# 4 Sannsynlighet

## 4.1 Simulering av stokastiske forsøk

GeoGebra er et fint verktøy til å simulere ulike stokastiske forsøk. Fordelen her er at du kan gjøre mange forsøk nok så rakst. Du kan også lett behandle resultatene. Vi starter med et velkjent forsøk: vi kaster en terning.

### Eksempel 4.1

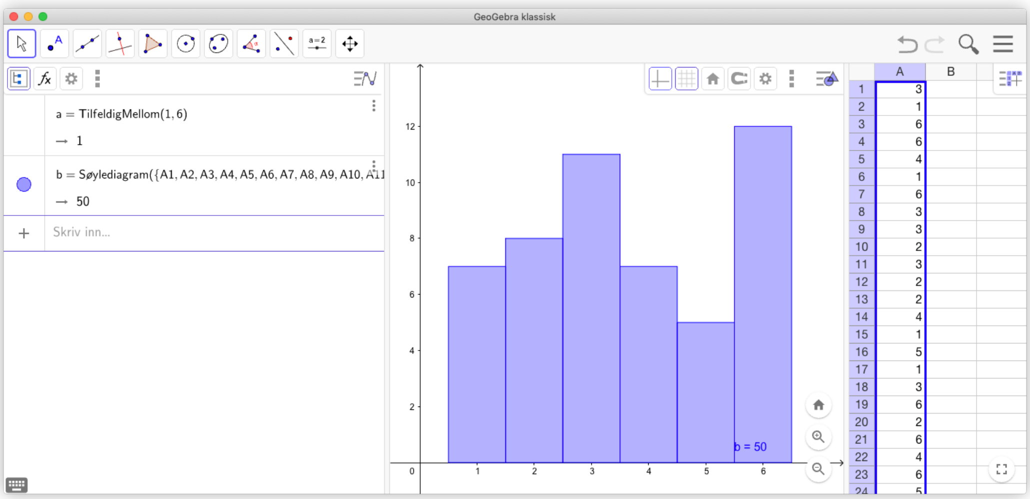
Simuler et terningkastforsøk.

**Løsning:**
Skriv inn kommandoen `TilfeldigMellom(1,6)`. Du får da et tilfeldig tall blant tallene 1, 2, 3, 4, 5 og 6. Ønsker du å gjenta forsøket kan du klikke inn i algebrafeltet (der du skrev `TilfedigMellom(1, 6)`) og trykke enter. Da gjøres utregningen på nytt. Alternativt kan du skrive inn `OppdaterKonstruksjonen()` i algebrafeltet og trykk enter.

I stedet for å gjøre ett kast om gangen, så kan vi bruke regnearket til å gjøre mange kast samtidig.

– Åpne regnearket og skriv inn `=TilfeldigMellom(1,6)` i celle A1. Autokopier så denne cellen ned til og med celle A50.

– Lag søylediagram av disse tallene ved å bruke verktøyet *Analyse av en variabel* til å lage søyledigram som illustrerer utfallene.



**Tips!** Du kan også lage lister med tilfeldige tall ved å bruke kommandoen

```
Følge(TilfeldigMellom(1,6), k, 1, 50)
```

Du får da 50 tilfeldige tall mellom 1 og 6.

Lager du en glider $n$ med for eksempel heltallsverdier fra for 1 til 10000 kan du nå simulere $n$ slike kast ved å skrive inn kommandoen

```
Følge(TilfeldigMellom(1,6), k, 1, n)
```

### Oppgave 4.1

Simuler et forsøk der du kaster 2 terninger 100 ganger og der du ser på summen av antall øyne i hvert kast. Hvordan fordeler resultatet seg på to øyne, tre øyne, ..., 12 øyne?

