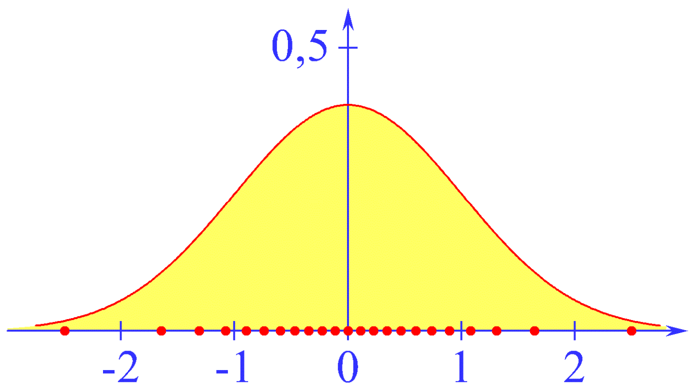
Statistik

for

matematik på B-niveau i hf



2014 Karsten Juul

TEST

FORDELINGER

En tidligere version af dette hæfte har skiftet adresse til

http://mat1.dk/statistik_for_matematik_paa_b_niveau_i_hf_2013.pdf

TEST

1 Stikprøver.

Nogen på et gymnasium mener at der er forskel på hvad piger og drenge mener om et bestemt spørgsmål. For at undesøge denne hypotese, spørger vi nogle piger og drenge.

1.1 Hvad er populationen?

De ting eller personer som vi vil påstå noget om, kaldes populationen.

Er det alle personer i europa som nu er mellem 10 og 20 år?
Er det alle elever på vores gymnasium?
Eller?

Når vi laver en statistisk undersøgelse, skal vi **skrive** en præcisering af

hvad det er for en population vi vil påstå noget om.

1.2 Hvad er stikprøven?

Vi undersøger kun en lille del af hele populationen.
De personer vi får et svar fra (eller de ting vi undersøger), kaldes stikprøven.

Når vi laver en statistisk undersøgelse, skal vi **skrive** en præcisering af

hvordan vi har valgt stikprøven.

Det er **IKKE nok at skrive**:

”Vi har spurgt 47 elever på vores gymnasium.”

Det er **nok at skrive**

”Den 20. februar mellem kl. 8:50 og 9:10 spurgte vi de 47 elever der sad på gangen, og vi fik svar fra dem alle. 10 af drengene og 8 af pigerne var fra 3g FY, 13 af drengene og 16 af pigerne var fra 3g Fy.”

eller

”Den 20. februar kl. 8:50 sendte vi en besked til alle elever på skolen. Stikprøven er de 47 elever der svarede inden kl. 10:00 den 22. februar.”

Disse to beskrivelser af en indsamling af stikprøve er så grundige at læseren kan se om der er grund til tro at der kan være systematiske fejl.

1.3 Systematiske fejl ved valg af stikprøven.

Eksempel 1

Population: Eleverne på vores gymnasium.

Stikprøve: Eleverne i en sproglig klase.

Her kan vi have lavet en systematisk fejl ved valg af stikprøven, for det kan være at en bestemt holdning oftere er blandt sproglige end blandt andre.

Eksempel 2

Hvis vi spørger elever pr. e-mail, og mange ikke svarer, så kan vi have lavet en systematisk fejl, for det er måske især elever med en bestemt holdning der svarer.

1.4 Tilfældige fejl ved valg af stikprøven.

Selv om vi vælger stikprøven tilfældigt blandt hele populationen,
er det ikke helt sikkert at den ligner populationen.

Det kan f.eks. være at vi tilfældigt har fået for mange ja-sigere med i stikprøven.

Det er muligheden for tilfældige fejl vi beskæftiger os med når vi udregner tallet p.

2 Hvad er sandsynlighed?

2.1 Eksempel.

At **sandsynligheden for at vinde = 25 %**
betyder at **vi vinder 25 % af gangene**.

2.2 Eksempel.

At **sandsynligheden for at en pose har 3 eller flere defekte = 4 %**
betyder at **4 % af poserne har 3 eller flere defekte**.

