

GSFE Datorövning 2

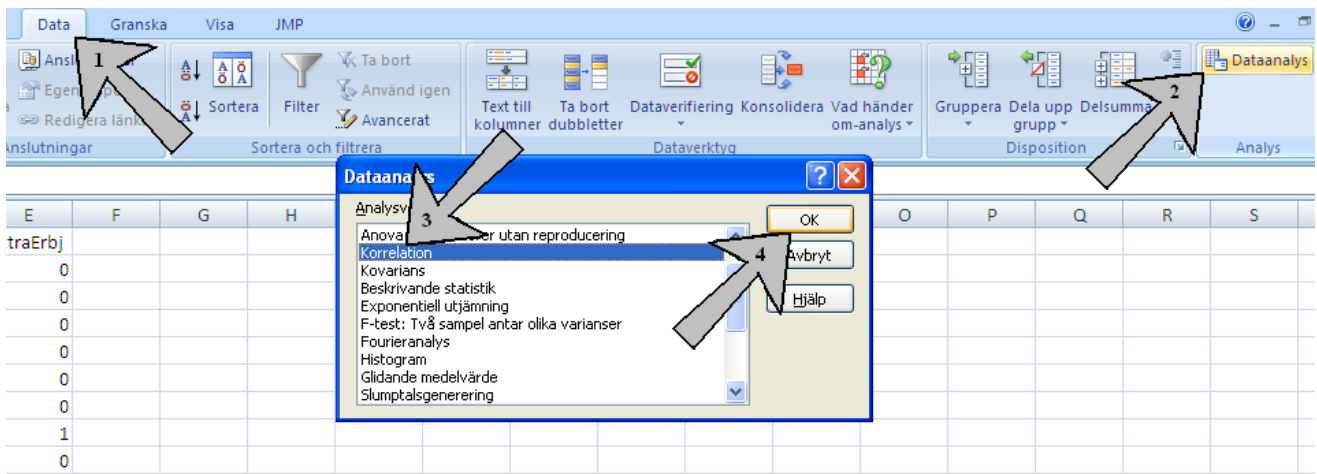
Korrelationsmatris och regression

Korrelationsmatriser i Excel 2007

Excel kan användas till mer än bara visuellt fina diagram och informativa tabeller. Excel kan även användas som statistikprogram för enkla beräkningar¹. Efter att man har samlat sitt datamaterial i ett fint Excel ark kan det vara en bra idé att undersöka sambandet mellan de olika variablerna med spridningsdiagram och korrelationer. Istället för att beräkna ena korrelationen efter den andra kan Excel snabbt och enkelt beräkna och sammanfatta alla korrelationer i en **korrelationsmatris**.

För att kunna skapa en korrelationsmatris måste verktyget 'Analysis ToolPak' vara aktiverat. Se "Utbildningsmaterial till D1" på kurshemsidan för instruktioner på hur verktyget aktiveras.

Gå in på fliken Data och klicka på knappen dataanalys. Hitta "Korrelation" på listan och klicka på OK



¹ Om riktiga statistikprogram önskas rekommenderas SAS, SPSS, Stata, JMP, R och MatLab

Forsätt genom att välja er indataområde (klicka på knappen med röda pilen och markera önskat data). Om ni markerar det streckade området, dvs. variabel namnen, måste ni klicka på "Etiketter på första raden", annars kommer Excel klaga eftersom det inte går att beräkna korrelation på icke-numeriska värden.

OBS! Det går inte att markera flera olika områden, dvs alla variabler som Excel ska beräkna korrelation på måste vara bredvid varandra. Det går t ex inte att markera *Antal* och *ÅterPris*, hoppa över *KonkPris* och sedan ha med *AnnonSK*.

Det är rekommenderat att ni väljer nytt kalkylblad som ni kan namnge som ni vill.

Avsluta genom att klicka på knappen OK.

	A	B	C	D	E
1		<i>Antal</i>	<i>ÅterPris</i>	<i>KonkPris</i>	<i>AnnonSK</i>
2	Antal	1			
3	ÅterPris	-0,171	1		
4	KonkPris	0,743941	0,283266	1	
5	AnnonSK	0,528631	0,331247	0,887674	1