### Oppgave 4.2

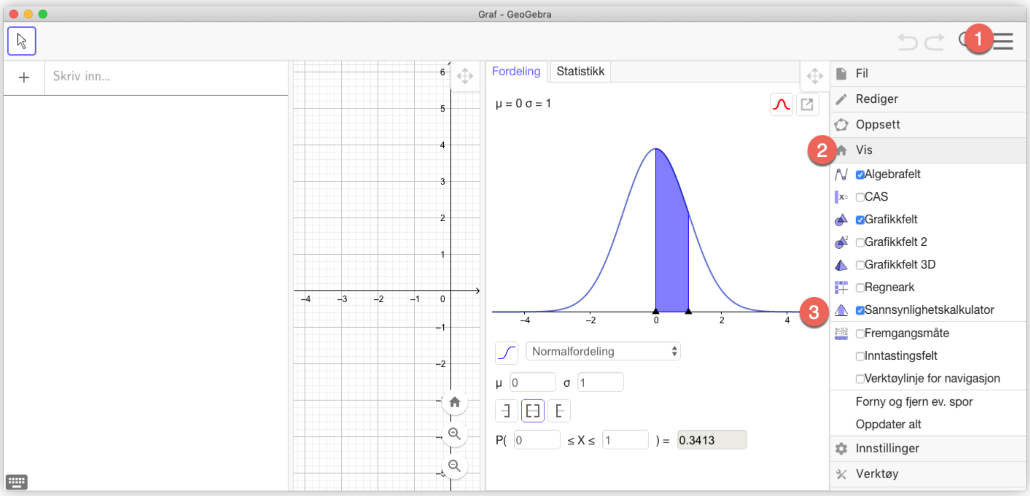
Simuler et forsøk der du kaster to mynter.

**Tips!** Du kan la 0 stå for kron og 1 stå for mynt og bruke kommandoen

```
TilfeldigMellom(0,1)
```

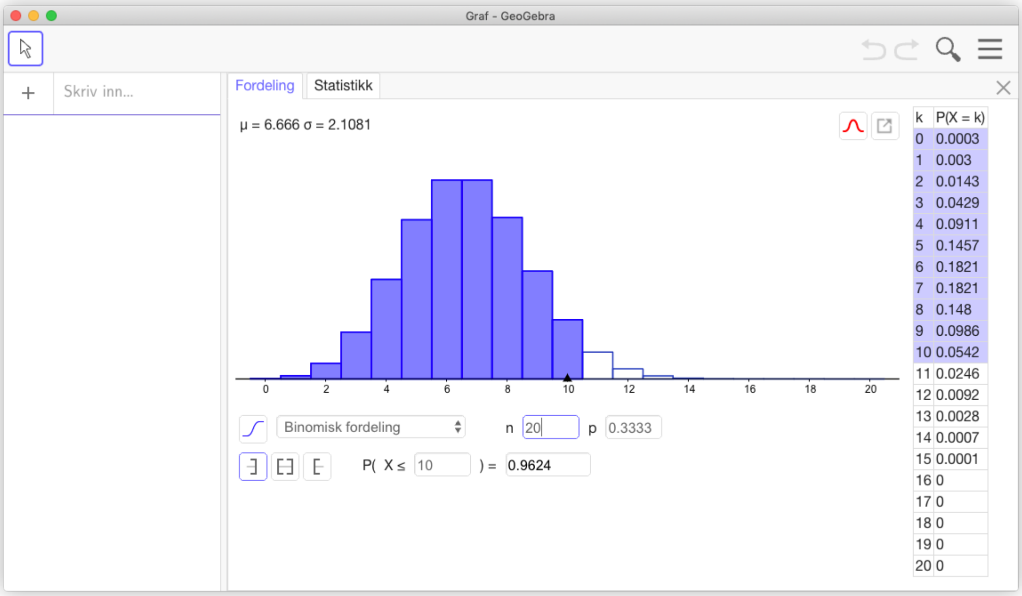
## 4.2 Sannsynlighetskalkulatoren

GeoGebra har et nyttig verktøy for utregning og presentasjon av sannsynlighetsfordelinger. Du finner dette verktøyet under «Vis» på menylinjen:




Figur 4.1: *Sannsynlighetskalkulator* finner du under «Vis» på menylinjen.

Det er et stort utvalg av fordelinger du kan bruke, men i videregående skole er det nok binomisk, hypergeometrisk og normalfordeling som er viktigst. Figur 4.2 vises binomisk fordeling.




Figur 4.2: GeoGebra gjør utregninger og visualiserer svaret fint.