3 Test af hypotese.

3.1 Signifikansniveau.

Hypotese: *Halvdelen af brikkerne er gule.*

Vi tager en stikprøve for at teste hypotesen.
Vi får en stikprøve der ligger langt fra hypotesen.
Vi udregner at hvis hypotesen er rigtig,

så er sandsynligheden kun 3 %
for at få en stikprøve der ligger så langt eller længere fra hypotesen,

Hvis 3 % er mindre end det valgte signifikansniveau, så forkaster vi hypotesen.
Hvis vi har valgt signifikansniveau = **5 %** , så forkaster vi da 3 % < **5 %**.
Hvis vi har valgt signifikansniveau = **1 %** , så forkaster vi ikke da 3 % ≥ **1 %**.

3.2 Hvornår har vi vist noget med en test?

Kun når vi forkaster en hypotese, har vi vist noget.

Hvis vi ikke forkaster hypotesen, har vi IKKE vist at hypotesen sandsynligvis er rigtig.
Dette skyldes at der altid er flere indbyrdes modstridende hypoteser som ikke forkastes af stikprøven.

4. Test for uafhængighed i 2×2 tabel.

Vi vil teste om piger og drenge har samme holdning til et spørgsmål.

Hypotese: *Andelen der siger ja, er ens for piger og drenge.*

Hypotesen er altså at svaret er uafhængigt af om det er en pige eller en dreng.

I den slags test vi laver her, gælder altid: Hypotesen er at noget er ens.

Vi vælger: signifikansniveau = 5 %

Nogle tilfældigt udvalgte elever får stillet samme spørgsmål.
De svarede sådan:

Faktiske tal	ja	nej
piger	**87**	**46**
drenge	**71**	**21**

4.1 Sådan udregner vi FORVENTEDE TAL.

For at teste hypotesen, udregner vi først noget vi kalder de forventede tal, dvs. hvordan svarene skulle være fordelt mellem de fire felter hvis ja-andelen i tabellen skulle være ens for piger og drenge.

For at kunne udregne de forventede tal udregner vi først følgende tal:

Antal piger:	$87+46 = 133$
Antal drenge:	$71+21 = 92$
Antal ja-sigere:	$87+71 = 158$
Antal nej-sigere:	$46+21 = 67$
Antal piger og drenge:	$133+92 = 225$

Disse tal skriver vi i et skema:

	ja	nej	i alt
piger			133
drenge			92
i alt	158	67	225

I tabellen med forventede tal skal andelen af ja-sigere være den samme for piger og drenge, så da 158 af de 225 elever svarede ja, skal vi i denne tabel skrive at $\frac{158}{225}$ af de 133 piger svarede ja:

forventet antal ja-svar fra piger:	$133 \cdot \frac{158}{225} = 93{,}40$
forventet antal nej-svar fra piger:	$133-93{,}40 = 39{,}60$
forventet antal ja-svar fra drenge:	$158-93{,}40 = 64{,}60$
forventet antal nej-svar fra drenge:	$92-64{,}60 = 27{,}40$

Her er de fire forventede værdier skrevet ind i skemaet:

Forventet	ja	nej	i alt
piger	**93,40**	**39,60**	133
drenge	**64,60**	**27,40**	92
i alt	158	67	225

4.2 Sådan udregner vi χ^2.

Symbolet χ^2 læses sådan: *ki i anden*

χ^2 er et tal der udtrykker

afstanden mellem de faktiske tal og de forventede tal .

For hvert af de fire felter udregner vi

$$\frac{(faktisk - forventet)^2}{forventet}$$

Ved at lægge disse fire tal sammen får vi tallet χ^2 som er afstanden mellem de faktiske tal og de forventede tal:

$$\chi^2 = \frac{(87-93{,}40)^2}{93{,}40} + \frac{(46-39{,}60)^2}{39{,}60} + \frac{(71-64{,}60)^2}{64{,}60} + \frac{(21-27{,}40)^2}{27{,}40}$$

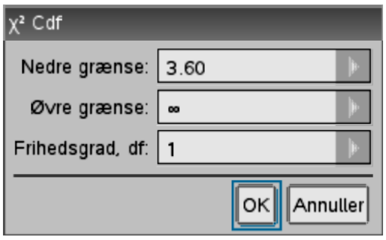
$$\chi^2 = 3{,}60$$

4.3 Sådan udregner vi *p*.

Ovenfor udregnede vi at afstanden mellem faktiske og forventede tal er $\chi^2 = 3{,}60$.

