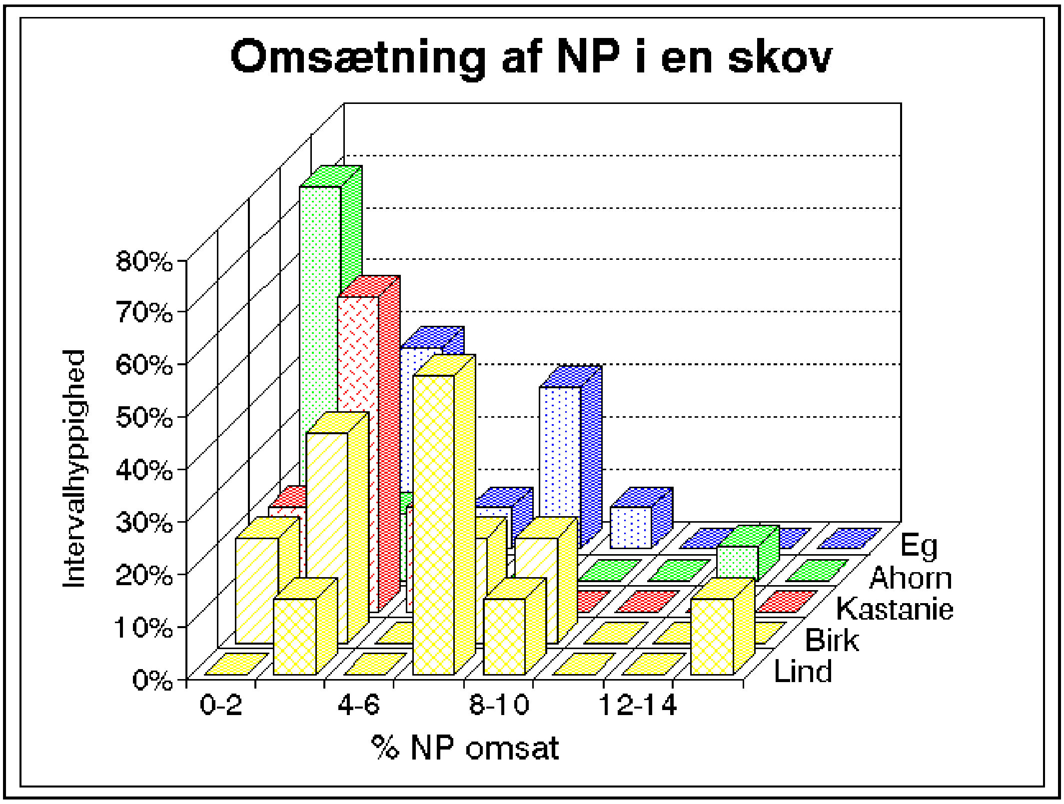
Om vurdering af talmateriale



1998

Indhold

	Indledning		*side 1*
1	**Resultatbehandling**		
	I	*Skemaopstilling*	*side 2*
	I	*Variation og spredning*	*side 3*
	III	*Grafisk vurdering*	*side 5*
2	**Statistisk analyse**		
	I	*Test af normalfordeling*	*side 6*
	II	*F-test og t-test*	*side 6*
	III	*Variansanalyse*	*side 8*
	IV	*Lineær tilpasning*	*side 9*
3	**Regneark**		
	I	*Definitioner*	*side 13*
	II	*Import af tal*	*side 13*
	III	*Formler*	*side 14*
	IV	*Grafer*	*side 14*
	V	*Statistik*	*side 15*
	Register		*side 19*

□ *Thorkild Steenberg* □ *Århus Akademi* □

1998

Indledning

Resultater af biologiske forsøg eller undersøgelser vil ofte have form af en række måledata eller beregninger på basis af målinger.
Hvordan overskues og behandles sådanne forsøgsresultater?
Ofte ønsker man svar på spørgsmålet: "Er der forskel på resultaterne af denne og hin måling?"

I det følgende anvendes som eksempel målinger af hvor mange procent af blad-nettoproduktionen fra forskellige træarter i en skov, der er fortæret af planteædende insekter:

Eg:	8,4	4,8	7,7	2,3	3,8	6,06	3,44	7,14	0
	3,2	6	0	3,3					
Ahorn:	0,57	0,91	2,5	0,88	0,39	13,51	0,22	2,5	1,67
	5,93	0,04	0,017	0,92	1,26	1,118	0,45		
Kastanie:	2,04	3,68	3,83	5,88	0,005				
Birk: 2,92	2,19	9,44	7,85	0,63					
Lind:	3,25	7,222	6,227	7	9,333	6,35	15,909		

Regneark er meget velegnet til bearbejdning af sådanne forsøgsresultater, men lommeregner med statistiktaster kan også anvendes.

Forsøget skal give svar på to spørgsmål:

1 hvor meget af bladbiomassen er fortæret af planteædere? - og
2 er der forskel på plantearterne med hensyn til hvor meget, der er ædt?

Resultatbehandling

I Opstil resultater i et skema

Systematisér og skab overblik over resultaterne ved at skrive dem op i en tabel (figur 1) - brug regnearksfaciliteterne til at sætte overskrift, skillelinier, fed skrifttype, antal decimaler, m.m. (se Brug af regneark side 11).

Svaret på spørgsmål 1 kan man nu få ved at beregne gennemsnit af resultaterne (figur 2).