## Eksempel 4.2

Ved en stor videregående skole blir det brukt en nettbasert ressursside. Bruk av ressurssiden forutsetter at hver elev har installert et bestemt program på datamaskinen sin. I klasse 2b fikk 15 av 27 elever hjelp av IKT-seksjonen med installeringen av programmet. Resten av elevene installerte det selv. Det trekkes tilfeldig ut 10 elever i klasse 2b.

a) Finn sannsynligheten for at 6 av de 10 elevene fikk hjelp av IKT-seksjonen.

b) Bestem sannsynligheten for at minst 2 av de 10 elevene installerte programmet selv.

Ved skolen måtte 30 % av alle elevene få hjelp av IKT-seksjonen for å komme inn på ressurssiden.

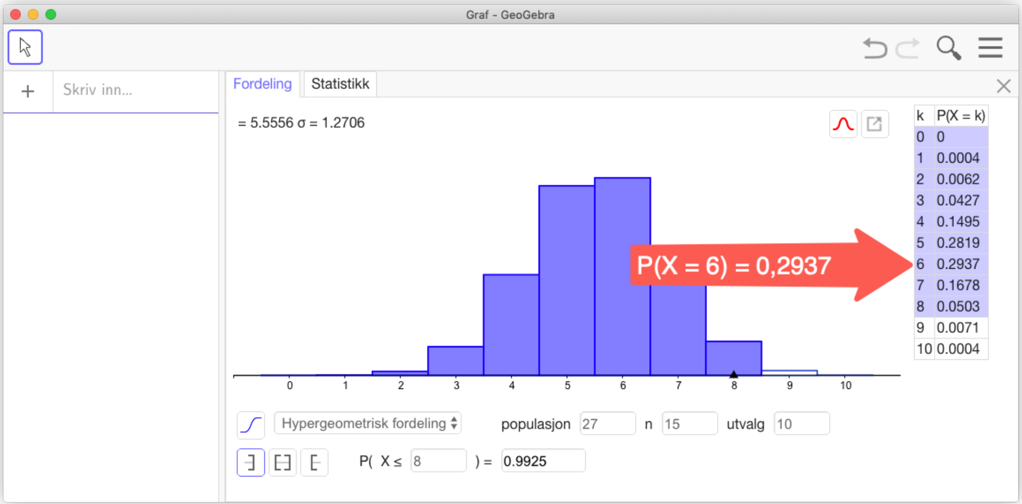
c) Hva er sannsynligheten for at 9 av 24 tilfeldig valgte elever har fått hjelp av IKT-seksjonen? Forklar hvilke forutsetninger du må legge inn for å kunne regne binomisk.

d) Hva er sannsynligheten for at minst 9 av 24 tilfeldig valgte elever har fått hjelp av IKT-seksjonen?

**Løsning:**

a) Vi åpner sannsynlighetskalkulatoren og velger «Hypergeometrisk fordeling» og skriver inn 27 for populasjon, 15 for antall elever som har fått hjelp og 10 for utvalget vi gjør. Vi ser at $P(X = 6) = 0{,}2937$. Se figure 4.3.

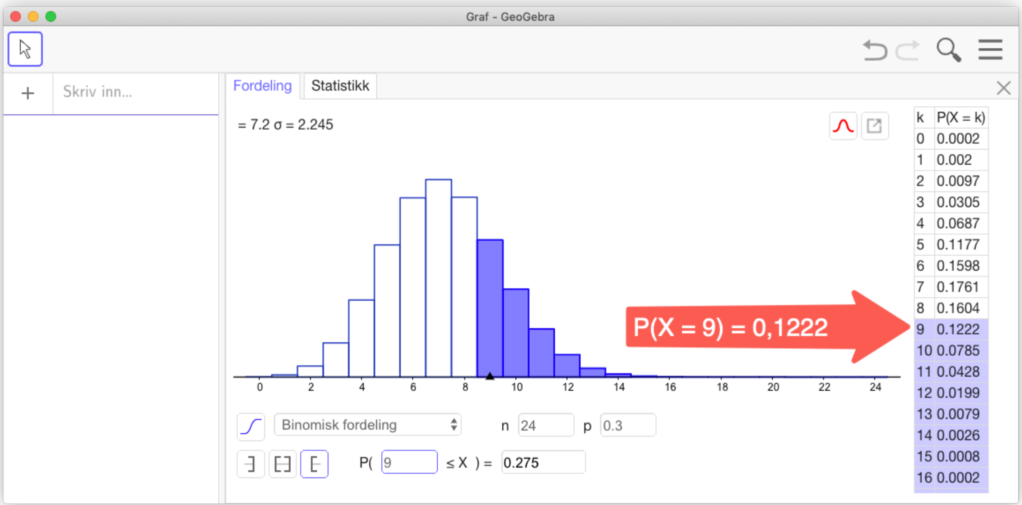
b) Siden $P(\text{minst 2 installerte programmet selv}) = P(\text{høyst 8 fikk hjelp})$, så ser vi at dette blir $P(X \leq 8) = 0{,}9925$. Se figur 4.3.




