

---

ANVÄNDBARA MATLABKOMMANDON I KURSEN  
MATEMATISK STATISTIK FÖR B, K, N, BME OCH KEMISTER; FMSF70 &  
MASB02

---

## Statistiska kommandon

Förutom de 'vanliga' funktionerna i MATLAB finns diverse 'verktyglådor' att köpa till. Med dessa 'specialverktyg' för olika tillämpningsområden ökar MATLABs användbarhet ytterligare. Med *Statistics Toolbox* kan du exempelvis generera slumpstal, beräkna fördelningsfunktioner, sannolikhetsfunktioner och kvantiler för diverse olika fördelningar. Genom kommandot `help stats` får du en lista över de funktioner som ingår i Statistics Toolbox. Använd sedan hjälpfunktionen `help` för att ta reda på in- och utparametrar m.m. Om du t.ex. skulle vara intresserad av funktionen `norminv`, som finns i listan, får du veta mera genom kommandot `help norminv`. Några viktiga kommandon finns sammanfattade i följande tabell:

---

Fördelning	Slumpstal	Fördelningsfunktion	Sannolikhetsfunktion	Kvantil
Normal	<code>normrnd</code>	<code>normcdf</code>	<code>normpdf</code>	<code>norminv</code>
Likformig	<code>unifrnd</code>	<code>unifcdf</code>	<code>unifpdf</code>	<code>unifinv</code>
Binomial	<code>binornd</code>	<code>binocdf</code>	<code>binopdf</code>	<code>binoinv</code>
Hypergeometrisk	<code>hygernd</code>	<code>hygecdf</code>	<code>hygepdf</code>	<code>hygeinv</code>
Poisson	<code>poissrnd</code>	<code>poisscdf</code>	<code>poisspdf</code>	<code>poissinv</code>
Exponential	<code>exprnd</code>	<code>expcdf</code>	<code>exppdf</code>	<code>expinv</code>
Gamma	<code>gamrnd</code>	<code>gamcdf</code>	<code>gampdf</code>	<code>gaminv</code>
Weibull	<code>weibrnd</code>	<code>weibcdf</code>	<code>weibpdf</code>	<code>weibinv</code>

---

## Användbara Matlabkommandon i kursen

**För mer fullständig information se `help` för respektive kommando!**

### Generella kommandon

- `load filnamn` – laddar in filen `filnamn` till Matlabs arbetsområde
- `save filnamn v1 v2 ...` – sparar variablerna `v1`, `v2`, etc. i filen `filnamn.mat`
- `whos` – ger aktuella variabler i Matlabs arbetsområde
- `clear` – rensar bort alla variabler ur arbetsområdet

**Skapa variabler**

- `x=[2 8 5 12]` – genererar en radvektor med talen 2, 8, 5 och 12
- `x=[1:100]` – genererar en radvektor med talen 1, 2, 3, . . . , 100
- `zeros(m,n)` – ger en  $m \times n$  matris av enbart nollor
- `ones(m,n)` – ger en  $m \times n$  matris av enbart ettor
- `size(X)` – ger storleken på matrisen  $X$
- `length(x)` – ger storleken på vektorn  $x$

**Skapa delmatriser**

- `X(:,j)` – ger den  $j$ :te kolonnen i  $X$
- `X(i,:)` – ger den  $i$ :te raden i  $X$
- `y=x(villkor)` – ger de värden på  $x$  som uppfyller villkor, vilket kan t ex vara av formen  $>$ ,  $<$ ,  $=$ . Ex: `I y=x(x>=0)` tilldelas  $y$  alla  $x$ -värden  $\geq 0$ .