% NP omsat gennem græsningsfødekæderne i en skov

Eg	**Ahorn**	**Kastanie**	**Birk**	**Lind**
8,400	0,570	2,040	2,920	3,250
4,800	0,910	3,680	2,190	7,222
7,700	2,500	3,830	9,440	6,227
2,300	0,880	5,880	7,850	7,000
3,800	0,390	0,005	0,630	9,333
6,060	13,510			6,350
3,440	0,220			15,909
7,140	2,500			
0,000	1,670			
3,220	5,930			
6,000	0,040			
0,000	0,017			
3,030	0,920			
	1,260			
	1,118			
	0,450			

Figur 1.

Resultater opstillet i skemaform.

gns.	**4,3**	**2,1**	**3,1**	**4,6**	**7,9**
total gns.	**4,0**				

Figur 2.

Udregnede gennemsnit.

Delkonklusion 1

Der omsættes 4,3 % Egeblade - 2,1 % Ahornblade - 3,1 % Kastanieblade - 4,6 % Birkeblade og 7,9 % Lindeblade, eller der omsættes i totalgennemsnit 4,0 % bladnettoproduktion af konsumenterne i en skov.

II Resultaternes variation

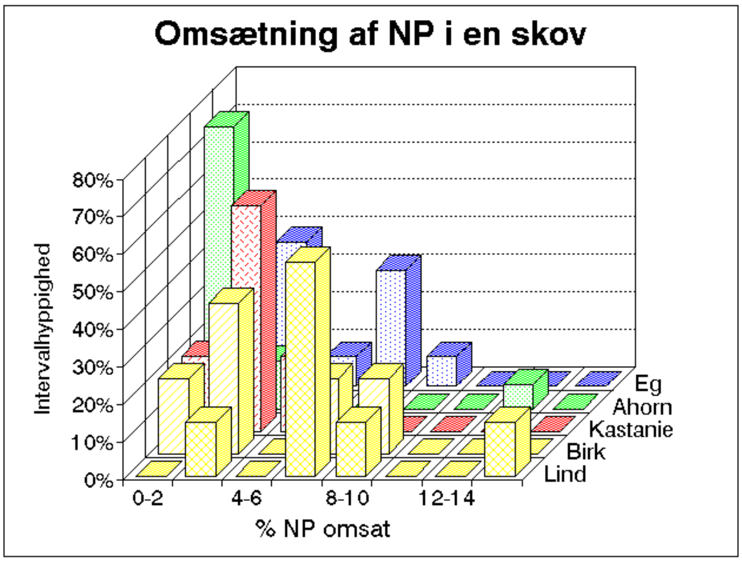
Spørgsmål 2 kan ikke umiddelbart besvares. Resultaterne tyder på, at Lind og Ahorn adskiller sig fra de øvrige, medens der kun er små eller ingen forskelle mellem Eg, Kastanie og Birk. Om der er reel forskel på gennemsnitsværdierne afhænger af hvor "godt" de er bestemt, dvs. hvor stor variation, der er imellem de enkelte måleresultater.

interval	Eg nedre grænse	Hyp-pighed	Hyp-pighed
0-2	0	2	15%
2-4	2	5	38%
4-6	4	1	8%
6-8	6	4	31%
8-10	8	1	8%
10-12	10	0	0%
12-14	12	0	0%
14-16	14	0	0%
	i alt	13	100%

Det kan give et overblik at tegne stolpediagrammer (histogrammer) over resultaternes fordeling.
Vælg et passende antal, lige store intervaller og optæl hvor mange værdier, der falder i hvert interval. Værdier der er lig med et intervals øvre grænse medtages i dette interval, medens værdier der er lig med nedre intervalgrænse medtages i forrige interval [1].
Lav fordelingstabeller og regn %-fordelingen ud (figur 3, se også side 15).

Figur 3. *Eksempel på fordelingstabel. Intervalhyppighederne er angivet som antal og procent. (NB! fordelingstabellen er lavet med regnearkstandarden for intervalgrænser).*

Tegn et diagram med intervaller afsat på x-aksen og intervalhyppighed i % afsat på y-aksen (figur 4).
Selv om regnearket kan lave flotte tredimensionale diagrammer, er det oftest mere overskueligt at anvende todimensionale diagrammer.



Figur 4. *Intervalhyppigheder (%) indtegnet i et stolpe-diagram.*

1 åÒãÕŸ ÕõÒãúÒøú' É øÛÜãÛÒøŸ ÒãýÛãúÛÕ ÐûõÛ úÛã ÙãõÛøãÒõÙÐãÒÿÛ ÕõÒãúÒøú ÃÛú ãÛúøÛ ÜøŮãÕÛ ÙãŸÿÝúÛøÛõ Ù ÙãõÛøýÒÿÿÛõ'

Resultaterne fordeler sig mere eller mindre jævnt på begge sider af en gennemsnitsværdi. Hvis man forestiller sig et ideelt, meget stort forsøgsmateriale ville resultaterne fordele sig symmetrisk omkring gennemsnitsværdien og udgøre det man kalder en normalfordeling.

Det faktiske måleresultat kan så betragtes som en stikprøve af dette ideelle, normalfordelte forsøgsmateriale, og det man behøver er et mål for stikprøvens pålidelighed som udtryk for denne fordeling.

