KONFIDENSINTERVALL. (Av Göran Rundqvist, KTH.)

En bra metod som alltid fungerar är att HÄRLEDA konfidensintervallen. Som du kommer att förstå av fortsättningen utgår man då (vanligen) från skattningsvariabeln som efter normering får en känd fördelning. Sedan ställer man upp en lämplig sannolikhet och ordnar om i den så att parametern hamnar mellan två värden (tvåsidigt k.i.). Ensidiga k.i. är ovanligare, för att härleda dem utgår man i den lämpliga sannolikheten från EN olikhet.

För att kunna sköta omordningen ovan rätt måste man kunna hantera vissa olikheter. Vad man behöver veta kan sammanfattas på följande sätt: Om $a < b < c$ gäller och man använder en VÄXANDE funktion g på a,b och c ändras inte olikhetstecknen dvs man får $g(a) < g(b) < g(c)$, men om g är AVTAGANDE så kastas olikhetstecknen om dvs då får man $g(a) > g(b) > g(c)$.

Ett viktigt specialfall av detta sista är då man multiplicerar med -1 (vilket svarar mot den avtagande funktionen $g(x) = -x$). Ett annat fall när olikhetstecknen kastas om är vid invertering (dvs $g(x) = 1/x$). Exempel: $2 < 3 < 4 \Rightarrow 1/2 > 1/3 > 1/4$. Vid rotutdragning (som svarar mot den växande funktionen $g(x) = \sqrt{x}$) ändras däremot inte olikhetstecknen.

Följande snabbmetod som grundar sig på det ovanstående kan det dock vara bra att kunna utantill: $-a < \frac{x-m}{b} < a \Rightarrow x - ab < m < x + ab$.

Räkna som övning tal 12.21 i kursboken. Använd FS 11.2 d och snabbmetoden ovan. Härvid motsvaras x av $\bar{X} - \bar{Y}$, m av $\mu_A - \mu_B$, b av s$\sqrt{\frac{1}{9} + \frac{1}{16}}$ och a av $t_{0.005}(23)$.

Vi ska nu fortsätta talet genom att även härleda ett 90 % igt k.i. för σ. Vi skattar σ med S i FS 11.2 b dvs skattningen av σ fås genom en sammanvägning av de skattningar av σ som kommer från respektive stickprov. OBS. För att sammanvägningen ska vara meningsfull är det viktigt att σ är lika för de bägge stickproven så att 5.0 och 7.1 skattar SAMMA σ. Om vi normerar S till $\frac{23S^2}{\sigma^2}$ så får detta enligt 11.2 b en $\chi^2(23)$-fördelning. Tabell 4 i FS ger nu $90\% = P(13.1 < \frac{23S^2}{\sigma^2} < 35.2) =$ (Invertering!) $= P(1/13.1 > \frac{\sigma^2}{23S^2} > 1/35.2) =$ (Multiplikation med $23S^2 > 0$!) $= P(23S^2/13.1 > \sigma^2 > 23S^2/35.2) =$ (rotutdragning!) $= P(\sqrt{23S^2/13.1} > \sigma > \sqrt{23S^2/35.2})$. Vi ersätter nu S^2 med $S^2_{obs} = \frac{(9-1)(5.0)^2+(16-1)(7.1)^2}{9+16-2}$ och får (om vi vänder intervallet rätt) $5.21 < \sigma < 8.54$ (90 %) som är svaret.

Standardavvikelse är en term som används i många olika betydelser. Ett enkelt exempel:

Ex 1. $X_1, X_2 \cdots X_{22}$ är oberoende och likafördelade. Vi sätter $\mathrm{E}(X_i) = \mu$ och $D(X_i) = \sigma$. Data: $\sum_{i=1}^{22} x_i = 1187$ och $\sum_{i=1}^{22} x_i^2 = 66526$. (Här är x_i= observerade värdet på X_i.)