**Deskriptiv statistik**

- `sum(X)` – ger en radvektor vars element är de olika kolonnsummorna i  $X$
- `sum(r.*r)` – ger kvadratsumman av elementen i vektorn  $r$  (bra då man vill beräkna  $\sigma$ -skattning ur residualer i linjär regression)
- `mean(X)` – ger en radvektor innehållande aritmetiska medelvärdena för varje kolonn i matrisen  $X$
- `std(X)` – ger en radvektor innehållande standardavvikelserna för varje kolonn i matrisen  $X$
- `cov(X)` – beräknar kovariansmatrisen för variablerna som är kolonner i  $X$
- `corrcoef(X)` – beräknar korrelationsmatrisen för variablerna som är kolonner i  $X$
- `normplot(x)` – ritar ut datamaterialet i  $x$  på ett normalfördelningspapper
- `hist(x,n)` – ritar ut ett  $n$ -intervalls histogram för elementen i vektorn  $x$  (utelämnas  $n$  blir antalet intervall 10)
- `histfit(x,n)` – ritar ut ett  $n$ -intervalls histogram för elementen i vektorn  $x$ , samtidigt ritas en normalfördelning med parametrar  $\text{mean}(x)$  och  $\text{std}(x)^2$  i figuren

**Fördelningar**

- `exprnd(a,m,n)` – ger en  $m \times n$  matris av exponentialfördelade slumpstal med parameter  $a$
- `rand(m,n)` – ger en  $m \times n$  matris av rektangelfördelade slumpstal i intervallet (0,1)
- `normrnd( $\mu, \sigma, m, n$ )` – ger en  $m \times n$  matris av normalfördelade slumpstal från  $N(\mu, \sigma^2)$

- `normpdf(x, μ, σ)` – beräknar täthetsfunktionen av en normalfördelning med väntevärde  $\mu$  och varians  $\sigma^2$  i punkten  $x$  (om  $x$  är en vektor beräknas täthetsfunktionen för samtliga element)
- `normcdf(x, μ, σ)` – beräknar  $P(X \leq x)$  då  $X \in N(\mu, \sigma^2)$

### Rita figurer

- `hold on` – håller kvar aktuellt grafikfönster så att man kan rita in flera figurer i samma fönster
- `hold off` – avslutar kvarhållningen av grafikfönster
- `subplot(m, n, p)` – där  $m$ ,  $n$  och  $p$  är ensiffriga heltal, delar upp grafikfönstret i ett  $m \times n$  nät av små grafikfönster och väljer ut fönster nr  $p$  som aktuellt ritfönster. Delfönstren numreras från vänster till höger, uppifrån och ned.
- `subplot` – återställer det normala fallet, det vill säga ett odelat fönster.
- `axis([x_min x_max y_min y_max])` – sätter axlarnas skalor
- `plot(x, y, '*')` – ger ett tvådimensionellt diagram med  $*$  som tecken (man kan t ex välja på  $., *, o, x, +$ )
- `plot(x(villkor), y(villkor), '*')` – plottar  $x$  mot  $y$  men tar endast med de talpar för vilka villkor gäller. Ex: `plot(x(x>0), y(x>0), '*')` ritar enbart ut de talpar för vilka  $x > 0$ .
- `plot3(x, y, z)` – ger en kurva genom punkterna som definieras av  $(x_i, y_i, z_i)$ . Vektorerna  $x$ ,  $y$  och  $z$  måste ha samma längd.
- `plot3(x, y, z, '*')` – ritar som föregående men ger  $*$  i stället för heldragen linje.

### Regressionskommando

- `[b, bint, r, rint, stats]=regress(y, X, α)` – utför regressionsanalys där  $y$  är beroende variabel (responsvariabel),  $X$  är designmatrisen som innehåller de oberoende  $x$ -variablerna som kolonner och  $\alpha$  är konfidensgraden i konfidensintervallen `bint`. Observera att första kolonnen i  $X$  måste vara en kolonnvektor av ettor (kan fås genom `ones(m, 1)` där  $m$  är antalet observationer av  $y$ ). Som utargument fås  $b$  – skattningar av  $\beta$ -parametrarna, `bint` – konfidensintervall för  $\beta$ -parametrarna,  $r$  – residualerna, `rint` – konfidensintervall för residualer samt `stats` som bl a innehåller  $R^2$  – ”förklaringsgraden”
- `reggui(x, y)` – specialskriften `m`-fil som utför enkel linjär regression och polynomregression, där  $y$  är beroende variabel av  $x$ . Grafisk beskrivning av residualer, `normplot`, prediktionintervall och kalibreringsintervall. Ingen designmatris behövs.

