



[Vis alt](#)

# Statistiske analyseverktøy

Microsoft Excel inneholder et sett med dataanalyseverktøy kalt Analyseverktøy, som forenkler arbeidet med å utføre komplekse statistiske eller tekniske analyser. Du legger inn data og parametere for hver analyse, og verktøyet bruker passende statistiske eller tekniske makrofunksjoner og viser deretter resultatet i en utdatatabell. Enkelte verktøy genererer diagrammer i tillegg til utdatatabellene.

**Relaterte regnearkfunksjoner** Excel tilbyr mange andre statistiske, økonomiske og tekniske regnearkfunksjoner. Noen av de statistiske funksjonene er innebygd, og andre blir tilgjengelige etter at du har installert analyseverktøyet.

**Tilgang til dataanalyseverktøyene** Analyseverktøyet inneholder verktøyene som er beskrevet nedenfor. Du får tilgang til disse verktøyene ved å klikke **Dataanalyse** på **Verktøy**-menyen. Hvis kommandoen **Dataanalyse** ikke er tilgjengelig, må du laste inn [tilleggsprogrammet](#) Analyseverktøy.

[Anova](#)

Analyseverktøyene Anova inneholder ulike verktøy for å analysere varians. Verktøyet som skal brukes, avhenger av antall faktorer og antall utdrag fra populasjonen du vil teste.

**Anova: En-faktoranalyse** Dette verktøyet utfører en enkel variansanalyse av data for to eller flere utvalg. Analysen tester hypotesen at hvert utvalg er trukket fra samme underliggende sannsynlighetsfordeling, mot den alternative hypotesen at underliggende sannsynlighetsfordelinger ikke er like for alle utvalg. Hvis det bare er to utvalg, kan regnearkfunksjonen TTEST like godt brukes. Hvis det er mer enn to utvalg, finnes det ingen passende generalisering av TTEST, og en-faktoranalysen kan brukes i stedet.

**Anova: To-faktoranalyse med tilbakelegging** Dette analyseverktøyet er nyttig når data kan klassifiseres etter to forskjellige dimensjoner. I et eksperiment for å måle plantehøyde, kan for eksempel plantene ha forskjellige gjødselstyper (for eksempel A, B og C) og også forskjellige temperaturer (for eksempel lav og høy). For hvert av de 6 mulige gjødsel- eller temperaturparene har vi et likt

antall observasjoner av plantehøyde. Ved å bruke dette Anova-verktøyet kan vi teste:

1. Om plantehøyden for de forskjellige gjødselstypene er trukket fra samme underliggende populasjon. Temperaturene ikke tatt hensyn til i denne analysen.
2. Om plantehøyden for de forskjellige temperaturnivåene er trukket fra samme underliggende populasjon. Gjødselstypene blir ikke tatt hensyn til i denne analysen.
3. Om det er gjort rede for virkningene av forskjellene mellom gjødselstypene i trinn 1 og temperaturene i trinn 2, med de 6 utvalgene som representerer alle parverdiene for gjødsel eller temperatur trukket fra samme populasjon. Den alternative hypotesen er at det finnes virkninger som skyldes bestemte gjødsels -eller temperaturpar over og ovenfor forskjellene basert bare på gjødsel eller bare på temperatur.

Inndataområde		
	Gruppe 1	Gruppe 2
Forsøk 1	75	58
	68	56
	71	61
Forsøk 2	75	60
	66	62
	70	60
	68	59
	68	68

**Anova: To-faktoranalyse uten tilbakelegging** Dette analyseverktøyet er nyttig når data klassifiseres etter to forskjellige dimensjoner, som i tilfellet i To-faktoranalyse med tilbakelegging. I dette verktøyet går vi imidlertid ut ifra at det bare finnes én observasjon for hvert par (for eksempel hvert gjødsels- eller temperaturpar i eksemplet ovenfor). Ved å bruke dette verktøyet kan vi bruke testene i trinn 1 og 2 i Anova To-faktoranalyse med tilbakelegging, men vi har ikke nok data til å utføre testen i trinn 3.

## [Korrelasjon](#)

Regnearkfunksjonene KORRELASJON og PEARSON beregner begge korrelasjonskoeffisienten mellom to målingsvariabler når målingene av hver variabel observeres for hvert av N-emnene (når en observasjon for et emne mangler, blir emnet ikke tatt hensyn til i analysen). Korrelasjonsanalyseverktøyet er spesielt nyttig når det er flere enn to målingsvariabler for hvert av N-emnene. Det gis en utdatatabell, en korrelasjonsmatrise som viser KORRELASJON-verdien (eller PEARSON-verdien) som gjelder for hvert mulige

målingsvariabelpar.