Tallet p er sandsynligheden hvis hypotesen er rigtig,
for at afstanden er 3,60 eller større.

Vi kan få Nspire til at udregne p ved i beregningsmenuen at vælge
Statistik / Fordelinger / χ^2 Cdf... og udfylde sådan:



På Nspire-lommeregneren kan vi skrive ∞ ved hjælp af π-tasten.
På computeren skriver vi infinity i stedet for ∞ da det på computeren ikke er muligt at bruge tegnpaletten i χ^2-vinduet (med mindre vi har valgt Vis / Håndholdt).

Vi får $p = 5{,}8\,\%$.

4.4 Hvordan kan vi skrive konklusionen?

Da sigifikansniveauet er 5 % , og p ikke er mindre end 5 % , kan vi ikke forkaste hypotesen.

Stikprøven giver ikke belæg for at hævde at andelen der siger ja,

er forskellig for piger og drenge.

4.5 Misforstå ikke procenterne.

5,8 % er IKKE sandsynligheden for at hypotesen er rigtig.

94,2 % er IKKE sandsynligheden for at hypotesen er forkert.

De 5,8 % er udregnet under den forudsætning at hypotesen er rigtig
og er sandsynligheden for at få en stikprøve hvis afvigelse fra hypotesen
er så stor som eller større end afvigelsen i den stikprøve vi fik.

5 Tal der er normalfordelt.

Vi måler højden af alle drenge i 3.g .

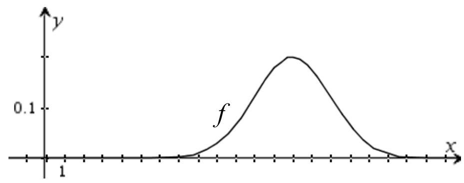
De målte tal afsætter vi som prikker på en tallinje.

Tallene ligger ikke jævnt fordelt. De er fordelt sådan:

Der er et sted på tallinjen hvor tallene ligger tæt.
Jo længere man kommer fra dette sted, jo mindre tæt ligger tallene.

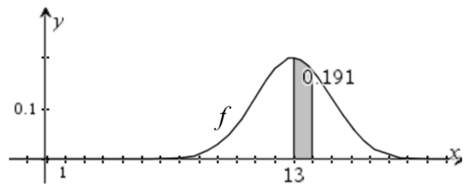
Når vi måler ting af samme slags, vil tallene ofte være fordelt på en bestemt måde som vi kalder **normalfordelt.** F.eks. er højderne af drengene i 3.g normalfordelt.

6 Graf for tal der er normalfordelt.



Grafen til venstre viser hvordan nogle målte tal er fordelt.

Hvis vi afsætter tallene som prikker på x-aksen, vil prikkerne ligge tættest der hvor y er størst, dvs. nær 13 på x-aksen ligger prikkerne tættest.

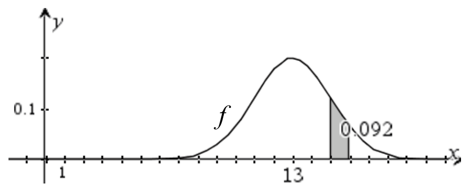


Nær 11 på x-aksen er y mindre, så nær 11 vil prikkerne ligge mindre tæt.

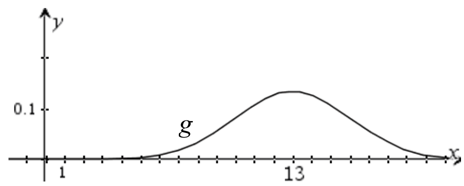
Nær 6 vil prikkerne ligge meget mindre tæt da y er tæt på 0.

Arealet under grafen er $1 = 100\,\%$.

I intervallet $13 \leq x \leq 14$ er arealet under grafen 0,191 , dvs. i dette interval ligger 19,1 % af de målte tal.