σ ovan är X_i:nas standardavvikelse. Detta är en (här) okänd konstant som man skulle kunna kalla för den teoretiska standardavvikelsen. σ uppskattas med s i FS 8. Vi får s= $\sqrt{\frac{1}{21}(66526 - \frac{(1187)^2}{22}})$ $= 10.9 (\approx \sigma)$

Detta s som uppskattar σ kallas ofta också för standardavvikelsen, men borde rätteligen kallas för stickprovs-standardavvikelsen. s är ett mått på hur mycket observationerna i genomsnitt sprider sig kring sitt medeltal, vilket framgår av den första formeln för s^2 i FS 8. (Förklaringen till $\frac{1}{n-1}$ istället för $\frac{1}{n}$ finns i tillägg 11 i mitt tärningskompendium.)

Om man nu också vill skatta väntevärdet μ så använder man medeltalet $\bar{X} = \frac{1}{22}\sum_{i=1}^{22} X_i$. Här skriver vi X_i och inte x_i eftersom vi vill se på skattningen som den ser ut INNAN vi gjort några observationer. Det är lätt att visa att E($\bar{X}$) $= \mu$ vilket betyder att skattningen är väntevärdesriktig, men man kan också (via variansen) bestämma standardavvikelsen för SKATTNINGEN av μ dvs bestämma D($\bar{X}$). Vi får D ($\bar{X}$) $= \frac{\sigma}{\sqrt{22}}$ (Jfr med parametrarna för N-fördelningen i FS 11.1 a !) En siffermässig uppskattning av denna sista standardavvikelse är vad som kallas för skattningens medelfel. Vi får här $\frac{\sigma}{\sqrt{22}} \approx \frac{10.9}{\sqrt{22}} = 2.32$.

Sammanfattningsvis är alltså medelfelet (=2.32) en uppskattning av noggrannheten i SKATTNINGEN av μ medan s (=10.9) är en uppskattning av sannolikhetsfördelningens standardavvikelse. Se upp med detta! Ett VANLIGT FEL är att s och medelfelet förväxlas.

När man ska beräkna k.i. för väntevärden med approximativ konfidensgrad använder man andra ställen i FS än FS 11. Ofta används CGS (FS 5) som gör att man kan räkna med approximativ normalfördelning. Typiskt i dessa sammanhang är att man vanligen gör ännu en approximation genom att skatta standardavvikelsen i normalfördelningen.

Ex.2 Beräkna ett $\approx$ 95 %-igt k.i. för μ i exempel 1.

Lösning: Enligt FS 5 gäller att $\sum_{i=1}^{22} X_i$ är $\approx$ N(22μ, $\sigma\sqrt{22}$). Om vi här approximerar σ med s (=10.9) ger detta att $\sum_{i=1}^{22} X_i$ är $\approx$ N(22 μ, 51.13). Division med 22 gör att både väntevärdet och standardavvikelsen divideras med 22 vilket ger $\bar{X} \approx$ N(μ, 2.32). Tabell 2 ger nu att 95% $\approx$ P($-1.96 < \frac{\bar{X}-\mu}{2.32} < 1.96$) = (Enligt snabbmetoden ovan)= $P(\bar{X} - 1.96 \cdot 2.32 < \mu < \bar{X} + 1.96 \cdot 2.32)$. Insättning av $\bar{X}_{obs} = 1187/22 = 53.95$ ger $I_\mu : 53.95 \pm 4.55(\approx 95\%)$ som är svaret. (Observera att $1.96 \cdot 2.32 = \lambda_{0.025}\cdot$ medelfelet.)

Här är ett annat exempel där man istället använder FS 6 och också utnyttjar en viktig sats för poisson-fördelningar.

Ex. 3. Under ett år har man för varje kvartal noterat antalet olyckor (av något slag). Som modell antar vi att de fyra observationerna är observationer på oberoende Poisson-fördelade variabler med olika väntevärden dvs X_i är Po(μ_i) där $\mu_i = E(X_i) = V(X_i)$ enligt FS 3. Våra observationer är: $x_1 = 19, x_2 = 8, x_3 = 12, x_4 = 26$