Korrelasjonskoeffisienten er som kovariansen en måling av hvor mye to målingsvariabler “varierer sammen”. I motsetning til kovariansen skaleres korrelasjonskoeffisienten slik at verdien er uavhengig av enhetene de to målingsvariablene er uttrykt i (hvis for eksempel de to målingsvariablene er vekt og høyde, forblir korrelasjonskoeffisienten uendret hvis vekten konverteres fra pund til kilogram). Verdien av alle korrelasjonskoeffisienter må være mellom -1 og +1.

Du kan bruke korrelasjonsanalyseverktøyet til å undersøke hvert målingsvariabelpar for å avgjøre om de to målingsvariablene viser samvariasjon, det vil si om store verdier i en variabel henger sammen med store verdier i den andre (positiv korrelasjon), om små verdier i en variabel henger sammen med store verdier i den andre (negativ korrelasjon), eller om verdiene i variablene ikke er relatert til hverandre (korrelasjon nær null).

## [Kovarians](#)

Korrelasjons- og kovariansverktøyene kan begge brukes ved samme innstillinger, når du har observert N forskjellige målingsvariabler for et sett personer. Korrelasjons- og kovariansverktøyene har begge en utdatatabell, en matrise som viser henholdsvis korrelasjonskoeffisienten eller kovariansen mellom hvert målingsvariabelpar. Forskjellen er at korrelasjonskoeffisienter er skalert til å ligge mellom -1 og +1. De tilsvarende kovariansene skaleres ikke. Både korrelasjonskoeffisienten og kovariansen er målinger av hvor mye to variabler “varierer sammen”.

Kovariansverktøyet beregner verdien av regnearkfunksjonen KOVARIANS for hvert målingsvariabelpar (direkte bruk av KOVARIANS i stedet for kovariansverktøyet er et fornuftig alternativ når det bare er to målingsvariabler, det vil si at  $N=2$ ). Oppføringen i diagonalen i kovariansverktøyets utdatatabell i rad  $i$ , kolonne  $i$  er kovariansen målingsvariabelen  $i$ -th med seg selv. Dette er bare populasjonsvariansen for variabelen slik det beregnes i regnearkfunksjonen, VARIANSP.

Du kan bruke kovariansverktøyet til å undersøke hvert målingsvariabelpar for å avgjøre om to målingsvariabler viser samvariasjon, det vil si om store verdier i en variabel henger sammen med store verdier i den andre (positiv kovarians), om

små verdier i en variabel henger sammen med store verdier i den andre (negativ kovarians) eller om verdiene i variablene ikke er relatert til hverandre (kovarians nær null).

## [Deskriptiv statistikk](#)

Analyseverktøyet for deskriptiv statistikk genererer en rapport over univariat statistikk for data i inndataområdet. Denne rapporten gir informasjon om sentraltendens og spredning for dataene.

## [Eksponentiell glatting](#)

Analyseverktøyet for eksponentiell glatting predikerer en verdi basert på prognosen for forrige periode, justert for feil i den forrige prognosen. Dette verktøyet bruker glattingskonstanten  $a$ . Størrelsen på denne konstanten avgjør i hvilken grad prognoser reagerer på feil i forrige prognose.

**Obs!** Verdier mellom 0,2 og 0,3 er fornuftige glattingskonstanter. Disse verdiene indikerer at gjeldende prognose bør justeres med 20 til 30 prosent for feil i forrige prognose. Større konstanter gir raskere svar, men kan gi uberegnelige prognoser. Mindre konstanter kan resultere i lang ventetid på prognoseverdiene.

## [F-test: Verktøy for variansanalyse med to utvalg](#)

F-test: Verktøy for variansanalyse med to utvalg utfører en F-test med to utvalg for å sammenligne variansen i to populasjoner.

Du kan for eksempel bruke F-test-verktøyet for eksempler med tider i svømming for hvert av to lag. Verktøyet gir testresultatet for nullhypotese, det vil si at disse to utvalgene kommer fra fordelinger med lik varians, mot alternativet at variansen ikke er lik i underliggende fordelinger.

