

## Statistik og forbandet løgn

Af Ib Marcussen og Troels V. Østergaard

Der er løgn, forbandet løgn og statistik. Den gamle vittighed handler om den politiske brug af statistiske undersøgelser. Når man spørger folk, kan man som bekendt få de svar, man gerne vil have.

Men der er også en anden form for statistik, nemlig den der hjælper forskere til at undgå forkerte konklusioner, og det der er værre. I den medicinske forskning accepterer man kun resultater, hvis de er statistisk signifikante, og forsøgene yderligere er gennemført som dobbelt blinde. Det sidste betyder, at hverken forsøgspersonerne eller forskeren ved, hvem der har fået placebo. Grunden til, at forsøgene skal være dobbelt blinde, er, at al erfaring viser, at man – selv som forsker – meget let kan blive vildledt af sine forestillinger.

### Statistikløs værktøjskasse

Når vi taler geologi, er der ikke en forsøgspersons psyke at tage hensyn til, men som regel er det samme person, der laver "målinger" og tolker resultaterne. I den situation må det være ekstra vigtigt, at man sikrer, at ens resultater er statistisk signifikante.

Nu er det ikke så mange steder i geologien, hvor der er behov for eller mulighed for at gennemføre en statistisk analyse. Det er formentlig forklaringen på, at kun meget få geologer har statistik i deres "værktøjskasse". Det er ærgerligt – især hvis det leder til, at man mistolker sine resultater.



En tur på stranden med øje for stenene viser med stor tydelighed, at stenmateriale i de danske istidsaflejringer er heterogent. Med statistiske metoder kan man vise, at det er så heterogent, at ethvert forsøg på at bruge stentællinger som stratigrafisk redskab er dømt til at mislykkes. (Foto: TVØ)

### Tolkning af gletscherudbredelse

Et eksempel på et område, hvor brugen af statistik havde været på sin plads, er i tolkningen af, hvordan gletschere har bredt sig over Danmark i løbet af Weichsel. Som bekendt har man først brugt ledeblokke siden stentællinger og i de nyeste undersøgelser fingrusanalyser – tællinger af bjergartstyper

i fingrusfraktionen – til korrelation. Man har i princippet brugt resultaterne på samme måde, som man bruger ledefossiler, men uden at sikre, at analyserne er statistisk signifikante.

Da tællingerne – ledeblokke, sten, fingrus – er det eneste grundlag for forestillingerne om flere isstrømme fra forskellige retninger:

## Vi borer over hele landet..!

- Kerneboringer
- Hulsneglsboringer
- Højslevboring
- Tørboring
- Luftslylleboring
- Skylléboring

- ring og hør nærmere...



POUL CHRISTIANSEN A/S  
Brøndborer- &  
Ingeniørfirma  
7840 Højslev  
Tlf. 97 53 52 22

- din sikkerhed for erfaring og kompetence...

Norsk is, NØ-is, Baltisk osv. er det en bemærkelsesværdig undladelse. Og hvad værre er, går man analyserne efter, viser det sig, at de ikke er statistisk signifikante.

En gennemgang af fingrusanalyser publiceret af Houmark-Nielsen (1987) viser for eksempel følgende (Marcussen et al. 2006): Af 51 sæt multiple prøver (prøver fra samme morænebænk på én lokalitet) var kun 2 statistisk set identiske. For de øvrige 49 sæt var sandsynligheden for, at prøverne i hvert sæt stammede fra samme population, mindre end 5 %.

Sammenligner man prøver fra morænebænke, der er tolket som stammende fra forskellige isfremstød, og som findes på samme lokalitet, bliver resultatet ikke meget

anderledes: Af 94 sæt var sandsynligheden for, at prøverne skulle stamme fra samme population, mindre end 5 % i de 92. Det er forventeligt, men siger ikke meget, da multiple prøver fra samme moræne er lige så heterogene.