Man är intresserad av att bestämma ett $\approx$ 95 %-igt k.i. för: a. Skillnaden mellan kvartal 4 och kvartal 1 dvs för $\mu_4 - \mu_1$. b. Medelantalet olyckor under året dvs för $\frac{1}{4}(\mu_1 + \cdots + \mu_4)$.
Lösning: a. $X_4 - X_1$ skattar $\mu_4 - \mu_1$. Eftersom μ_i:na är väntevärden ska de skattas med motsvarande medeltal vilket här blir lika med den observation vi har. Eftersom $\mu_4 \approx 26 \geq 15$ har vi enligt FS 6 att poissonfördelningen för X_4 får approximeras med en normalfördelning, dvs vi har $X_4 \approx N(\mu_4, \sqrt{\mu_4})$. Om vi här även approximerar standardavvikelsen får vi $X_4 \approx N(\mu_4, \sqrt{26})$. På samma sätt får vi $X_1 \approx N(\mu_1, \sqrt{19})$. Eftersom en linjär sammansättning av oberoende N-fördelade variabler är N-fördelad ger detta att $X_4 - X_1$ är $\approx$ N($\mu_4 - \mu_1, \sqrt{26+19}$). (OBS. att $V(X_4 - X_1) = V(X_4) + V(X_1) = 26 + 19$.) Vi fortsätter nu precis som i ex.2 och får $I_{\mu_4-\mu_1} : 26 - 19 \pm 1.96 \cdot \sqrt{26+19}(\approx 95\%)$ etc

b.$\frac{1}{4}(X_1 + \cdots + X_4)$ skattar $\frac{1}{4}(\mu_1 + \cdots + \mu_4)$. Nu går det inte att göra som i a eftersom X_2 och X_3 inte uppfyller approximationsvillkoret $\mu \geq 15$. Istället använder man först satsen att en summa av oberoende poissonfördelade är poissonfördelad.(OBS. att satsen BARA gäller för summor.) Om summan kallas Y gäller att Y är Po ($\mu_1 + \cdots + \mu_4$) $\approx$ Po (19+8+12+26)= Po(65). Samma approximationer som i a ger nu Y$\approx N(\mu_1 + \cdots \mu_4, \sqrt{65})$. Fortsätter man sedan som i ex.2 får man det sökta intervallet: $\frac{65}{4} \pm 1.96 \cdot \frac{\sqrt{65}}{4}$ dvs $16.25 \pm 3.95(\approx 95\%)$.

Ex.4 $X_1, X_2 \cdots X_5$ är oberoende alla med fördelningen N($3\theta, \theta$). Data: $\sum_{i=1}^{5} x_i = 18.35$. Beräkna ett k.i. för θ med den EXAKTA konfidensgraden 90 %.

Lösning: I alla tidigare k.i. som vi bestämt kunde man istället för att härleda intervallen ha använt FS 12. Här går inte detta utan man MÅSTE härleda. Det är dock inte så svårt. Enligt FS 11.1 a gäller att $\bar{X}$ är N($3\theta, \frac{\theta}{\sqrt{5}}$). Tabell 2 ger nu 90% = $P(-1.6449 < \frac{\bar{X}-3\theta}{\theta/\sqrt{5}} < 1.6449)$. Sätter vi in $\bar{X}_{obs} = 18.35/5 = 3.67$ kan vi göra följande omformning: $-1.6449 < \frac{3.67-3\theta}{\theta/\sqrt{5}} < 1.6449 \Leftrightarrow \frac{-1.6449}{\sqrt{5}} < \frac{3.67-3\theta}{\theta} < \frac{1.6449}{\sqrt{5}} \Leftrightarrow \frac{-1.6449}{\sqrt{5}} < \frac{3.67}{\theta} - 3 < \frac{1.6449}{\sqrt{5}} \Leftrightarrow \frac{-1.6449}{\sqrt{5}} + 3 < \frac{3.67}{\theta} < \frac{1.6449}{\sqrt{5}} + 3 \Leftrightarrow 2.264 < \frac{3.67}{\theta} < 3.736$ (invertera!) $\Leftrightarrow 1/2.264 > \frac{\theta}{3.67} > 1/3.736 \Leftrightarrow 3.67/2.264 > \theta > 3.67/3.736 \Rightarrow$ Svar: $0.98 < \theta < 1.62$ (90%).