

## Laboration 1

### i 5B1512, Grundkurs i matematisk statistik för ekonomer

Namn: ..... Elevnummer: .....

Laborationen syftar till att ge information och träning i Excels rutiner för sannolikhetsteori och statistisk behandling av data.

1. Använd en webbläsare för att besöka Statistiska centralbyråns statistikdatabas på webbadress [www.ssd.scb.se](http://www.ssd.scb.se). Under området Finansmarknad finner du uppgifter om bostadsinstitutens tillgångar och skulder.

Använd databasen för att bestämma Stadshypotek AB:s och Länsförsäkringar Hypotek AB:s utlåningar till hushåll i (miljoner) svenska kronor under perioden januari till och med december 2006. Håll ned CONTROL-tangenten samtidigt som du markerar fälten med musen för att välja flera fält.

Ladda ned datamängden som en Excel-fil.

2. Öppna filen i Excel och kopiera de 24 värdena i cellerna D6–O7 till cell B2–C13 i ett nytt ark.

Detaljerade instruktioner: När du öppnat den nedladdade Excel-filen, markera cellerna D6–O7 med musen. Välj Copy under Edit-menyn eller tryck Control+C. Skapa ett nytt ark genom att välja New under File-menyn.

Markera cellen B2 i det nya arket och välj Paste special under Edit-menyn. Se till att alternativet Transpose är markerat.

3. Skriv in "Jan" för januari i cell A2, "Feb" i cell A3, och så vidare till "Dec" i ruta A13. Ange "Stadshypotek" i cell B1 och "Länsförsäkringar" i C1.

### Beskrivande statistik

4. Du skall nu beräkna den totala utlåningen under året för de två instituten i cell B14 (för Stadshypotek) och C14 (Länsförsäkringar)..

Skriv "Totalt" i cell A14. Markera cell B14 och klicka på summaknappen på menyraden (den med ett  $\Sigma$  på). Då läggs formeln =SUM(B2:B13) in i cellen. I cellen C14 kan du istället skriva formeln =SUM(C2:C13).

5. Beräkna där den genomsnittliga månadsvisa utlåningen under 2006 med funktionen AVERAGE. Skriv "Medel" i cell A15 och formeln =AVERAGE(B2:B13) i cell B15. Om man inte vill skriva in formeln kan man använda funktionsknappen (den med  $f(x)$  på). Tryck på knappen och välj Statistical-kategorin. Dubbelklicka på funktionen AVERAGE. Markera sedan med musen de celler med data som skall behandlas, i detta fall B2 till B13.

Beräkna den genomsnittliga månadsutlåningen för Länsförsäkringar i cell C15.

Vilket bolag har den största genomsnittliga utlåningen? .....

6. Skriv "Standardavvikelse" i cell A16. Beräkna i cell B16 stickprovsstandardavvikelsen

$$\sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

för observationerna i cell B2 till B13 med hjälp av funktionen STDEV. Den används på samma sätt som AVERAGE och kan antingen skrivas in direkt eller med hjälp av funktionsknappen.

- Bestäm stickprovsstandardavvikelsen för länsförsäkringar i cell C16.  
 Vilket bolag har den största spridningen? .....
7. Om man begränsar sig till att se datamängden som en population, en totalundersökning av utlåningen till hushåll för det givna bolaget under det givna året, skall man istället beräkna populationsstandardavvikelsen
- $$\sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2}.$$
- Detta gör man med funktionen **STDEVP**. Ange denna i cell B17 för Stadshypotek och C17 för Länsförsäkringar.  
 Populationsstandardavvikelsen för  $n$  observationer är en faktor mindre än stickprovsstandardavvikelsen. Ange denna faktor här .....  
 och verifiera detta genom att beräkna kvoten mellan B17 och B16 eller C17 och C16.
8. Skriv "Varians" i cell A18. Beräkna i cell B18 och B19 stickprovs- respektive populationsvarianserna för månadsutlåningen med funktionerna **VAR** och **VARP**. Gör sedan samma sak för Länsförsäkringar i cell C18 och C19.
9. Vilket bolag har den relativt största variationen? .....  
 Besvara detta genom att beräkna variationskoefficienterna för bolagen i cell B20 (Stadshypotek) och C20 (Länsförsäkringar). Ange variationskoefficienten i procent.  
 Variationskoefficienten för det ena bolaget är en faktor större än den för det andra bolaget. Ange denna faktor. ....  
 Hur påverkas denna faktor av valet av populationsspridning eller stickprovsspridning? .....
10. Ändra dina observationer till miljarder svenska kronor istället för miljoner genom att dividera de  $12 \times 2$  värdena i B2–C13 med 1000.  
 (a) Hur förändrades medelvärdena? .....  
 (b) Hur förändrades standardavvikelserna? .....  
 (c) Hur förändrades varianserna? .....
11. Ett alternativt spridningsmått är variationsbredden (range). Använd funktionerna **max** och **min** för att beräkna variationsbredderna för Stadshypotek och Länsförsäkringar i cell B21 respektive C21. Skriv "Variationsbredd" i A21.
12. Ett alternativt mått på genomsnittlig månadsutlåning är medianen. Beräkna medianerna för dina observationsserier med funktionen **median** i cellerna B22 och C22. Skriv "Median" i A22.
13. Beräkna undre kvartil i månadsutlåningsfördelningen med funktionen **QUARTILE** i cell B23. Funktionen tar två argument; dels en referens till celler med data för vilken kvartilen skall beräknas, dels värdet 1 om den undre kvartilen skall beräknas eller 3 om den övre kvartilen skall beräknas.  
 Beräkna de undre kvartilerna i cell B23 och C23 och de övre kvartilerna i cell B24 och C24. Skriv "Q1" och "Q3" som rubriker i cell A23 respektive A24.  
 Vad beräknas med funktionsanropet **=QUARTILE(B2:B13,2)**? .....
14. Med funktionen **PERCENTILE** kan man även bestämma godtyckliga percentiler. Bestäm den övre 5 procents percentilen, dvs den månadsbelåning som är sådan att utlåningen är större än så endast i 5 procent av fallen. Ange dina svar i B25 och C25. Skriv "5%-percentil" i A25.  
 Vad beräknas med funktionsanropet **=PERCENTILE(B2:B13,0.50)**? .....