Figur 4.3: Sannsynlighetskalkulatoren viser at $P(X \leq 8) = 0{,}9925$ og $P(X = 6) = 0{,}2937$

c) Vi kan i dette tilfellet bruke en binomisk sannsynlighetsmodell under forutsetningen at sannsynligheten for at en elev har fått hjelp er 30 %, uavhengig av om de andre elevene som velges ut har fått slik hjelp.

Vi bruker *Sannsynlighetskalkulator* og finner at $P(X = 9) = 0{,}1222$. Det vil si at det er ca 12 % sannsynlig at 9 av 24 tilfeldig valgte elever har fått slik hjelp. Se figur 4.4.




Figur 4.4: Sannsynlighetskalkulatoren viser at $P(X \geq 9) = 0{,}275$ og $P(X = 9) = 0{,}1222$

d) Bruker *Sannsynlighetskalkulator* med binomisk fordeling og finner at $P(X \geq 9) = 0{,}275$. Se figur 4.4.

### Eksempel 4.3

Ved et helsestudio registrerte de kroppsvekten til alle de 320 kundene.

Gjennomsnittsvekten var 79,2 kg med et standardavvik på 6,4 kg. Vi antar at kroppsvekten er normalfordelt.

a) 1) Hvor stor andel av kundene veide mellom 75,0 kg og 85,0 kg?

2) Hvor stor andel av kundene veide over 100,0 kg?

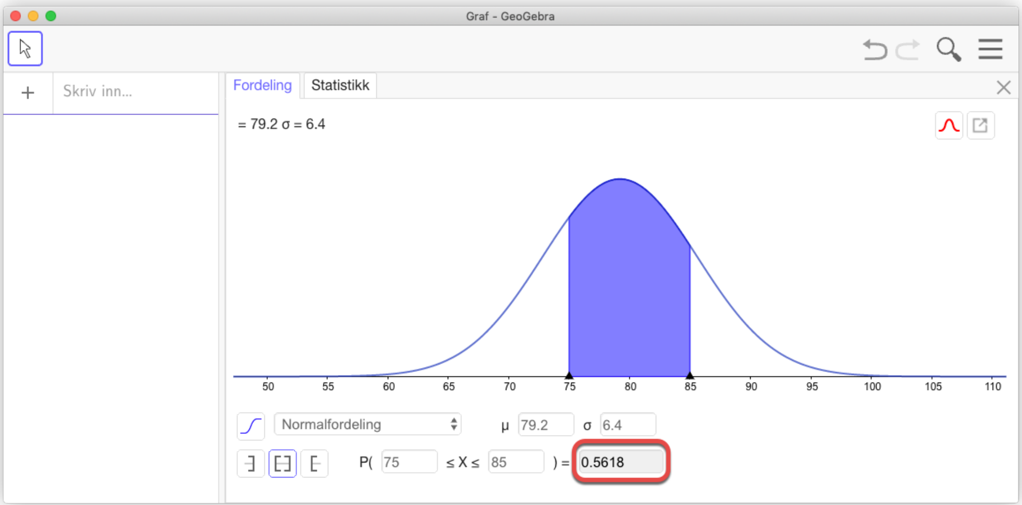
Helsestudioet vil undersøke om treningen påvirker kroppsvekten. De veier derfor 30 tilfeldig valgte kunder etter en periode med jevnlig trening. Gjennomsnittsvekten for disse 30 kundene er 76,0 kg. Vi antar at standardavviket er uendret.

b) Sett opp en nullhypotese og en alternativ hypotese som passer til denne problemstillingen.

c) Undersøk om det er grunnlag for å hevde at gjennomsnittsvekten til kundene i helsestudioet har gått ned. Bruk et signifikansnivå på 5 %.