Beregner man den gennemsnitlige afvigelse mellem måleværdierne og gennemsnittet fås resultaternes **spredning (σ)**:

$$\sigma = \sqrt{\frac{\sum (\bar{x}-x_i)^2}{n-1}} \quad ; \quad \left[\begin{array}{lcl} \Sigma & = & \textit{summationstegn} \\ \bar{x} & = & \textit{gennemsnit} \\ x_i & = & \textit{måleværdi} \\ n & = & \textit{antal måleværdier} \end{array} \right]$$

Spredningen udtrykker resultaternes "samling" om gennemsnittet. Jo mindre σ er, des bedre er gennemsnitsværdien bestemt og man kan så tillade sig at betragte to gennemsnit som forskellige, hvis deres respektive spredningsintervaller (gns -σ til gns + σ) ikke overlapper eller kun overlapper lidt.
I figur 5 er resultaterne vist med udregnet spredning.

	Eg	**Ahorn**	**Kastanie**	**Birk**	**Lind**
gns	4,3	2,1	3,1	4,6	7,9
spredning	2,7	3,4	2,2	3,8	4,0
total gns.	4,0				
total spr.	3,7				

Figur 5. *Uddrag af resultatskema med udregnet spredning.*

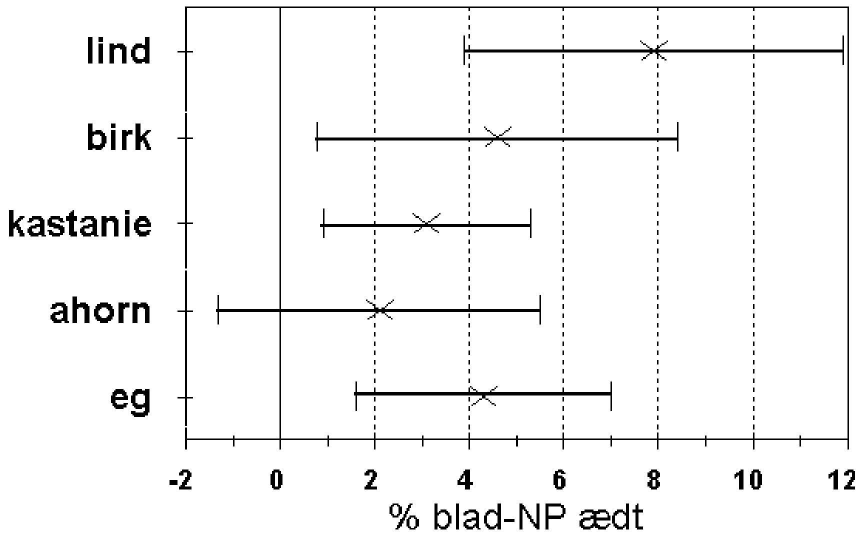
Delkonklusion 2

Man har valgt $\bar{x} \pm \sigma$ som en standard resultatangivelse.
Svaret på spørgsmål 1 (jvf side 2) bør derfor korrekt skrives:

Der omsættes 4,3 % ± 2,7 egeblade - 2,1 % ± 3,4 ahornblade - 3,1 % ± 2,2 kastanieblade - 4,6 % ± 3,8 birkeblade og 7,9 % ± 4,0 lindeblade eller totalt 4,0 % ± 3,7 af alle blade.

III Grafisk vurdering af resultater

Svaret på spørgsmål 2 kan man nu få ved at sammenligne spredningsintervallerne (gennemsnit - σ til gennemsnit + σ; figur 6) for forsøgsresultaterne, jvf. teksten side 4.



Figur 6. *Diagram der viser spredningsintervaller for forsøgsresultaterne (X markerer gns.).*

Resultaterne for Birk, Kastanie, Ahorn og Eg adskiller sig ikke indbyrdes fra hinanden. Lind adskiller sig kun usikkert fra hele gruppen. Derimod kan resultaterne for Lind og Ahorn med rimelighed betragtes som forskellige.

Slutkonklusion

Mellem 2,1 % og 7,9 % af blad-NP i en skov ædes af dyr i græsningsfødekæden.
Der er ingen tydelig forskel mellem resultaterne for Ahorn (2,1 %), Kastanie (3,1 %), Eg (4,3 %), Birk (4,6 %) og Lind (7,9 %), hvis man sammenligner dem indbyrdes; dog er der med rimelig sikkerhed en forskel mellem det mindste resultat: Ahorn (2,1 %) og det største resultat: Lind (7,9 %).

□ □ □

Statitisk analyse

Længere end ovenstående kan man ikke komme uden en egentlig statistisk analyse.

I Test af normalfordeling

Den grundlæggende forudsætning bør testes: nemlig at der er tale om normalfordelte værdier i en stikprøve.

Ud fra fordelingstabellerne (side 4) beregnes den kumulerede hyppighed ved at fordelingshyppighederne (%) lægges sammen løbende (1. felt alene, 2. felt lægges sammen med 1. felt, 3. felt lægges sammen med 1. og 2. felt, etc). Disse kumulerede hyppigheder afsættes derefter i et digram på statistikpapir (normalfordelingspapir). Ligger de afsatte punkter jævnt om - eller på - en ret linie er der tale om en normalfordeling.

II Afgørelse af om der er forskel på gennemsnitsværdier

Her må man skelne mellem 1) tilfælde hvor kun to gennemsnitsværdier skal sammenlignes og 2) tilfælde hvor flere end to gennemsnitsværdier skal sammenlignes.

II A Sammenligning af to middelværdier (*t*-test)

Den statistiske test der anvendes til sammenligning af to middelværdier (***t*-test**) har som forudsætning, at varianserne (**var = σ^2**) for de to stikprøver, der skal sammenlignes, er ens. Derfor må det først afgøres, om de foreliggende varianser er tilstrækkeligt ens.