15. Rita ett stapeldiagram över den månatliga utlåningen i miljarder kronor genom att markera området A2–C13 och välj Chart från Insert-menyn. Se till att få med månadsbeteckningarna som etiketter på staplarna. Notera att utlåningsbeloppen skiljer sig avsevärt mellan de olika företagen.

16. Ange de relativa utlåningsbeloppen per månad, relativt den totala årsutlåningen.

Markera området B2–C14 och välj Copy eller tryck Control+C. Ställ markören i cell E2 och välj Paste. Du har nu fått en kopia av dina värden. Värdena i E2–E13 skall nu delas med summan i B14. Gör detta genom att markera B14, välj Copy, markera E2–E13 och välj Paste special. Markera alternativet Values och operationen Divide. Detta delar samtliga värden i E2–E13 med innehållet i B14. Notera att värdena i E2–E13 summeras till 1, vilket syns i E14.

Upprepa operationen genom att i F2–F13 ge de relativa månadsutlåningsbeloppen för Länsförsäkringar och summan av dessa i F14.

I vilken månad sker den största relativa utlåningen till hushållen för Stadshypotek? .....

för Länsförsäkringar? .....

### Diskret sannolikhetsfördelning

17. Under en tidsperiod om ett år beskrivs antalet gånger en maskin havererar av fördelningen på ditt datablad.

Skapa ett nytt Excel-blad under menyn File. Skriv "Antal fel" i A1 och "Sannolikhet" i B1. Skriv sedan in i cellerna A2–7 värdena 0,1,...,5 och i cellerna B2–7 motsvarande sannolikheter från ditt datablad.

Summera värdena i cell B2–7 i B8 och kontrollera att summan är 1.

- (a) Illustrera sannolikhetsfördelningen genom att markera cellerna A1–B7 och rita ett stapeldiagram med Chart från Insert-menyn. ....

- (b) Beräkna sannolikheten att maskinen under ett år går sönder fler än 2 gånger. ....

- (c) Du skall nu beräkna populationsmedelvärdet i fördelningen. Lägg i cell C2 formeln =A2\*B2. Markera cellen och välj Copy eller tryck Control+C. Markera sedan cellerna C3–7 och välj Paste eller tryck Control+V. I cell C3 står nu formeln =A3\*B3 och så vidare till C7 där det nu står =A7\*B7. Summera värdena i C2–7 i cell C8. Ange detta genomsnittliga antal fel. ....

Du kan kontrollera rimligheten i ditt svar genom att betrakta figuren.

- (d) Du skall nu beräkna populationsvariansen och populationsstandardavvikelsen i fördelningen.

I cell D2 ange formeln =(A2-\$C\$8)^2. Markera cell D2, välj Copy eller tryck Control+C. Markera sedan cellerna D3–7 och välj Paste eller tryck Control+V. I kolumnen D:s celler står nu avvikelserna från medelvärdet och det är dessa som skall viktas med sannolikheterna i kolumn B.

Lägg i cell E2 formeln =D2\*B2. Markera cellen och välj Copy eller tryck Control+C. Markera sedan cellerna E3–7 och välj Paste eller tryck Control+V. I cell E3 står nu formeln =D3\*B3 och så vidare till E7 där det nu står =D7\*B7. Summera värdena i E2–7 i cell E8. Ange denna populationsvarians  $\sigma^2$  .....

- (e) Beräkna  $\sigma$  i cell E9 med formeln =SQRT(E8) och ange detta värde .....

