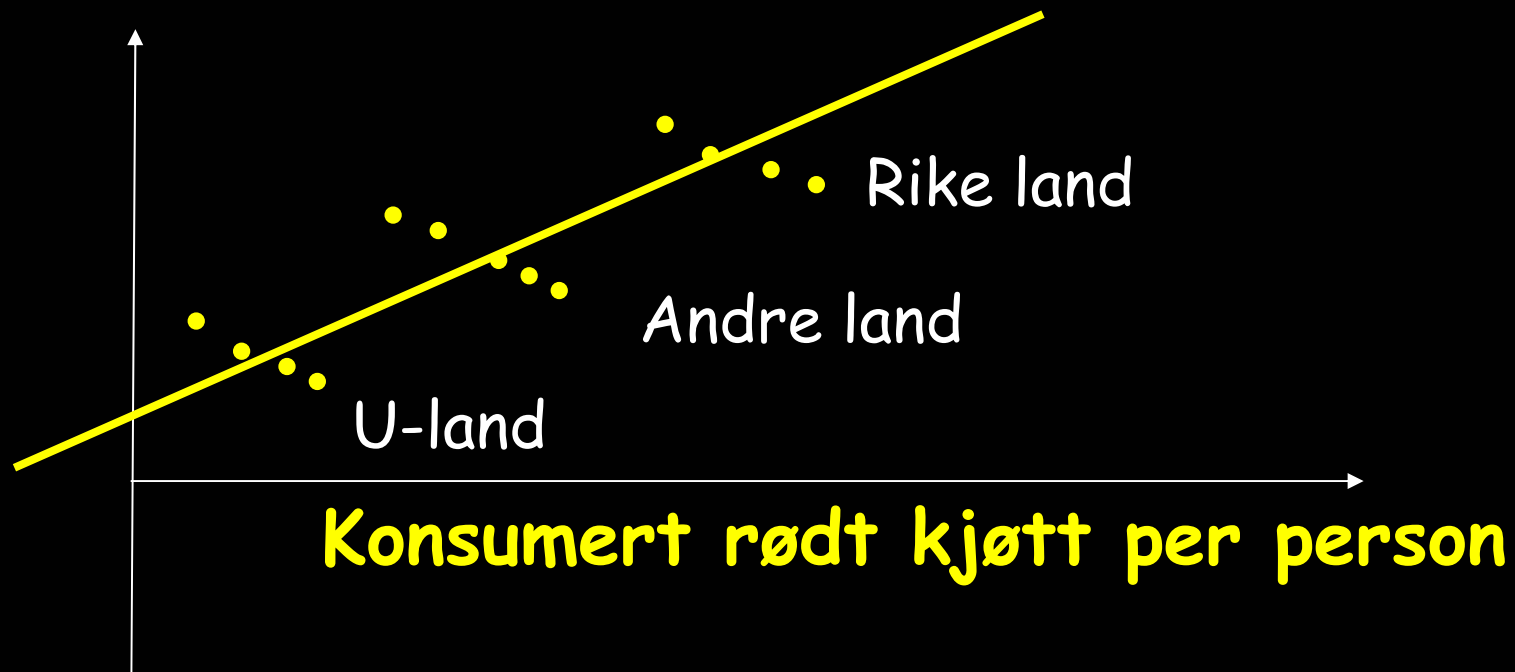


International passengers



$$\text{Levealder} = \beta_0 + \beta_1 \text{Kjøtt} + \varepsilon$$

Forventet levealder



Eksempel

Y = Demand for brasilian coffee

p_{bc} = Price on brasilian coffee

p_t = Price on tea

y_d = Mean disposable income

$$\hat{Y} = 9.1 + 7.8 p_{bc} + 2.4 p_t + 0.0035 y_d$$

Eksempel

Y = Demand for brasilian coffee

p_{bc} = Price on brasilian coffee

p_t = Price on tea

y_d = Mean disposable income

p_{cc} = Price on Columbian coffee

$$\hat{Y} = 9.1 + 7.8 p_{bc} + 2.4 p_t + 0.0035 y_d$$

$$\hat{Y} = 10.0 + 8.0 p_{cc} - 5.6 p_{bc} + 2.6 p_t + 0.003 y_d$$

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

Outlier



Leverage point



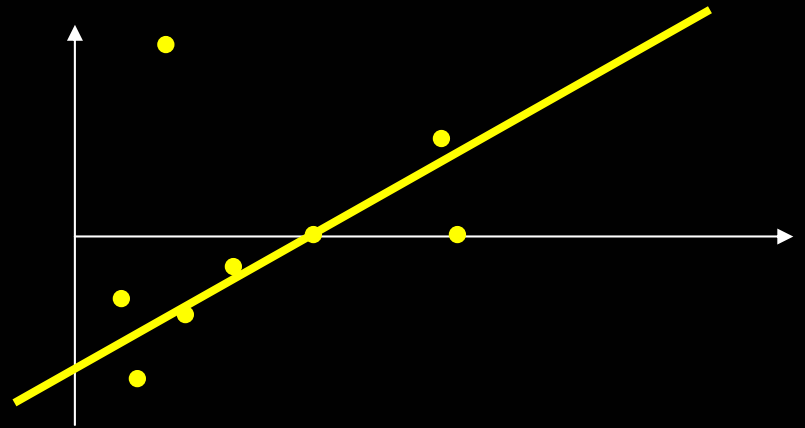
Estimeringsteknikker

$$Y = \beta_0 + \beta_1 x + \varepsilon$$

Finn det par $(\hat{\beta}_0, \hat{\beta}_1)$ som minimerer

$$\rho(\mathbf{e}_i) = \sum \omega_i |e_i|$$

- LAD: $\omega_i = 1$