Denne test kaldes en ***F*-test**

$$F = \frac{\text{største varians}}{\text{mindste varians}} \quad ; \begin{bmatrix} m = \text{antal målinger i tælleren} - 1 \\ n = \text{antal målinger i nævneren} - 1 \end{bmatrix}$$

Hvis $\sigma_1^2 = \sigma_2^2$ er $F = 1$. Det vil sige, hvis F værdien er tilstrækkeligt tæt på 1, er der ikke noget, der strider mod at acceptere, at de to stikprøvevarianser er ens.

F værdien kan slåes op i en tabel (m og n kaldes frihedsgrader), og sandsynligheden for at varianserne er ens kan aflæses. Alternativt kan man med de indbyggede statistiske funktioner i regnearket beregne ***F***-værdien og sandsynligheden på én gang (eksempel figur 7 og fremgangsmåde side 15).
Hvis ***F*** er mindre end den kritiske værdi - det samme som at sandsynligheden $P(F \leq f) \geq 0.05$ - accepteres det, at varianserne er ens.

F-Test: Varianser for to stikprøver	Eg	Ahorn
Middelværdi	4,299230769	2,092313
Varians	7,285174359	11,3305
Observationer	13	16
df	12	15
F	1,555282322	
P(F < = f) enhalet	0,207457029	
F-kritisk enhalet	2,017070295	

Figur 7. *Eksempel på **F**-test v.h.a. regneark.*

I figur 7 sammenlignes varianserne for eg og ahorn ved en ***F***-test.

F er mindre end den kritiske værdi, dvs det accepteres at varians-erne er ens.

Derefter fortsættes med den egentlige sammenligning af middelværdierne: en ***t*-test**

Hvis $gns_1 = gns_2$ er $t = 0$. Det vil sige, hvis $|t|$-værdien er tilstrækkeligt tæt på 0, er der ikke noget, der strider mod at acceptere, at de to gennemsnit er ens.

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{\left(\frac{1}{n_x} + \frac{1}{n_y}\right)\frac{\sum x_i^2 - n_x}{n}}}$$

***t**-Test: To stikprøver med ens varianser*	*Eg*	*Ahorn*
Middelværdi	*4,299230769*	*2,092313*
Varians	*7,285174359*	*11,3305*
Observationer	*13*	*16*
Samlet varians	*9,532579102*	
Hypotetisk mid-delforskel	*0*	
df	*27*	
t	*1,914316354*	
P(T < = t) enhalet	*0,033116602*	
t-kritisk enhalet	*1,703288444*	
P(T < = t) tohalet	*0,066233205*	
t-kritisk tohalet	*2,051830516*	

Figur 8. *Eksempel på **t**-test v.h.a. regneark.*

t-værdien kan slåes op i en tabel ($n_x + n_y - 2$ = antal friheds-grader) og sandsynligheden for at gennemsnittene er ens kan aflæses; men her er det absolut en fordel at bruge regnearkets statistiske funktioner! (se eksempel side 8 og fremgangsmåde side 16).

Hvis $|t|$ er mindre end den kritiske værdi - det vil sige det samme som at sandsynligheden $P(T \le t) \ge 0{,}05$ - accepteres, at gennemsnittene er ens.

I figur 8 sammenlignes middelværdierne for Eg og Ahorn

ved en ***t***-test.
t er større end den kritiske værdi, dvs de to middelværdier er ikke ens.

Slutkonklusionen på side 5 må altså revideres til, at der er forskel på resultaterne for Ahorn og Eg; men da man ikke kan afgøre om der er forskel mellem alle træerne ved at anvende ***t***-testen parvis på resultaterne, må man i stedet anvende en statistiktype, der giver mulighed for at sammenligne flere datasæt: en ***variansanalyse***.

II B Sammenligning af flere middelværdier (variansanalyse)

Man kan bruge en *variansanalyse* til at teste hypotesen at et antal stikprøver har det samme gennemsnit. Variansen mellem grupper sammenlignes med variansen inden for grupperne; der anvendes også her en ***F***-test:

$$F = \frac{\dfrac{\sum_1^G (\bar{x}_g - \bar{x})^2}{(G-1)}}{\dfrac{\sum_1^G \sum_1^{n_g} (x - \bar{x}_g)^2}{(N-G)}} \;;\qquad \left|\begin{array}{l} x = \textit{enkelte data} \\ \bar{x}_g = \textit{gruppegns.} \\ G = \textit{antal grupper} \\ N = \textit{total antal} \\ \bar{x} = \textit{total gns} \\ n_g = \textit{antal data i gruppen} \end{array}\right|$$

Tælleren udtrykker gruppegennemsnittenes afvigelse fra totalgennemsnit og nævneren udtrykker de enkelte datas afvigelse fra deres gruppegennemsnit. Hvis de to variansudtryk er ens er ***F*** = 1. Det vil sige, at hvis ***F***-værdien er tilstrækkeligt tæt på 1, er der ikke noget der strider mod at acceptere, at stikprøvegennemsnittene er ens. Hvis ***F*** er mindre end den kritiske værdi - det samme som at sandsynligheden P(***F***≤ f) ≥ 0,05 - accepteres det. at gennemsnittene er ens (se eksempel figur 9).