- (f) Varje gång maskinen havererar kostar det 5000 kronor i reparation och uteblivna intäkter. Du skall nu beräkna den genomsnittliga kostnaden och spridningen i kostnad för maskinhaverier under ett år.

- i. Byt ut värdena (antalet fel) i cell A2–A7 mot motsvarande kostnader, dvs multiplicera var och en av dem med 5000.

- ii. Ange den genomsnittliga kostnaden .....  
och ange hur den förhåller sig till det genomsnittliga antalet fel. ....

- iii. Ange standardavvikelsen för kostnaden .....  
och hur den förhåller sig till standardavvikelsen för antalet fel. ....

## Binomialfördelning

18. Ett flygplan med  $c$  tillgängliga platser trafikerar en given sträcka mellan två orter. Flygbolaget säljer  $n$  biljetter och räknar med att med sannolikhet  $p$  kommer en såld biljett inte att utnyttjas, oberoende av andra sålda biljetter. Genom att välja  $n > c$  överbokar man planet, men på grund av  $p < 1$  kommer planet ofta ändå inte att vara fullt.

Du finner dina värden på  $n$ ,  $p$  och  $c$  på ditt datablad. Antalet biljetter som utnyttjas beskrivs av en binomialfördelning med parametrar  $n$  och  $p$ .

- (a) Vad är det genomsnittliga antalet utnyttjade biljetter. ....
- (b) Vad är standardavvikelsen för antalet utnyttjade biljetter. ....
- (c) Excel-funktionen **BINOMDIST** beräknar binomialsannolikheter och tar fyra argument: **BINOMDIST(x,n,p,b)**.  $n$  och  $p$  är dina parametervärden. Om  $b$  är **FALSE** returnerar funktionen sannolikheten att få värdet  $x$ , om  $b$  är **TRUE** returnerar funktionen summan av sannolikheten upp till och med punkten  $x$ , dvs sannolikheten att få något av värdena  $0, 1, 2, \dots, x$ .  
 Utnyttja **BINOMDIST** för att beräkna sannolikheten att samtliga biljetter utnyttjas. ....  
 Beräkna sannolikheten att fler biljetter utnyttjas än det finns platser. ....
- (d) Varje såld biljett genererar en intäkt på 400 kronor. De resenärer som inte får plats på grund av överbokningen genererar en kostnad på 900 kronor (dvs nettointäkten på en överbokad biljett är  $-500$  kronor). Bestäm den förväntade intäkten som flygpolaget har per resa. ....

## Normalfördelning

19. Betrakta en standardnormalfördelning,  $N(0, 1)$ . Funktionen **NORMSDIST** ger sannolikheten att få högst värdet  $x$ , dvs arean under normalfördelningskurvan för området  $(-\infty, x)$ . Notera att tabell 1 i kurslitteraturen ger arean under normalfördelningskurvan för området  $(0, x)$ .

20. Antag att ränteförändringen i procentenheter under en tidsperiod beskrivs av en  $N(0, 1)$ -fördelning.

Använd funktionen **NORMSDIST** för att beräkna

- (a) sannolikheten att ränteförändringen under tidsperioden inte överstiger 0.52 procentenheter. ....
- (b) sannolikheten att ränteförändringen under tidsperioden inte är mindre än 1.74 procentenheter. ....
- (c) sannolikheten att ränteförändringen under tidsperioden är på intervallet  $(-0.48, 1.23)$  procentenheter. ....

21. Funktionen **NORMSINV** ger inversen till fördelningsfunktionen för standardnormalfördelningen  $N(0, 1)$ . Alltså, för ett värde  $0 < x < 1$  så returnerar **NORMSDIST(NORMSINV(x))** värdet  $x$ .

Använd funktionen för att beräkna 4%-, ....

5%- ....

och 0.2%-percentilerna ....

i standardnormalfördelningen, dvs punkter i fördelningen som överskrids endast med sannolikhet 4, 5 resp 0.2 procent.

22. Antag att den månatliga efterfrågan för en produkt beskrivs av en  $N(\mu, \sigma)$ -fördelning, dvs en normalfördelning med populationsmedelvärde  $\mu = 1200$  enheter och populationsstandardavvikelse  $\sigma = 100$  enheter.

Använd funktionen **NORMDIST** för att beräkna sannolikheten att efterfrågan under kommande månad

- (a) kommer att överstiga 1000 enheter .....
  - (b) kommer att vara mellan 900 och 1100 enheter .....
- 23. Funktionen `NORMINV` ger inversen till fördelningsfunktionen för normalfördelningen med populationsmedelvärde  $\mu$  och populationsstandardavvikelse  $\sigma$ .  
 Använd  $\mu = 1200$  och  $\sigma = 100$  från uppgiften ovan samt funktionen `NORMINV` för att beräkna
  - (a) hur många enheter man kommer att sälja med minst 90 % sannolikhet. ....
  - (b) ett antal enheter så stort att endast med sannolikhet 1 % kommer försäljningen att vara så stor. ....