I intervallet $15 \leq x \leq 16$ er arealet under grafen 0,092 , dvs. i dette interval ligger 9,2 % af de målte tal.



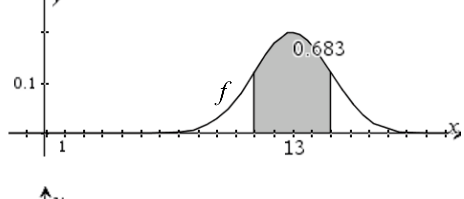
Grafen for g er bredere end grafen for f. Til gengæld er den lavere da arealet under grafen skal være 1.

Når grafen er bredere, ligger tallene mere spredt.

7 Middelværdi for tal der er normalfordelt.

Middelværdien af de målte tal er 13, da grafen er symmetrisk om $x=13$. Dette gælder både for f og g.

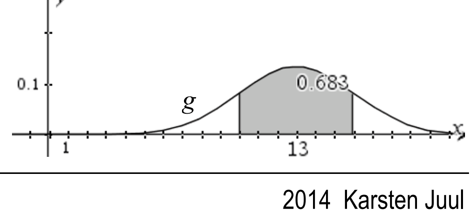
8 Spredning for tal der er normalfordelt.



På f-grafen ser vi at hvis vi fra middelværdien går 2 ud til begge sider, så får vi 68,3 % af tallene med.

Vi siger at tallenes spredning er 2 fordi vi skal gå 2 ud fra middelværdien for at få 68,3 % af tallene med.

På g-grafen ser vi at spredningen er 3.



9 Forskrift for funktionen der viser fordelingen af tal der er normalfordelt.

Når tal er normalfordelt med middelværdi m og spredning s, så vil funktionen f der viser fordelingen, have følgende forskrift:

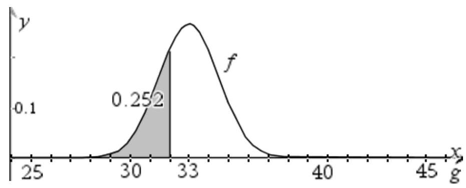
$$f(x) \;=\; \frac{1}{s\cdot\sqrt{2\pi}}\,\mathrm{e}^{-\frac{1}{2}\left(\frac{x-m}{s}\right)^2}$$

Denne forskrift skal du IKKE huske!

10 En anvendelse af normalfordeling.

Ved fremstillingen af en vare er der på grund af tilfældigheder stor forskel på vægtene af varerne.
Vægtene er normalfordelt med spredning 1,5 gram.
Varer der vejer under 32 gram skal kasseres.
(Jo tungere varen er, jo dyrere er det at fremstille den).

Opgave: Hvor stor en del af varerne skal kasseres hvis maskinen indstilles til at fremstille varer med middelværdi 33 gram ?

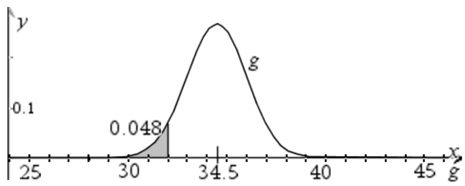


Svar: Lad f være funktionen der viser fordelingen af tal der er normalfordelt med middelværdi $m = 33$ og spredning $s = 1{,}5$.
Arealet under f-grafen i intervallet $-\infty < x < 32$ er

$$\int_{-\infty}^{32} f(x)\,dx \;=\; 0{,}252$$ *Udregnet af Nspire.*

så 25,2 % af varerne skal kasseres.

Opgave: Hvor stor en del af varerne skal kasseres hvis maskinen indstilles til at fremstille varer med middelværdi 34,5 gram ?