- MKM: $\omega_i = |e_i|$
- ...



Minste kvadratavvik

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

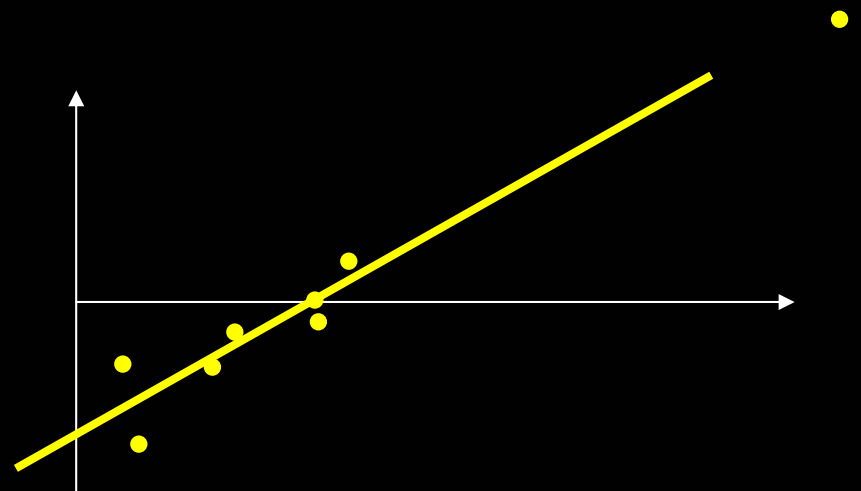
$$\hat{\beta}_0 = \bar{y} - \hat{\beta}\bar{x}$$

Minste kvadratavvik

$$\hat{\beta}_1 = \sum_{i=1}^n w_i \left(\frac{y_i - \bar{y}}{x_i - \bar{x}} \right)$$

der

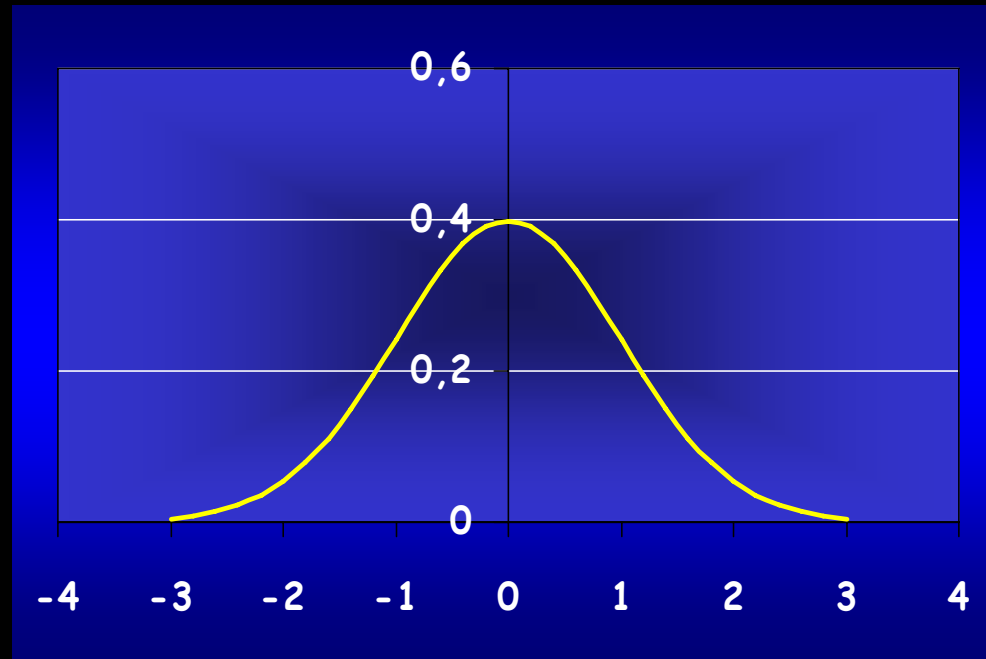
$$w_i = \frac{(x_i - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$



Estimeringsmetoder

- Least Absolute Deviation Regression (Boscovitch, 1757)
- Least Squares Regression (Gauss, 1795 & Legendre, 1805)
- Bayesian Regression (Bayes, 1760)
- Sufficiency Principle (Fisher, 1922)
- Invariance Principle (Hotelling, 1931 & Pitman, 1938)
- Non-parametric Regression (Jeffreys, 1939)
- Conditional Principle (Fisher, 1956)
- Maximum Likelihood Principle (Fisher, 1956)
- Ridge Regression (Hoerl, 1962)
- M-, L- og R-Regression (Huber, 1964 & 1981)
- Principal Components Regression (??)