Verktøyet beregner verdien  $f$  av F-statistikken (eller F-forhold). En  $f$ -verdi nær opptil 1 beviser at underliggende populasjonsvarianser er like. Hvis  $f < 1$  “P(F  $\leq f$ ) one-tail” i utdatatabellen, er sannsynligheten for å observere en verdi i F-statistikken mindre enn  $f$  når populasjonsvariansene er like, og “F Critical one-tail” gir kritisk verdi mindre enn 1 for valgt signifikansnivå, Alfa. Hvis  $f > 1$ , “P(F  $\leq f$ ) one-tail”, er sannsynligheten for å observere en verdi i F-statistikken

større enn  $f$  når populasjonsvariansene er like og "F Critical one-tail" gir kritisk verdi større enn 1 for Alfa.

## [Fourier-analyse](#)

Fourier-analyseverktøyet løser problemer i lineære systemer og analyserer periodiske data ved at dataene transformeres ved hjelp av metoden "rask Fourier-transformasjon" (Fast Fourier Transform). Dette verktøyet støtter også inverse transformasjoner, der inversen av transformerte data returnerer de opprinnelige dataene.

Inndataområde		Utdatatabell
Tid	Frekvens	
Domene	Domene	
Data	Utdata	
1	3	
1	1.707106769-1.707106769i	
1	-1	
0	0.292893231+0.292893231i	
0	1	

## [Histogram](#)

Analyseverktøyet for histogrammer beregner enkeltfrekvenser og kumulative frekvenser for et celleområde med data og dataintervaller. Dette verktøyet genererer data for antall forekomster av verdier i et datasett.

I en klasse med 20 studenter kan du for eksempel være interessert i fordelingen av karakterer. En histogramtabell presenterer øvre og nedre karaktergrense, og antall karakterer mellom laveste grense og gjeldende grense. Den enkeltkarakter som forekommer hyppigst, utgjør modalverdien for dataene.

## [Glidende gjennomsnitt](#)

Analyseverktøyet for glidende gjennomsnitt projiserer verdier i prognoseperioden, basert på gjennomsnittsverdien av variablene over et bestemt antall foregående perioder. Et glidende gjennomsnitt gir trendinformasjon som et enkelt gjennomsnitt av alle historiske verdier ville skjule. Du kan bruke dette verktøyet til å forutsi salg, til vareopptelling eller andre trender. Hver prognoseverdi er basert på følgende formel.

$$F_{(t+1)} = \frac{1}{N} \sum_{j=1}^N A_{t-j+1}$$

der:

- $N$  er antall tidligere perioder du vil ha med i det glidende gjennomsnittet
- $A_j$  er den faktiske verdien på tidspunkt  $j$
- $F_j$  er den prognostiserte verdien på tidspunkt  $j$

### [Generering av tilfeldige tall](#)

Analyseverktøyet for generering av tilfeldige tall fyller et område med uavhengige, tilfeldige tall, trukket fra en eller flere fordelinger. Du kan beskrive elementer i en populasjon med en sannsynlighetsfordeling.

Du kan for eksempel bruke en normalfordeling til å beskrive høyden på individene i en populasjon, eller bruke en Bernoulli-fordeling med to mulige resultater til å simulere en populasjon av mynt- og kronekast.

### [Rangering og persentil](#)

Analyseverktøyet for rangering og persentil lager en tabell med ordenstallrangering og prosentdelrangering for hver verdi i et datasett. Du kan analysere den relative rangeringen for verdier i et datasett. Dette verktøyet bruker regnearkfunksjonene RANG og PROSENTDEL. RANG gjør ikke rede for uavgjorte verdier. Hvis du ønsker å gjøre rede for uavgjorte verdier, bruker du regnearkfunksjonen RANG sammen med korrigeringsfaktoren foreslått i hjelpefilen for RANG.

### [Regresjon](#)

Regresjonsanalyseverktøyet utfører lineær regresjonsanalyse ved hjelp av "minste kvadraters metode" til å tilpasse en linje til et sett observasjoner. Du kan analysere hvordan en enkelt variabel påvirkes av verdiene til en eller flere uavhengige variabler.

Du kan for eksempel analysere hvordan ytelsen til en idrettsutøver påvirkes av faktorer som alder, høyde og vekt. Du kan knytte deler av prestasjonene til hver av de tre faktorene, basert på et sett prestasjonsdata, og deretter bruke resultatet til å predikere prestasjonene til en ny idrettsutøver som ikke har vært testet.

Regresjonsverktøyet bruker regnearkfunksjonen RETTLINJE.

## Utvalg

Utvalgsverktøyet lager et utvalg fra en populasjon ved å behandle inndataområdet som en populasjon. Når populasjonen er for stor til at du kan behandle eller lage diagram av det, kan du bruke et representativt utvalg. Du kan også lage et utvalg som bare inneholder verdier fra en bestemt del av en syklus, hvis du tror at inndataene er periodiske.