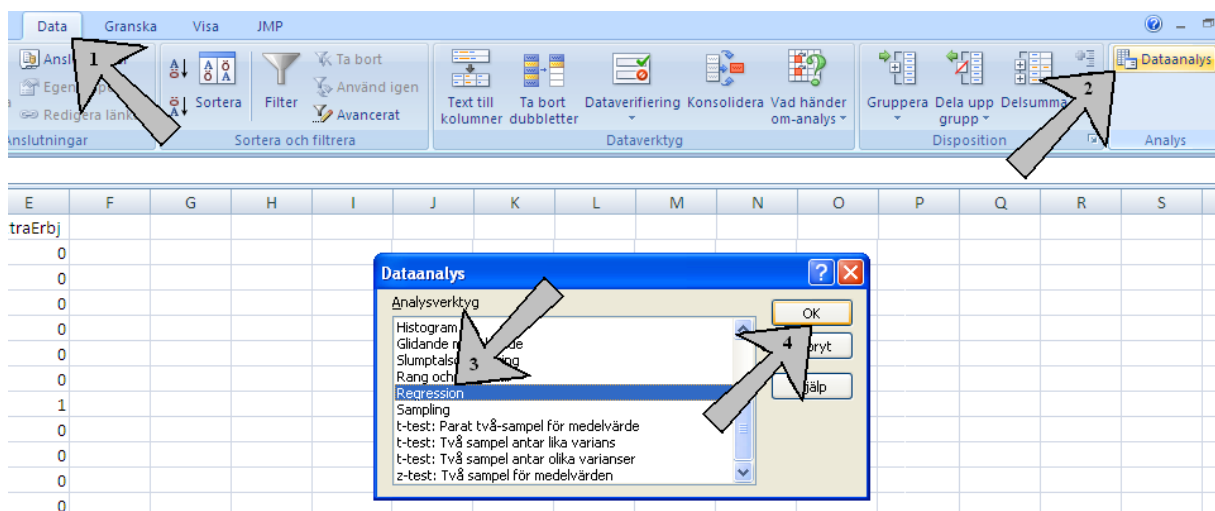
Resultatet ni ska få består av en diagonal rad av 1:or (eftersom korrelation mellan t ex *Antal* och *Antal* är 1 då de är samma), med tomma celler till höger (eftersom matrisen är symmetrisk) och korrelationerna till vänster.

Regressionsanalys i Excel 2007

Korrelationsmatrisen har nu gett oss mått på hur starkt sambandet är mellan de olika variablerna. T ex $\text{Korr}(\text{KonkPris}, \text{Antal}) = 0,743741$ säger oss att när ena ökar, ökar också den andra. Men vi kan inte uttala oss om hur mycket. Inte än. För att få reda på förändringen i *Antal* när *KonkPris* ökar med en enhet kommer vi utföra **regressionsanalys** i Excel.

Glöm inte bort att verktyget 'Analysis ToolPak' måste vara aktiverat för att kunna se knappen dataanalys. Se "Utbildningsmaterial till D1" på kurshemsidan för instruktioner på hur verktyget aktiveras.

Gå in på fliken Data och klicka på knappen dataanalys. Hitta "Regression" på listan och klicka på OK



Börja med att välja Y-indataområde, detta kommer vara er undersökningsvariabel. Sedan väljer ni er X-indataområde, vilket är er/-a förklaringsvariabel/-ler. Väljer ni bara en X-variabel utför ni en enkel linjär regression och om ni väljer flera X-variabler utför ni en

	A	B							
1	UTDATASAMMANFATTNING								
2									
3	Regressionsstatistik								
4	Multipel-R	0,872435338							
5	R-kvadrat	0,761143418							
6	Justerad R-kv	0,722926365							
7	Standardfel	60,49610218							
8	Observationer	30							
9									
10	ANOVA								
11		<i>fg</i>	<i>KvS</i>	<i>Mkv</i>	<i>F</i>	<i>p-värde för F</i>			
12	Regression	4	291557,4072	72889,3518	19,91632942	1,77213E-07			
13	Residual	25	91494,45947	3659,778379					
14	Totalt	29	383051,8667						
15									
16		<i>Koefficienter</i>	<i>Standardfel</i>	<i>t-kvot</i>	<i>p-värde</i>	<i>Nedre 95%</i>	<i>Övre 95%</i>	<i>Nedre 95,0%</i>	<i>Övre 95,0%</i>
17	Konstant	565,8119269	394,0829112	1,435768745	0,163463343	-245,8170149	1377,440869	-245,8170149	1377,440869
18	ÅterPris	-42,32797437	11,83646309	-3,576066099	0,001458206	-66,70562622	-17,95032252	-66,70562622	-17,95032252
19	KonkPris	88,45601975	15,01443814	5,891397264	3,79555E-06	57,53320582	119,3788337	57,53320582	119,3788337
20	AnnonsK	-0,070962168	0,031858343	-2,227428099	0,035154598	-0,136575652	-0,005348683	-0,136575652	-0,005348683
21	ExtraErbj	11,9375077	27,2188699	0,4385747	0,664736511	-44,12080376	67,99581916	-44,12080376	67,99581916
22									

multipl linjär regression. När ni väljer X-indataområde måste ni som tidigare markera med ett fält, det går inte att hoppa över t ex kolumner. Om ni markera variabelnamnen måste ni klicka på "Etiketter" och resterande data ni markerar måste vara siffror. Välj nytt kalkylblad med valfritt namn och avsluta genom att trycka på knappen OK.

Tips: Öka kolumnbredden till 12-15