**Løsning:**

a) 1) Bruker *Sannsynlighetskalkulator* med Normalfordeling som vist på figur 4.5. Vi ser at $P(75{,}0 \le X \le 85{,}0) = 0{,}5618$. Det vil si at 56,2 % av kundene veier mellom 75,0 kg og 85,0 kg. Det vil med andre ord si 180 personer.




Figur 4.5: Vi ser at $P(75{,}0 \le X \le 85{,}0 = 0{,}5618$. Legg merke til hvordan dette visualiseres med Gauss-kurven.

2) Vi bruker *Sannsynlighetskalkulator* og ser at $P(X \ge 100) \approx 0{,}0006$. Dette tilsvarer mye mindre enn 1 person. Vi kan derfor konkludere med at ingen av kundene er mer enn 100 kg.

b) Vi lar $\overline{X}$ være gjennomsnittsvekten[1] til de 30 tilfeldig valgte personene.

$$\text{Nullhypotse } H_0 : \mu_{\overline{X}} = 79,2$$
$$\text{Alternativ hypotese } H_1 : \mu_{\overline{X}} < 79,2$$

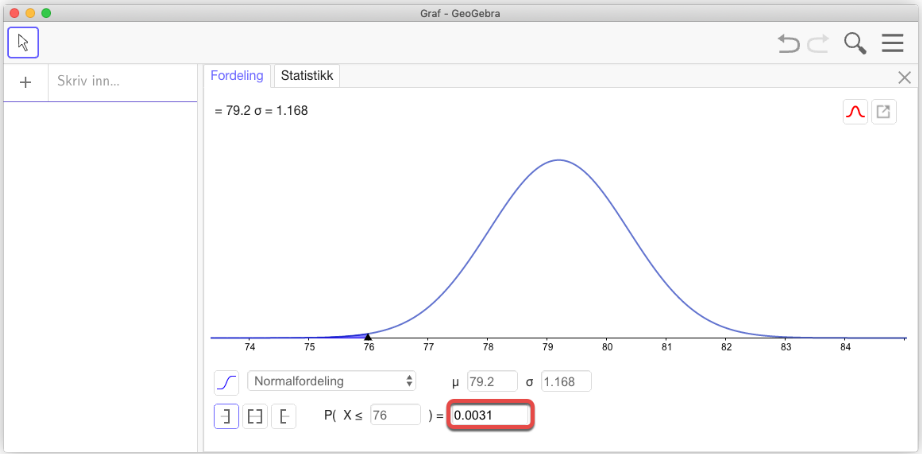
c) Vi antar nullhypotesen gjelder og ønsker å finne ut hva sannsynligheten er for at gjennomsnittsvekten til 30 tilfeldig valgte er mindre eller lik 76,0 kg. Da er forventningsverdien $E(\overline{X}) = 79{,}2$ (siden vi antar nullhypotesen er sann) og standardavviket er

$$SD(\overline{X}) = \frac{6,4}{\sqrt{30}} = 1{,}168$$

Bruker *Sannsynlighetskalkulator* og finner at

$$P(X \leq 76{,}0) = 0{,}0031 = 0{,}31\,\%$$

Vi ser at dette er veldig lite, så innenfor et vanlig signifikansnivå på 5 % kan vi forkaste nullhypotesen. Vi kan med andre ord konkludere med at treningen har virket!




Figur 4.6: *Sannsynlighetskalkulator* viser at $P(\overline{X} \leq 76{,}0) = 0{,}0031$.

---

**Oppgave 4.3 — Eksempelsett S2, oktober 2008.**
Levetiden til en spesiell motor antas å være normalfordelt med en forventningsverdi på 10 år og et standardavvik på 2 år.

a) Finn sannsynligheten for at

1) motoren fungerer mindre enn 8 år

2) motoren fungerer mellom 8 og 11 år

[1] Vi kan også se på variabelen $Y$ = summen av vektene til kundene og bruke at $E(Y) = 30 \cdot \mu = 2376$ og $SD(Y) = \sqrt{30}\sigma = 35{,}05$. Da skal vi regne ut $P(Y \leq 2280)$.