#### Et beregningseksempel

Beregningerne er for så vidt enkle. Man sætter tælle tallene op i en tilfældighedstabel og bruger  $\chi^2$  som testfordeling, som vist nedenfor. (Se for eksempel Krumbein, W.C. & Graybill, F.A., 1965: An introduction to statistical models in geology, McGraw-Hill for nærmere beskrivelse). Metoden er specielt egnet til diskrete variable som netop tælle tal.

#### En forældet tradition?

Vi har en hundredårig tradition for at tolke de danske istidsdannelse fra Weichsel i en model, der siger isfremstød fra forskellige retninger. Hvor længe kan man undlade at tage stilling til, at det datagrundlag, som tolkningen hviler på, ikke lever op til moderne videnskabelig standard?

#### Litteratur:

Houmark-Nielsen, M. 1987: Pleistocene stratigraphy and glacial history of the central part of Denmark. *Bulletin of the Geological Society of Denmark* 36.  
 Marcussen et al 2006: De danske glacial dannelser: undersøgelsesmetoder og resultater. *Geologisk Tidsskrift*.

Tabel 1. Fingrusanalyser af moræneler fra Nordsamsø (observerede værdier, O) fra Houmark-Nielsen (1987). Alle 5 prøver er taget i samme morænebænk (multiple prøver). Tallene er beregnet ud fra de procenttal og totaltal, der er angivet af Houmark-Nielsen. De afrundingsfejl, der kan være opstået, anses for ubetydelige.

#### Observerede værdier

Prøvenr.	Kvarts	Flint	Kryst.	Sed.	Paleoz.	Kretassisk	Sum
305	18	93	150	45	27	33	366
306	8	70	113	14	23	27	255
307	15	93	137	47	20	61	373
308	6	102	157	54	29	45	393
309	19	111	145	34	34	53	396
Sum	66	469	702	194	133	219	1783

Tabel 2. Hvis man forudsætter, at alle bjergartstyper forekommer med samme frekvens i alle prøver, kan man beregne en forventet værdi, E, som række-sum  $\times$  kolonne-sum / totalsum.

#### Forventede værdier

Prøvenr.	Kvarts	Flint	Kryst.	Sed.	Paleoz.	Kretassisk	Sum
305	13,5	96,3	144,1	39,8	27,3	45	366
306	9,4	67,1	100,4	27,7	19	31,3	255
307	13,8	98,1	146,9	40,6	27,8	45,8	373
308	14,5	103,4	154,7	42,8	29,3	48,3	393
309	14,7	104,2	155,9	43,1	29,5	48,6	396
Sum	66	469	702	194	133	219	1783

Tabel 3. For hver celle beregnes nu  $(O-E)^2/E$ . Tallet angiver, hvor meget den observerede værdi afviger fra den forventede. Hvis de to er lige store, bliver værdien 0. Summen fra alle celler kaldes  $\chi^2$ . Der findes tabeller over  $\chi^2$ , som angiver, hvor stor sandsynligheden er for, at  $\chi^2$  bliver så stor som en vis værdi på grund af statistisk variation, selvom alle tællinger er fra samme population. Man skal dog også tage hensyn til antallet af frihedsgrader, som er lig  $(\text{antal rækker} - 1) \times (\text{antal kolonner} - 1)$ . I eksemplet her er  $\chi^2$  beregnet til 38,8, og antallet af frihedsgrader er 20. Et tabelopslag viser, at en så stor værdi af  $\chi^2$  kun forekommer en gang ud af 100 på grund af statistisk variation, hvis prøverne er fra samme population. Sandsynligheden for, at det er tilfældet, er med andre ord mindre end 1 %.

#### $\chi^2$ -bidrag

Prøvenr.	Kvarts	Flint	Kryst.	Sed.	Paleoz.	Kretassisk	Sum
305	1,5	0,1	0,2	0,7	0	3,2	5,7
306	0,2	0,1	1,6	6,8	0,8	0,6	10,2
307	0,1	0,3	0,7	1	2,2	5	9,3
308	5	0	0	3	0	0,2	8,3
309	1,3	0,4	0,8	1,9	0,7	0,4	5,5
Sum	8,1	1	3,3	13,4	3,7	9,4	38,8