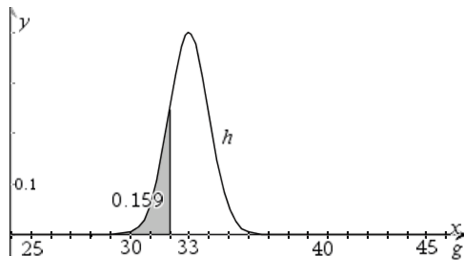


Svar: Lad g være funktionen der viser fordelingen af tal der er normalfordelt med middelværdi $m = 34{,}5$ og spredning $s = 1{,}5$.
Arealet under g-grafen i intervallet $-\infty < x < 32$ er

$$\int_{-\infty}^{32} g(x)\,dx \;=\; 0{,}048$$ *Udregnet af Nspire.*

så 4,8 % af varerne skal kasseres.

Opgave: En ny maskine laver varer der er normalfordelt med spredning 1 gram. Hvor stor en del af varerne skal kasseres hvis maskinen indstilles til at fremstille varer med middelværdi 33 gram ?



Svar: Lad h være funktionen der viser fordelingen af tal der er normalfordelt med middelværdi $m = 33$ og spredning $s = 1$.
Arealet under h-grafen i intervallet $-\infty < x < 32$ er

$$\int_{-\infty}^{32} h(x)\,dx \;=\; 0{,}159$$ *Udregnet af Nspire.*

så 15,9 % af varerne skal kasseres.

11 χ^2-fordeling.

Nogle tal er normalfordelt med middelværdi 0 og spredning 1.
På figur 9 har vi tegnet nogle få af disse tal som prikker.
Disse få af tallene har vi valgt sådan at de giver et indtryk af hvordan alle tallene er fordelt.

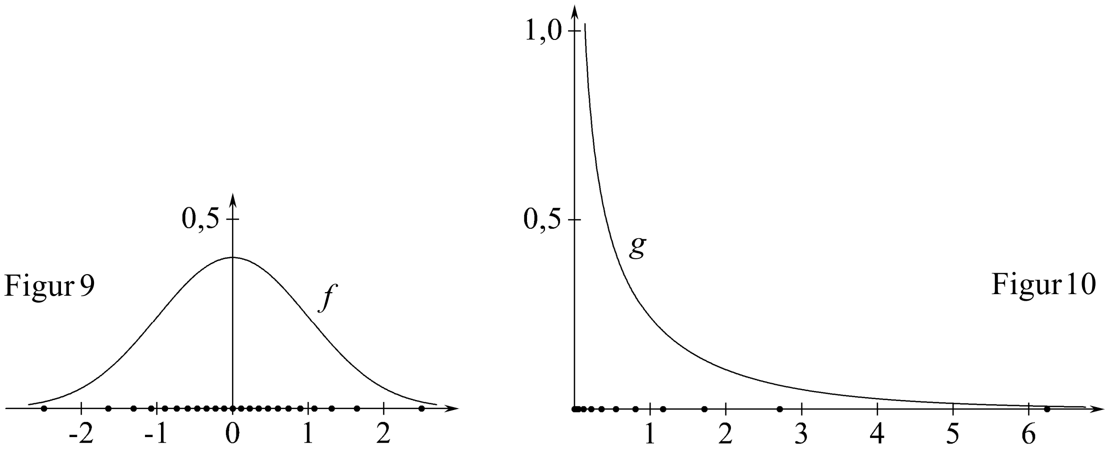
Hvert af tallene opløfter vi til anden:

$(-0{,}6)^2 = 0{,}36$

$2{,}5^2 = 6{,}25$

osv.

De tal vi får på denne måde, er fordelt som antydet på figur 10. Vi vil finde forskriften for funktionen g der viser hvordan disse tal er fordelt. (Kun få af tallene er vist som prikker).



Figur 9

Figur 10

g-grafen viser hvordan tallene χ^2 fra 4.2 er fordelt, dvs.

**hvis vi mange gange tager en stikprøve og hver gang udregner χ^2 ,
så vil vi få nogle tal der er fordelt som g-grafen viser.**

Da vi i afsnit 4 tog en stikprøve hvor $\chi^2 = 3{,}60$, kunne vi altså have udregnet p sådan:

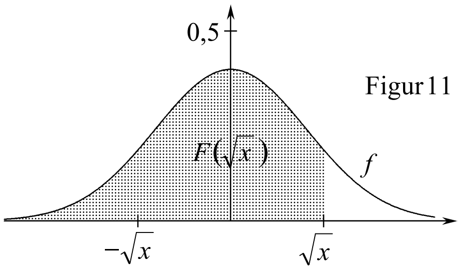
$$p = \int_{3,60}^{\infty} g(x)\,dx = 0{,}058 \ .$$

Her har vi brugt forskriften for g som står i linje (8) nederst i afsnit 12 .