Motorer som blir defekte før garantitiden går ut, blir erstattet av produsenten. Firmaet som produserer motorene, ønsker ikke å erstatte mer enn 3 % av de motorene som blir defekte.

b) Hvor lang garantitid bør de da tilby?

I firmaet er de usikre på om forventet levetid er så lang som 10 år. De registrerer levetiden i antall år på 10 tilfeldig valgte motorer:

8,3 9,2 7,3 10,1 9,5 8,7 8,4 10,0 9,1 9,4

De antar fortsatt at levetiden til motoren er normalfordelt med standardavvik på 2 år.

c) Still opp en nullhypotese $H_0$ og en alternativ hypotese $H_1$ for denne problemstillingen.

d) Velg et signifikansnivå på 5 % og undersøk om firmaet må forkaste hypotesen $H_0$.

Kommentar: GeoGebra hjelper oss ikke til å finne standardavviket i oppgave d). Dersom vi ser på gjennomsnittet $\overline{X}$ av de 10 motorene, så vil

$$E(\overline{X}) = \mu = 10 \quad \text{og} \quad SD(\overline{X}) = \frac{\sigma}{\sqrt{n}} = \frac{2}{\sqrt{10}} \approx 0,63245$$

## 4.3 Statistikk i sannsynlighetskalkulatoren

I Sannsynlighetskalkulator er det en egen flik som heter *Statistikk*. Her byr GeoGebra på en del tester og estimater. En del av disse ligger nok utenfor læreplanmåla i matematikk for fagene i skolen, men kan være interessant å se nærmere på likevel. Unntaket er muligens Matematikk X hvor elevene skal kunne gjøre utvalgsundersøkelser.

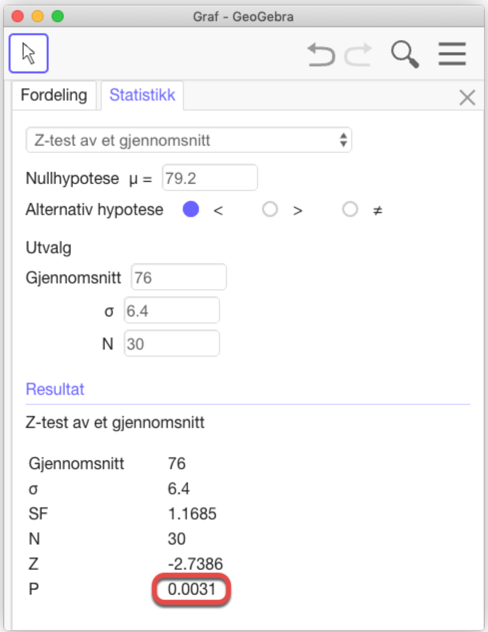
### Eksempel 4.4

Vi viser i dette eksempelet hvordan hypotesetesten i eksempel 4.3 kan løses i GeoGebras *Statistikk*-verktøy.

Vi lar $\overline{X}$ være gjennomsnittsvekten til 30 tilfeldig valgte kunder. Vi setter opp nullhypotese og alternativ hypotese:

$$H_0 : \mu_{\overline{X}} = 79{,}2$$
$$H_1 : \mu_{\overline{X}} < 79{,}2$$

Sentralgrenseteoremet gir at $\overline{X}$ er tilnærmet normalfordelt. Vi bruker et signifikansnivå på 5% og antar nullhypotesen er sann. Vi velger «Z-test av et gjennomsnitt» i *Statistikk* i Sannsynlighetskalkulator og fører inn de kjente tallene. Vi haker av for « < » i alternativ hypotese.




Vi ser at vi får $P = 0{,}0031$. Det vil si at det er 0,31% sannsynlig at gjennomsnittsvekten til 30 tilfeldig valgte medlemmer er mindre eller lik 76 kg. Vi forkaster derfor nullhypotesen og konkluderer med at det er gode grunner til å si at treningen har ført til at gjennomsnittsvekten har gått ned.

Merk at GeoGebra regner ut standarverdien i dette tilfellet. Det vil si at GeoGebra regner ut det som ofte kalles Standard Feil: $\mathrm{SF}(\overline{X}) = \frac{\sigma}{\sqrt{n}}$.