ïÒøÙÒãÕÒãÒÿÞÕÛ ÛãÕÙúÙÜ

ÇøÝđđÛø	ÀãõÒÿ	ÎÝÃ	ÇÛããÛÃ¸ ÕãÙõ	ïÒøÙÒãÕ
ÆÜ	[illegible]	[illegible]	[illegible]	[illegible]
ÀüÐøã	[illegible]	[illegible]	[illegible]	[illegible]
ÊÒÕõÒãÙÛ	[illegible]	[illegible]	[illegible]	[illegible]
àÙøŸ	[illegible]	[illegible]	[illegible]	[illegible]
êÙãú	[illegible]	[illegible]	[illegible]	[illegible]

ÀãÒÿÞÕÛ	Îî	úû	Ëî	æ	è¸ýŮøúÙ	æ¸ŸøÙõÙÕŸ
ËÛÿÿÛÃ ÜøÝđ¸ đÛø	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
ÉãúÛã ûÐø ÜøÝđđÛø	[illegible]	[illegible]	[illegible]			

Figur 9: *Eksempel på variansanalyse v.h.a. regneark*

Regnearkudgaven af *variansanalysen* har den begrænsning, at der skal være samme antal data i alle de grupper, som man ønsker at sammenligne.

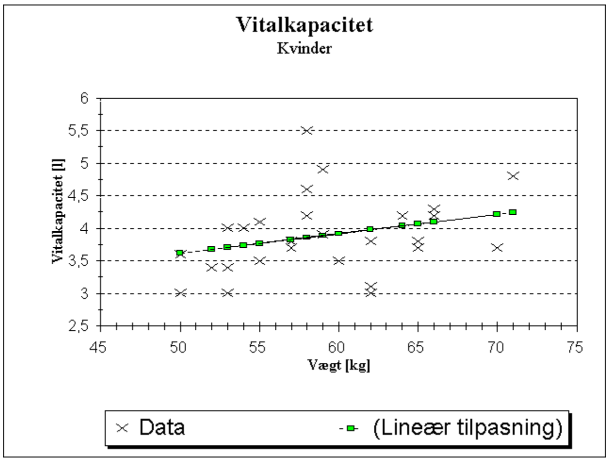
I figur 9 er der derfor kun medtaget de fem første data for hver træart i variansanalysen.

F er større end den kritiske værdi; det vil sige at gennemsnitsværdierne i eksemplet ikke er ens.

Ud fra en delmængde af resultaterne må slutkonklusionen på side 5 altså revideres til at der alligevel er forskel på hvormeget af de fem træers NP, der fortæres af insekter.

III Lineær tilpasning (regression)

Ofte har man i biologiske forsøg samtidige målinger af to eller flere data, som man vil undersøge den indbyrdes sammenhæng imellem.



Figur 10 *Vitalkapacitet som funktion af vægt hos kvinder*

Figur 10 viser som eksempel målinger af vægt og vitalkapacitet (dvs forskel mellem maksimal indånding og maksimal udånding) hos kvinder.

Målingerne er afsat i et x-y diagram (se side 15) og opgaven består i at bestemme et udtryk for den rette line, der bedst muligt tilpasses punktmængden.

Man vælger linien således, at summmen af kvadratet på afstanden mellem de beregnede y-koordinater og de faktiske y-koordinater bliver mindst mulig:

$$\sum (y_i - y_{bi})^2 \; ; \; \left[\begin{array}{l} y_i = \textit{målt y-koordinat} \\ y_{bi} = \textit{beregnet y-koordinat} \end{array} \right]$$

Metoden kaldes *lineær tilpasning* efter mindste kvadraters metode.
Ligningen for linien er $y = \boldsymbol{\alpha} x + \boldsymbol{b}$ og hældningskoeeficienten $\boldsymbol{\alpha}$ og skæringspunktet med y-aksen ***b*** kan beregnes efter følgende udtryk:

$$\alpha = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} \quad , \quad b = \bar{y} - \alpha\bar{x} \quad ; \quad \left[\begin{array}{rl} \bar{x} & = \textit{gns. af x-data} \\ \bar{y} & = \textit{gns. af y-data} \\ x_i, y_i & = \textit{datasæt} \end{array} \right]$$

Regnearket kan beregne hældningskoeeficent og skæringspunkt med y-aksen (se fig 11 og fremgangsmåde side 16).

Desuden får man en beregning af korrelationskoefficienten ***r***, som angiver graden af indbyrdes afhængighed mellem de to variable: ***r*** = 0 ingen afhængighed til ***r*** = 1 total afhængighed.
Det er vigtigt at understrege, at der ikke nødvendigvis er tale om en årsagssammenhæng mellem de to variable selv om korrelationskoefficienten er høj.

ÛÜøÛÕÕÙÐãÕúÒõÒ ûÐø
ÛŸÕÛÃđÛÿ Ù ûÙÜÝø ˝˙

ÐãÕõÒãõ	¸˛˝¸κκˇ¸
ÎõÒãúÒøúÒûý˛ ûÐø ¸ÛÕõÙÃÒõ	¨˛˘˘˛˘˘
ýÒúøÒõÛõ đ	¨˙ß˛˝˘κ
ÀãõÒÿ Ð ÕÛøýÒõÙÐãÛø	¸ß
æøÙ ÛúÕÜøÒúÛø	¸¯

¸ŸÐÛûûÙ ÙÛãõ Ûø	¨˙¸κˇßß
ÎõÒãúÒøúÒûý˛ ûÐø ŸÐÛûûÙ ÙÛãõ	¨˙˝κ˝ˇ˛

Figur 11 *Eksempel på lineær tilpasning v.h.a. regneark*

Sandsynligheden for at en ***r***-værdi er forskellig fra 0 kan findes med en en ***t***-test:

$$t = \sqrt{\nu \; r^2 \; (1 - r^2)} \; ; \quad \nu = n - 2$$

Hvis $t > t_k$ (alternativt $0{,}025 \le P(T \le t) \le 0{,}0975$) kan det ikke accepteres at ***r*** er lig med 0 (dvs. ***r*** signifikant forskellig fra 0).
Se tabel over udvalgte værdier af t_k og n side 16.