Standard avvik i $N(\mu, \sigma^2)$



- 1 SD → 68,3 %
- 2 SD → 95,4 %
- 3 SD → 99,7 %

The subexponential distributions

$$\lim_{x \rightarrow \infty} \frac{P\{X_1 + X_2 + \dots + X_n > x\}}{P\{\max(X_1, X_2, \dots, X_n) > x\}} = 1$$

Max(X_1, \dots, X_n)

Gumbel: $F(x) = P(X \leq x) = e^{-e^{-x}}; \forall x$

Fréchet: $F(x) = P(X \leq x) = e^{-x^{-\alpha}}; x \geq 0$

Weibull: $F(x) = P(X \leq x) = e^{-(-x)^{-\alpha}}; x \leq 0$

Ekstremalprosesser

X_1, \dots, X_n iid.

A record X_n occurs if $X_n > \max(X_1, \dots, X_{n-1})$

X_1 is by definition a record.

The record counting process

$$N_1 = 1$$

$$N_n = 1 + \sum_{i=2}^n I\{X_n > \max(X_1, \dots, X_{n-1})\}$$

Then

$$EN_n = \sum_{i=1}^n \frac{1}{i}$$

$$\text{Var}N_n = \sum_{i=1}^n \left(\frac{1}{i} - \frac{1}{i^2}\right)$$

The record counting process

$n=10^k$									
$k=$	1	2	3	4	5	6	7	8	9
EN_n	2.9	5.2	7.5	9.8	12.1	14.4	16.7	19.0	21.3

The number of records of iid data grows very slowly!

Probability distributions

- The subexponential distributions as realistic models for heavy-tailed random variables.
- The α -stable distributions for $\alpha < 2$ as the limit laws for sums of infinite-variance iid random variables.
- The Fréchet, the Weibull, and the Gumbel distributions, as limit laws for maxima of iid random variables.
- The Normal distribution as limit law for sums of iid, finite variance random variables.

Stochastic processes

- Poisson process
- α -stable processes ($0 < \alpha < 2$) and Brownian motion
- More general processes using the above as input.

Bruk av informasjon Wason (1966)

2 4 6

Hvilken regel ligger bak denne rekken med tall?

Prøv en ny rekke med tre tall.

Jeg skal fortelle om den er riktig eller ikke.

$a < b < c$

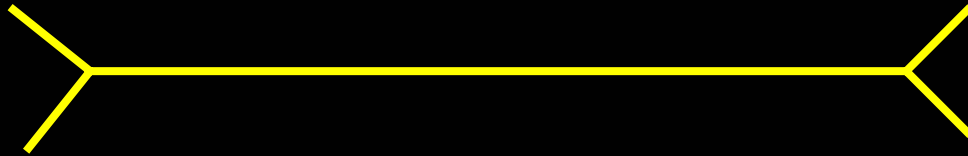
- Bare 25 % klarte oppgaven.
- Man lette etter bevis for å bekrefte den regel de opprinnelig trodde var riktig.
- De prøvde ikke eksempler som ville motbevise denne regelen.
- Etter at de fikk vite at foreslått regel var gal, prøvde 50 % fremdeles å få bekreftet at deres foreslåtte regel var riktig.

Bruk av informasjon

Vi kan bruke tilgjengelig informasjon feil, og vi kan la være å bruke tilgjengelig informasjon.

- Vi lager oss kun et lite antall mulige diagnoser, gjerne basert på kun en begrenset del av den tilgjengelige informasjon.
- Vi avviser/ignorerer informasjon som peker i en annen retning.
- Vi velger ut informasjon som bekrefter de antagelser vi har.
- Vi har en selektiv og forsterkende oppfatning.

Müller - Lyer figurene



Forankring

Svar i løpet av 5 sekunder.

Hva er $8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1$?

Median = 2 250

Hva er $1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6 \cdot 7 \cdot 8$?

Median = 512

Korrekt svar = 40 320

Forankring

Hvor stor andel av landene i FN er afrikanske?

Gruppe 1 Median = 25%

10% ble "tilfeldig valgt" som utgangspunkt.

Gruppe 2 Median = 45%.

65% ble "tilfeldig valgt" som utgangspunkt.