I oppgave 4.3 ble det gjort et utvalg på 10 tilfeldige motorer. I oppgaven står det at vi kan gå ut fra et standardavvik på 2 år. Uten denne opplysningen måtte vi ha gjort et estimat av standardavviket. Vi kan beregne standardavviket til tallene vi fant, nemlig standardavviket til:

$$8{,}3 \quad 9{,}2 \quad 7{,}3 \quad 10{,}1 \quad 9{,}5 \quad 8{,}7 \quad 8{,}4 \quad 10{,}0 \quad 9{,}1 \quad 9{,}4$$

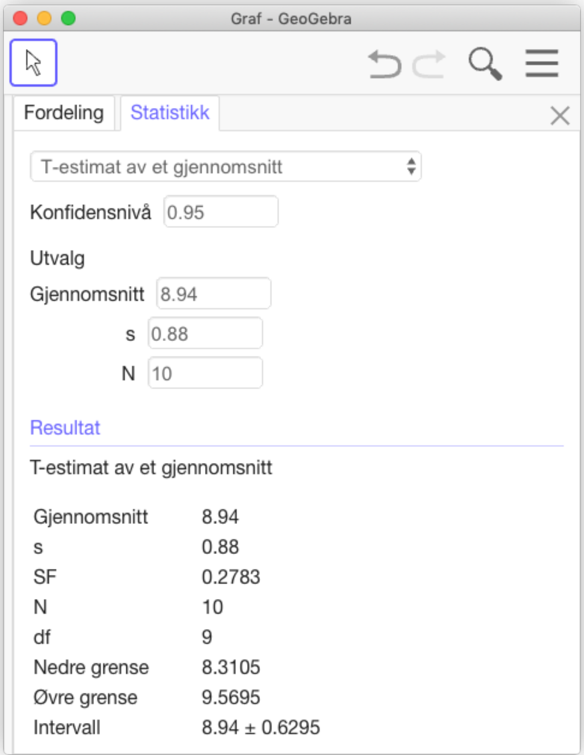
Siden vi i dette tilfelle har et utvalg, må vi bruke det som kalles *utvalgsstandardavvik*. I GeoGebra kan vi regne ut dette ved å bruke kommandoen

```
s=stavvp(8.3, 9.2, 7.3, 10.1, 8.7, 8.4, 10, 9.1, 9.4)
```

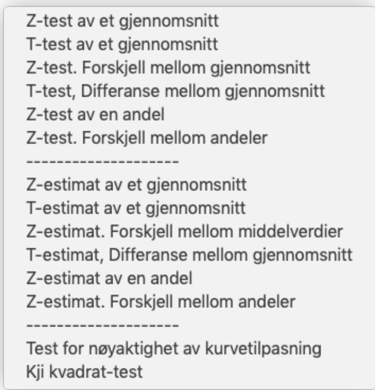
Forskjellen på vanlig standardavvik $\sigma$ og utvalgsstandardavvik $s$ er at vi i førstnevnte har hele populasjonen, mens i sistnevnte har en del av populasjonen. I vårt eksampel blir $s = 0{,}88$. Dette er større enn 0,63245, slik som i oppgaven. Men dette kan skyldes tilfeldige variasjoner, noe som selvsagt også får innvirkning på $s$. Når GeoGebra regner ut standardavvik $\sigma$ og utvalgsstandaravvik $s$ bruker programmet formelene

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 \qquad\qquad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

Dersom vi ikke kjenner standardavviket, må vi ta til gode med å bruke vårt estimat $s$ for standardsavviket. Når vi i slike situasjoner skal gjøre en hypotesetest, kan vi gjøre et T-esimat av et gjennomsnitt. Bruker vi et konfindensnivå på 95% får vi at konfidensintervallet blir [8,31 , 9,57].




Det er flere slike tester og estimater som kan gjøres i GeoGebra. På bildet til nedenfor ser du de som er tilgjengelige i Sannsynlighetskalkulator. Det ligger utenfor målsettingen til denne boken å gå inn på hver av de ulike testene.




Figur 4.7: Det fins flere ferdigprogrammerte tester vi kan bruke i GeoGebra sin Sannsynlighetskalkulator.