Ligningen for eksemplet i figur 10 bliver ved indsætning af værdierne i figur 11:

$$\text{vitalkap}_{\text{kvinder}} = 0{,}0298 * \text{vægt}_{\text{kvinder}} + 2{,}1299; \qquad r^2 = 0{,}085179;$$

men udføres en ***t***-test på korrelationskoefficienten som beskrevet ovenfor fås at ***r*** ikke adskiller sig fra 0; det vil sige at datamaterialet ikke er tilstrækkeligt til at påvise en sammenhæng mellem vitalkapacitet og vægt.

□ □ □

Regneark

Brug af regneark til beregninger (QuattroPro windows-version)

Afsnittet indeholder de vigtigste regnearksdefinitioner og regnearksfunktioner (inklusive statistiske) til talbehandling samt formler til beregning af gennemsnit og spredning udfra fordelingstabeller (grupperede data).
Desuden vejledning til import af datasæt fra databaser eller datafangstmoduler og vejledning til at oprette diagrammer over datamaterialet..

Eksemplerne i diagramdelen knytter sig til foregående tekst eller anvender data fra forsøget "Kropsmål og proportioner"; dvs målinger af højde, vægt, blodtryk, puls m.m.

□ □ □

DEFINITIONER OG FUNKTIONER

I DEFINITIONER

BLOK

En eller flere celler eller rækker eller kolonner

MARKER BLOK

Peg på øverste celle, hold venstre museknap nede og træk til nederste celle.
Slip museknappen (blokken er markeret med en afvigende farve).
Eller peg på øverste celle, tryk < ***skift F7*** > og dernæst < ***pil ned*** > / < ***pil højre*** > sålænge det er nødvendigt.

ÆNDRE BLOKKENS EGENSKABER

Klik med højre museknap på en celle eller blok, vælg < ***egenskaber*** >.
Her kan vælges skrifttype, skriftstørrelse, fed, kursiv, farve, format, linietegning, m.m.

KOLONNEBREDDE

Kolonnebredden kan automatisk afpasses efter det bredeste celleindhold med ikonen: **< ↔ >.**

KOPIÉR

Markér celle eller celler der skal kopieres, tryk på < ***kopiér*** > ikonen. Marker øverste celle, der skal kopieres til og tryk på < ***indsæt*** > ikonen.

FLYT

Peg på cellen eller blokken med venstre museknap. Hold knappen indtrykket og vent et lille øjeblik, indtil en lille hånd viser sig.
Flyt cellen eller blokken - med museknappen indtrykket - til den nye placering.
Slip museknappen.

II IMPORT AF TALMATERIALE

Import af kommaseparerede filer fra databaser eller datafangstmoduler.
Importen foregår i tre trin.

1) midlertidig ændring af regnearkets decimaltegn og adskillelsestegn, så det passer til databasens,
2) selve importen,
3) tilbageføring af regnearkets decimaltegn og adskillelsestegn til sædvanlig standard.

1) Vælg < ***Egenskaber*** > ***/applikation /international.***

Der fremkommer en liste med forskellige muligheder. Vælg kombinationen med decimalpunktum og kommaseparering: 1 234.56 (a1,a2). Afslut med < ***ok*** >.

2) Vælg < ***Værktøj*** > ***/importér.***

a Erstat *.prn i filruden med *.kom
b Vælg drev og bibliotek (fx a:\).
  Udpeg den ønskede fil.
c Afkryds "komma og anførselstegn".
d Afslut med < ***ok*** >.

3 Vælg < ***Egenskaber*** > ***/applikation /international.***

Vælg kombinationen med decimalkomma og punktumseparering: 1 234,56 (a1.a2). Afslut med < ***ok*** >.

Efter dataimporten kan regnearket tilrettes med overskrifter, flytning af importerede overskrifter, rettelse af fejlimporterede tegn (ø, ≥, ≤ og æ), m.m.

III FORMLER

1 Beregning af gennemsnit for en blok: @AVG(blok).

Vælg en fri celle under eller til højre for blokken.
Skriv @AVG("første celle"."sidste celle"), og tryk < ***retur*** > .
Blokken kan også udpeges med musen eller med < ***skift F7*** > og piletasterne.

2 Beregning af spredning for en blok: @STDS(blok).

Vælg en fri celle under eller til højre for blokken.
Skriv @STDS("første celle"."sidste celle"), og tryk < ***retur*** > .
Blokken kan også udpeges med musen eller med < ***skift F7*** > og piletasterne.

3 Summering af kolonner eller rækker: **<** ***autosum*** **>** **(Σ)** ikonen.

Afmærk kolonnen eller rækken + én ekstra celle.
Tryk på **Σ** ikonen.