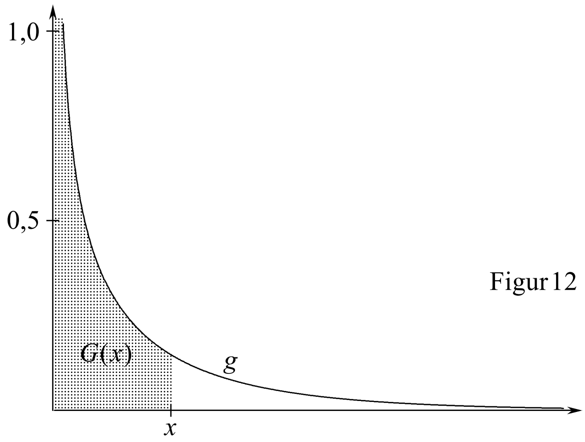
12 Forskrift for *g*.

f og g er funktionerne fra afsnit 11. Arealfunktionerne for f og g kalder vi F og G, dvs.

$F(x)$ = gråt areal på figur 11 og $G(x)$ = gråt areal på figur 12 .



Figur 11

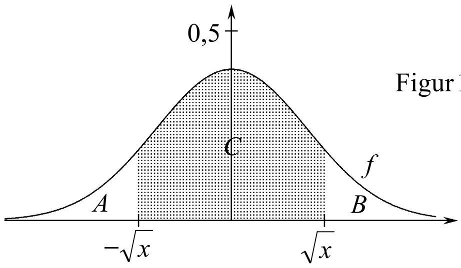


Figur 12

Tallene i intervallet $0 \le t \le x$ stammer fra tallene i intervallet $-\sqrt{x} \le t \le \sqrt{x}$, så

gråt areal på figur 12 = gråt areal på figur 13

dvs.



Figur 13

(1) $G(x) = C$

Af figur 11 og 13 ser vi at $C = F(\sqrt{x}) - A$.
Vi indsætter dette i (1):

(2) $G(x) = F(\sqrt{x}) - A$

Da *f*-grafen er symmetrisk, er $A = B$, så i (2) kan vi erstatte A med B:

(3) $G(x) = F(\sqrt{x}) - B$

Hele arealet under *f*-grafen er 1, så af figur 13 og 11 ser vi at $B = 1 - F(\sqrt{x})$.
Dette indsætter vi i (3):

(4) $G(x) = F(\sqrt{x}) - (1 - F(\sqrt{x}))$

Vi reducerer (4) og får:

(5) $G(x) = 2F(\sqrt{x}) - 1$

Da G er arealfunktion for g, er

(6) $g(x) = G'(x)$

Af (5) og (6) får vi

(7) $g(x) = (2F(\sqrt{x}) - 1)'$

Nspire udregner højresiden og får

(8) $g(x) = \dfrac{\sqrt{2} \cdot e^{\frac{-x}{2}}}{2 \cdot \sqrt{\pi \cdot x}}$ *Denne formel skal du IKKE huske!*

I rammen ses hvordan vi taster for at få (8).

Define $f(x)=\dfrac{1}{\sqrt{2 \cdot \pi}} \cdot e^{\frac{-x^2}{2}}$ ▸ *Udført*

fs står for **F**

Define $fs(x)=\int_{-\infty}^{x} f(t)\, dt$ ▸ *Udført*

Define $g(x)=\dfrac{d}{dx}\left(2 \cdot fs(\sqrt{x})-1\right)$ ▸ *Udført*

$g(x) = \dfrac{\sqrt{2} \cdot e^{\frac{-x}{2}}}{2 \cdot \sqrt{\pi \cdot x}}$ ⚠

$p=\int_{3.6}^{\infty} g(x)\, dx = 0.05778$

Stikordsregister