Hvis for eksempel inndataområdet inneholder salgstall for et kvartal, blir utvalgsverdier fra samme kvartal lagt inn i utdataområdet.

## T-test

Analyseverktøyene i T-test med to utvalg tester likhet i populasjonsgjennomsnittet som ligger under hvert utvalg. De tre verktøyene har forskjellige forutsetninger: at populasjonsvariansene er like, at populasjonsvariansene ikke er like, og at de to utvalgene representerer observasjoner av de samme elementene før behandling og etter behandling.

For alle de tre verktøyene nedenfor beregnes en t-statistikkverdi,  $t$ , som vises som “t Stat” i utdatatabellene. Avhengig av dataene kan denne t-verdien være negativ eller ikke-negativ. Under forutsetningen av at de underliggende populasjonene har samme gjennomsnitt, hvis  $t < 0$ , “P(T  $\leq$  t) one-tail”, er sannsynligheten for at en verdi i t-statistikken vil bli observert som si være mer negativ enn  $t$ . Hvis  $t \geq 0$ , “P(T  $\leq$  t) one-tail”, er sannsynligheten for at en verdi i t-statistikken vil bli observert som er mer positiv enn  $t$ . “t Critical one-tail” gir en avkuttet verdi, slik at sannsynligheten for å observere en verdi i t-statistikken større enn eller lik “t Critical one-tail”, er Alfa.

I “P(T  $\leq$  t) two-tail” er sannsynligheten for at en verdi i t-statistikken som er større i absolutt verdi enn i  $t$ , vil observeres. “P Critical two-tail” gir en avkuttet verdi, slik at sannsynligheten for en observert t-statistikk som er større i absolutt verdi enn “P Critical two-tail”, er Alfa.

**T-test: To utvalg med antatt like varianser** Dette analyseverktøyet utfører en t-test med to utvalg av studenter. I denne typen t-test blir det antatt at de to datasettene kom fra fordelinger med like varianser. Testen kalles en homoskedastisk t-test. Du kan bruke denne t-testen til å avgjøre om det er sannsynlig at de to utvalgene kom fra fordelinger der populasjonen har samme



gjennomsnitt.

**T-test: To utvalg med antatt ulike varianser** Dette analyseverktøyet utfører en t-test for to utvalg av studenter. I denne typen t-test antas det at de to datasettene kom fra fordelinger med ulike varianser. Testen kalles en heteroskedastisk t-test. Slik som i tilfellet med lik varians ovenfor, kan du bruke denne t-testen til å avgjøre om det er sannsynlig at de to utvalgene har kommet fra fordelinger der populasjonen har samme gjennomsnitt. Bruk denne testen når det er forskjellige elementer i de to utvalgene. Bruk den parvise testen som er beskrevet under, når det er et enkelt sett elementer og de to utvalgene representerer målinger for hvert element før og etter en behandling.

Formelen nedenfor brukes til å finne den statistiske verdien  $t$ .

$$t' = \frac{\bar{x} - \bar{y} - \Delta_0}{\sqrt{\frac{S_1^2}{m} + \frac{S_2^2}{n}}}$$

Følgende formel brukes til å beregne graden av frihet, df. Fordi resultatet av beregningen vanligvis ikke er et helt tall, rundes df-verdien av til nærmeste heltall for å få en kritisk verdi fra t-tabellen. Regnearkfunksjonen i Excel, TTEST, bruker den beregnede df-verdien uten å runde av, siden det er mulig å beregne en verdi for TTEST uten en df med heltall. På grunn av disse forskjellige måtene å bestemme grad av frihet på, vil resultatene i TTEST og dette t-test-verktøyet være forskjellige fra tilfellet med ulik varians.

$$df = \frac{\left(\frac{S_1^2}{m} + \frac{S_2^2}{n}\right)^2}{\frac{(S_1^2/m)^2}{m-1} + \frac{(S_2^2/n)^2}{n-1}}$$

**T-test: Verktøy for analyse av gjennomsnitt for to parvise utvalg.** Du kan bruke en parvis test når det dannes naturlige observasjonspar i utvalgene, for eksempel når en utvalgsgruppe testes to ganger, før og etter et eksperiment. Dette analyseverktøyet og formelen utfører en partest med to utvalg for studenter for å avgjøre om observasjoner før en behandling og observasjoner etter en behandling kan komme fra fordelinger der populasjonen har samme gjennomsnitt. I denne t-testen antas det ikke at variansen i populasjonene er like.