4 Beregning af gennemsnit ud fra fordelingstabeller (grupperede data):

$$\bar{x} \approx \frac{\sum (n_x\, x)}{\sum n_x} \; ; \quad \left[\begin{array}{l} \bar{x} = \text{gennemsnit} \\ \sum = \text{summationstegn} \\ n_x = \text{antal i klasse} \\ x = \text{klassemidte} \end{array} \right]$$

Lav et antal kolonner til udregning af formelen.

5 Beregning af spredning ud fra fordelingstabeller (grupperede data):

$$\sigma \approx \sqrt{\frac{\sum (n_x\, x^2) - \dfrac{(\sum (n_x\, x))^2}{\sum n_x}}{(\sum n_x) - 1}} \; ; \quad \left[\begin{array}{l} \sigma = \text{spredning} \\ \sum = \text{summationstegn} \\ n_x = \text{antal i klasse} \\ x = \text{klassemidte} \end{array} \right]$$

Lav et antal kolonner til udregning af formelen.

IV GRAFER

1 Stolpediagrammer

Vælg < ***Graf*** **>** ***/ny.***

Peg på knappen < ***x-akse*** > . Pilen skifter til et miniature-regneark.
Klik på knappen og x-akse tallene (dvs intervallerne i fordelingstabellen, fx. figur 3, side 3) kan markeres (med musen eller < ***skift F7*** **>** og piletaster, jvf ovenfor).
Returner ved klik på pilen i grafbjælken ovenover regnearket.

Samme fremgangsmåde for 1. serie (hyppighed i %), 2. serie, … og evt. ledetekst (celler med forklaring af, hvad serierne viser).

Tryk < ***ok*** > og grafen vises.

Højreklik på grafen, vælg titler og skriv overskrift og tekst på akserne.
Afslut med < ***ok*** > .

2 x-y diagrammer

Vælg < ***Graf* > /*ny***.

Peg på knappen < ***x-akse*** >. Pilen skifter til et miniatureregneark.
Klik på knappen og marker x-akse tallene (fx højde i et højde-vægt diagram eller systolisk blodtryk i systolisk-diastolisk blodtryksdiagram) med musen eller < ***skift F7*** > og piletaster, jvf ovenfor.

Returner ved klik på pilen i grafbjælken ovenover regnearket.

Samme fremgangsmåde for 1. y-serie og evt. 2. y- serie og evt. ledetekst (celler med forklaring af, hvad serierne viser).

Hvis punktmængder skal vises separat for kvinder og mænd som i et højde-vægt diagram, skal vægtkolonnen i regnearket deles i to ved at fx mændenes vægt flyttes til en nyoprettet kolonne ved siden af den oprindelige vægtkolonne.

Alle højder (kvinder + mænd) afmærkes som én sammenhængende x-serie.
Den oprindelige vægtkolonne (nu kun med kvinde-vægte) markeres som 1. y-serie, men lige så langt som der er x-værdier, selv om de nederste celler er tomme.
Den nye vægtkolonne markeres som 2. y-serie; også lige så langt som der er x-vær-dier, selv om de øverste celler er tomme.

Tryk < ***ok*** > og grafen vises, men som et liniediagram.
Højreklik på grafen, vælg graftype og ret til x-y diagram.
Afslut med < ***ok*** > .

Højreklik på grafen, vælg linieserieegenskaber. Vælg liniestil, sæt den til ingen.
Ret evt mærkestil, størrelse og farve.
Afslut med < ***ok*** > .

Højreklik på grafen, væg titler og skriv overskrift og tekst på akserne. Afslut med < ***ok*** > .

V STATISTIK

1 Oprette fordelingstabeller til stolpediagrammer (jvf. **IV, 1**).

Vælg et passende antal, lige store intervaller og lad regnearket optælle hvor mange værdier, der falder i hvert interval.
Værdier der er lig med et intervals nedre grænse bliver medtaget i dette interval, medens værdier der er lig med øvre intervalgrænse bliver medtaget i næste interval.
(NB! dansk standard er den omvendte procedure: værdier der er lig ned et intervals øvre grænse medtages i dette interval, medens værdier der er lig med nedre intervalgrænse medtages i forrige interval).

A Opret en kolonne med nedre grænseværdier i de ønskede intervaller.

B Vælg **< værktøj > / analyseværktøj**
Der vises en ny knaplineal. Vælg histogram ikonen.

C
1. Markér datakolonnen som indlæsningsblok
2. Markér den netop oprettede grænseværdikolonne som intervalblok
3. Markér feltet lige over grænseværdikolonnen som udlæsningsblok
4. Afkryds evt kumulerede værdier
5. Tryk < **ok** > .

D Regnearket beregner og opretter en tabel med intervalhyppigheder (Bin = begyndelsesværdi i interval).

Før man går videre til at oprette et diagram, bør grænseværdierne erstattes af en intervalangivelse (fx nedre grænseværdi: 10 ▸ interval: 10 - 20).

E Tegn stolpediagrammet som beskrevet i **IV , 1**.

2 ***F***-test: test for at varianser i to stikprøver er ens

Vælg < ***værktøj* > */analyseværktøj.***

Der vises en ny knaplineal.
Vælg < ***F*** > . Udpeg 1. variabelblok (dvs 1. kolonne med data) med musen eller < ***skift F7*** > og piletasterne.
Dernæst 2. variabelblok. Hvis øverste linie indeholder kolonneoverskrifter afkrydses feltet: Labels.

Vælg et frit område nedenfor eller til højre for datablokkene og markér udlæsningsblok her.
Afslut med < ***ok*** > og regnearket opstiller en tabel med resultater.