### Några olika tester

- `[muhat, sigmahat, mucic, sigmacic]=normfit(x, alpha)` –  $x$  antas komma från en normalfördelning med väntevärde  $\mu$  och standardavvikelse  $\sigma$ . Väntevärdet skattas med `muhat` medan `mucic` anger ett konfidensintervall för  $\mu$  med konfidensgrad  $1-\alpha$ . Motsvarande storheter för normalfördelningens standardavvikelse  $\sigma$  anges i `sigmahat` och `sigmacic`.

- `[h sig ci]=ttest(x,M,alpha,tail)` – testar med ett "t-test" att väntevärdet är M på signifikansnivå alpha. Här antas att x är ett datamaterial från en normalfördelning. Mothypotesen kan vara olika beroende på värdet på tail. Utparametern ci ger ett tvåsidigt konfidensintervall för väntevärdet, h är 1 eller 0 beroende om nollhypotesen förkastas eller ej och sig är testets "P-värde".
- `[h sig ci]=ttest2(x,y,alpha,tail)` – testar med ett "t-test" att väntevärdena i två fördelningar är lika på signifikansnivå alpha. Här antas att x är ett datamaterial från  $N(\mu_1, \sigma^2)$  och y från  $N(\mu_2, \sigma^2)$ . Mothypotesen kan vara olika beroende på värdet på tail. Utparametern ci ger ett tvåsidigt konfidensintervall för skillnaden i väntevärden, h är 1 eller 0 beroende om nollhypotesen förkastas eller ej och sig är testets "P-värde".
- `styrka(sigma,n,alpha,ts)` – ritar styrkefunktionen vid hypotestest av  $\mu=\mu_0$  där  $\mu$  är väntevärdet i en normalfördelning med standardavvikelse *sigma*, n är antalet observationer som testet baserar sig på och *alpha* är den valda signifikansnivån. Mothypotesen kan väljas olika beroende på värdet på *ts*
- `anova1(X)` – gör en ensidig variansanalys mellan de variabler som utgör kolonner i X (här måste alla kolonner vara lika långa vilket de är ett balanserat försök)
- `anova1(X,group)` – gör en ensidig variansanalys som ovan men alla observationer ligger nu i kolonnvektorn X medan variabeln group identifierar vilken behandling (nivå) observationen tillhör (nödvändigt kommando om man har olika antal observationer för olika nivåer, dvs obalanserat försök)
- `anova2(X)` – gör en tvåsidig variansanalys i matrisen X där kolonner respektive rader utgör nivåer för de två faktorerna.

### Faktorförsök och responsytor

- `[mv,sv,effect,effectint,r,stats]= facdesign(y,combine,alpha,code)` – gör beräkningar i  $2^2$ - och  $2^3$ -försök. Observationerna ska ligga i vektorn y i "standardordning" ((1),(a),(b),(ab) om ett  $2^2$ -försök; (1),(a),(b),(ab),(c),(ac),(bc),(abc) om ett  $2^3$ -försök). Vid ett  $2^2$ -försök ska variabeln `combine`=[1, 1, 1, 1], medan vid ett  $2^3$ -försök ska `combine`=[1, 1, 1, 1, 1, 1, 1, 1]. Med inparametern `code` kan mängden utskrift väljas, '11' ger "allt".
- `[b,bint,sigma2,r]=respons(y,X,ind,alpha)` – skattar en responsyta  $y = b_0 + b_1 * x_1 + b_2 * x_2 + b_3 * x_1^2 + b_4 * x_2^2 + b_5 * x_1 * x_2$  där X-matrisen består av tre kolonner: först en med bara ettor, sedan  $x_1$  och  $x_2$ . Med inparametern `ind` kan man välja vilka parametrar som ska skattas i modellen (de andra sätts till 0). För en full polynommodell är alltså `ind`=[1, 1, 1, 1, 1, 1] medan om endast  $b_0$ ,  $b_1$  och  $b_2$  ska skattas används `ind`=[1, 1, 1, 0, 0, 0]