

Laboration 4

Icke-parametriska metoder, samt korrelation och regression.

Uppgift 1

Vid hypotesprövningar har vi att göra med två typer av fel, varav den ena feltypen är att förkasta en sann nollhypotes. Sannolikheten för att göra detta fel kallar vi för α , signifikansnivån. Genom att ange hur stor denna sannolikhet får vara (vi väljer signifikansnivå) begränsas sannolikheten för att förkasta en sann nollhypotes. Vid signifikansnivån 5% löper vi alltså högst 5% risk att förkasta nollhypotesen när den i själva verket är sann.

Här ska vi simulera fram ca 10-15 kast med ett symmetriskt mynt. Vore kronan helt symmetrisk borde ju hälften av kasten ge klave och hälften av kasten ge krona, men vi vet ju att slumpen kommer att ge oss andra resultat även för ett symmetriskt mynt. Utför denna procedur 50-70 gånger. (Detta görs blixtnabbt av Minitab: **Calc>Random Data>Bernoulli...**)

För att testa ifall myntet är symmetriskt kan vi använda oss av ett teckentest (**Stat>Nonparametrics>1-sample Sign...**). Testa samtliga omgångar med teckentest och ange för vilka utfall vi förkastar H_0 vid en vald signifikansnivå. Beräkna också andelen omgångar då nollhypotesen förkastas (d v s vi hävdar att myntet är osymmetriskt) på ex 10% nivå.

Jämför den valda signifikansnivån och det faktiska resultatet!

Uppgifterna 2-4 görs med hjälp av filen Fastighetsdata.MTW. Filen innehåller uppgifter om 28 fastigheter försålda av en fastighetsmäklare.

C1 = Pris i tusentals kronor

C2 = Tomtyta i m²

C3 = Boyta i m²

C4 = Husets ålder

C4 = Antal rum

C1 är att betrakta som undersökningsvariabel och de övriga som bakgrundsvariabler.

Uppgift 2

Gör en korrelationsmatris (**Stat>Basic Statistics>Correlation...**) och studera korrelationerna dels mellan pris och övriga variabler, men även mellan de övriga sinsemellan. Komplettera med spridningsgram för att se sambandet mellan punktmönster och mått.

Uppgift 3

Välj ut en av bakgrundsvariablerna och testa huruvida den har samband med prisvariabeln med hjälp av Homogenitetstest (**Stat>Tables >Cross Tabulation...**).

Innan du genomför testet måste du klassindela variablerna på lämpligt sätt! (**Data >Code >Numeric to Numeric**). ...och innan du gör det är det lämpligt att skaffa en överblick över respektive variabel så att du gör en rimlig klassindelning. Med tanke på att du endast har 28 observationer kan det inte bli fråga om väldigt många klasser!

Uppgift 4

Gör enkla regressionsmodeller med Pris som beroende variabel (*Respons*) och bakgrundsvariablerna som förklarande variabler (*Predictors*) och skatta dem, en i taget. (**Stat>Regression >Regression...**). Tolka koefficienterna i respektive regressionsskattning. Vad säger de om hur de olika variablerna påverkar priset?

Pröva även att ha mer än en förklarande variabel i modellen. Vad händer med koefficientskattningar, förklaringsgrad etc?