**Obs!** Blant resultatene som genereres av dette verktøyet er gruppevarians, et samlet mål på spredningen av data rundt gjennomsnittet, utledet fra formelen nedenfor.

$$S^2 = \frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2}$$

## [Z-test](#)

Z-Test: Verktøy for analyse av gjennomsnitt for to parvise utvalg utfører en z-test for to utvalg for gjennomsnitt med kjente varianser. Dette verktøyet brukes til å teste nullhypotesen om at det ikke er noen forskjell mellom populasjonsgjennomsnitt, enten mot en ensidet eller en tosidet alternativ hypotese. Hvis variansene er kjent, bør regnearkfunksjonen ZTEST brukes i stedet.

Ved bruk av Z-Test-verktøyet bør man være nøye på at man forstår utdataene. “P(Z <= z) one-tail” er egentlig P(Z >= ABS(z)), sannsynligheten for en z-verdi lengre fra 0 i samme retning som den observerte z-verdien når det ikke er noen forskjell mellom populasjonsgjennomsnittene. “P(Z <= z) two-tail” er egentlig P(Z >= ABS(z) or Z <= -ABS(z)), sannsynligheten for en z-verdi lengre fra 0 i hvilken som helst retning enn den observerte z-verdien når det ikke er noen forskjell mellom populasjonsgjennomsnittene. Det tosidige resultatet er bare det ensidige resultatet multiplisert med 2. Z-Test-verktøyet kan også brukes i tilfellet der nullhypotesen er at det er en bestemt ikke-null-verdi for forskjellen mellom de to populasjonsgjennomsnittene.

Du kan for eksempel bruke denne testen til å finne forskjellen i ytelse mellom to bilmodeller.



[Vis alt](#)

# Utføre en statistisk analyse

1. Velg **Dataanalyse** på **Verktøy**-menyen.

Hvis **Dataanalyse** ikke er tilgjengelig, kan du laste inn Analyseverktøyet.

[Hvordan?](#)

1. Velg **Tillegg** på **Verktøy**-menyen.
2. Merk **Analyseverktøy** i listen **Tilgjengelige tillegg**, og velg deretter **OK**.
3. Følg om nødvendig instruksjonene i installasjonsprogrammet.
2. I dialogboksen **Dataanalyse** merker du navnet på analyseverktøyet du vil bruke, og velger deretter **OK**.
3. Angi de analysealternativene du vil bruke, i dialogboksen til det valgte analyseverktøyet.

Du kan bruke **Hjelp**-knappen i dialogboksen hvis du vil ha mer informasjon om alternativene.

# Bibliografi over statistiske metoder og algoritmer

Følgende bok gir detaljert informasjon om algoritmer som er brukt til å lage analyseverktøy og -funksjoner i Microsoft Excel.

- Strum, Robert D. & Donald E. Kirk. *First Principles of Discrete Systems and Digital Signal Processing*. Reading, Mass.: Addison-Wesley Publishing Company, 1988.

Følgende bøker gir detaljert informasjon om statistiske metoder eller algoritmer som er brukt til å lage analyseverktøy og -funksjoner i Microsoft Excel.

- Abramowitz, Milton & Irene A. Stegun, eds. *Handbook of Mathematical Functions, with Formulas, Graphs, and Mathematical Tables*. Washington, D.C.: U.S. Government Printing Office, 1972.
- Box, George E.P., William G. Hunter & J. Stuart Hunter. *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building*. New York: John Wiley and Sons, 1978.
- Devore, Jay L. *Probability and Statistics for Engineering and the Sciences*. 4th ed. Wadsworth Publishing, 1995.
- McCall, Robert B. *Fundamental Statistics for the Behavioral Sciences*. 5th ed. New York: Harcourt Brace Jovanovich, 1990.
- Press, William H., Saul A. Teukolsky, William T. Vetterling & Brian P. Flannery. *Numerical Recipes in C: The Art of Scientific Computing*. 2nd ed. New York: Cambridge University Press, 1992.
- Sokal, Robert R. & F. James Rohlf. *Biometry: The Principles and Practice of Statistics in Biological Research*. 2<sup>nd</sup> ed. New York: W. H. Freeman, 1995.

# Feilsøking i forbindelse med dataanalyse

Gjelder verktøy i tillegget Analyseverktøy.

## **Data vises bare på det første regnearket**

Dataanalysefunksjonene kan bare brukes i ett regneark om gangen. Når du utfører dataanalyser i grupperte regneark, vises resultatet i det første regnearket, mens tomme, formaterte tabeller blir vist i resten av regnearkene. Når du skal utføre dataanalyser i resten av regnearkene, må du omberegne analyseverktøyet for hvert regneark.