Hvis P(F ≤f) ≥ 0,05 accepteres det, at varianserne er ens.

2 ***t***-test: test for at to gennemsnitsværdier er ens

Vælg < ***værktøj* > */analyseværktøj.***

Der vises en ny knaplineal.
Vælg < ***t*** > .
Afkryds feltet: "ens varianser".

Udpeg 1. variabelblok (dvs 1. kolonne med data) med musen eller < ***skift F7*** > og piletasterne.
Dernæst 2. variabelblok. Hvis øverste linie indeholder kolonneoverskrifter afkrydses feltet: Labels.

Vælg et frit område nedenfor eller til højre for datablokkene og markér udlæsningsblok her.
Afslut med < ***ok*** > og regnearket opstiller en tabel med resultater.

Hvis P(T ≤t) ≥ 0,05 accepteres det, at gennemsnittene er ens.

4 Variansanalyse: test for forskelle mellem flere stikprøver.

Vælg < ***værktøj* > */analyseværktøj.***

Der vises en ny knaplineal. Vælg < ***σ/1*** **>** *for én-faktoranalyse eller* **< *σ/3*** > for to-faktoranalyse.
Udpeg datablokken med musen eller < ***skift F7*** > og piletasterne. **NB! der skal være lige store antal observationer i de to eller flere kolonner med data, der indgår i analysen.**

Vælg et frit område nedenfor eller til højre for datablokkene og markér udlæsningsblok her. Afslut med < ***ok*** > og regnearket opstiller en tabel med resultater.

Hvis P(F ≤f) ≥ 0,05 accepteres det, at stikprøverne og dermed gennemsnittene er ens.

5 Lineær regression: tilpasning af en punktmængde til bedste rette linie.

Vælg < ***værktøj* > */avanceret matematik/regression.***

Der vises et nyt vindue. Udpeg den uafhængigt-variable blok (dvs x-aksen) med musen eller < ***skift F7*** > og piletasterne.
Dernæst den afhængigt-variable blok (dvs y-aksen).

Vælg et frit område nedenfor eller til højre for datablokkene og markér udlæsningsblok her.
Afslut med < ***ok*** > og regnearket opstiller en tabel med resultater.
Y-akse skæringspunkt er som standard sat til "beregnes"; hvis man vil have grafen til at gå gennem [0,0] skal afkrydsningen ændres.

Korrelationskoefficienten (r) angiver graden af indbyrdes afhængighed mellem de to variable: r = 0 ingen afhænghed til r = 1 total afhængighed.
Sandsynligheden for at en r-værdi er forskellig fra 0 kan findes med en en ***t***-test:

$$t = \sqrt{\nu\ r^2\ (1 - r^2)}\ ;\quad \nu = n - 2$$

Hvis $\boldsymbol{t} > \boldsymbol{t_k}$ (alternativt $0{,}025 \leq P(T \leq \boldsymbol{t}) \leq 0{,}0975$) kan det ikke accepteres at ***r*** er lig med 0 (dvs. ***r*** signifikant forskellig fra 0).

Tabel over udvalgte værdier af $\boldsymbol{t_k}$ og n	
$\boldsymbol{t_k}$	n
1,960	∞
1,980	120
2,048	30
2,101	20
2,306	10

□ □ □

Litteratur:

1 Bernhard Andersen & Erling B. Andersen:
Grundlæggende statistik
Gyldendal. 1978.

2 Bernhard Andersen & Erling B. Andersen:
Matematisk statistik i grundtræk
Gyldendal. 1973.

3 Arne Nielsen, Jørgen Hilden & Kirsten Fenger:
Statistik og sandsynlighed anvendt i medicin
FADL 2.udg. 1976.

4 Anon.
Applied Statistics with TI Programmable 58/59
Texas Instruments Inc. 1977.

□ □ □

Register

σ, spredning
beregning 4
definition 4
resultatangivelse 4
Blok
definition 13
flyt 13
kopiér 13
markér 13
ændre egenskaber ved 13
F-test
definition 6
regnearksfunktion 15
Fordelingstabel 3
beregning af gennemsnit 14
beregning af spredning 14
regneark 15
Gennemsnit 2
beregning 14
beregning (grupperede data) 14
resultatangivelse 4
t-test; sammenligning af to gennemsnit 6, 16
variansanalyse; sammenligning af flere gns. 8, 16
Hyppigheder
kumulerede 6
Intervalhyppighed 3
Konklusion
Delkonklusion 1 2
Delkonklusion 2 4
Slutkonklusion 5
Slutkonklusion, revideret 9
Lineær tilpasning 9
brug af regneark 16
definition 9
Måledata 1, 9
Normalfordeling 4
test af 6
Regneark 11
F-test 15
fordelingstabeller 15
lineær tilpasning (regression) 16
sammenligning af flere gennemsnit 16
sammenligning af to gennesmsnit 16
t-test 16
variansanalyse 16
Resultatangivelse 4
Spredning, σ
beregning 14
beregning fra grupperede data 14
definition 4
Spredningsintervaller 4
grafisk sammenligning af 5
Statitisk analyse 6
Stikprøve 4
Stolpediagrammer 3
i regneark 14
Summering
regnearksfunktion 14
t-test
beregning 6
regneark 16
Varians 6
Variansanalyse 8
regneark 16
Variation 3
x-y diagrammer
i regneark 15
lineær tilpasning 9

□ □ □