Rattfylleri - ett belysande exempel om hypotesprövning

(Hämtat från ett problem i Bloms bok).

För att skatta den sanna promillehalten θ i blodet på en misstänkt rattfyllerist tas 3 prov och man erhåller resultaten x_1, x_2, x_3. Man har vidare ansatt modellen att dessa mätvärden är utfall av oberoende stokastiska variabler X_1, X_2, X_3 som är $N(\Theta, 0.10)$. Detta innebär att man har modellen att mätvärdena är mätningar av den sanna promillehalten Θ men med normalfördelade slumpfel som alltså antas vara $N(0, 0.10)$. Att man ansåg sig känna spridningen berodde på att man gjort många mätningar med mätutrustningen och kunnat (nästan) exakt skatta mätfelsvariansen. För övrigt är numera mätutrustningen bättre, dvs har mindre slumpfelsvarians - mer om det senare!

Med utgångspunkt i dessa data avgörs om personen i fråga haft straffbar promillehalt eller ej. Beslutsregeln (lär) vara att om $\bar{x} = (x_1 + x_2 + x_3)/3 > 1.13$ så anses det styrkt att personen hade över 1.0 promille och döms! Detta utgör ett exempel på hypotesprövning. Gränsen 1.13 lär vara framtagen ur ovanstående modell för data enligt följande: Vi ställer upp nollhypotesen $H_0 : \Theta \leq 1.0$ och alternativhypotesen $H_1 : \Theta > 1.0$ (sortenhet promille) och utför en hypotesprövning på nivån 1%. Notera att H_0 svarar mot ”oskyldig” och H_1 mot ”skyldig”. I hypotesprövning har nollhypotesen H_0 alltid en fördel mot alternativhypotesen H_1 och detta motsvarar rättstrygghetens princip ”Man är oskyldig tills man överbevisats om motsatsen”. På basis av våra mätdata (här de tre promillehaltsbestämningarna) skall vi fatta endera av två beslut:

1) Förkasta H_0 till förmån för H_1 eller lite slarvigt uttryckt: H_1 är bevisad. Det svarar här mot beslutet ”Döm!”.

2) Förkasta ej H_0 till förmån för H_1. Detta svarar mot beslutet: ”Frikänn!” Ibland uttrycks detta (lite oegentligt) som att ”vi kan acceptera H_0”, men den korrekta slutsatsen är egentligen ”H_1 ej bevisad”. Man kan <u>inte</u> (även om det ibland görs!) betrakta detta som att H_0 skulle vara bevisad. Orsaken till detta är att skälet till att vi ej lyckades förkasta H_0 (dvs ej lyckades ”bevisa” H_1) ju kan vara att vi gjort ett dåligt försök (kanske med för få mätningar eller med för stora mätfel!) Om man nu <u>egentligen</u> var ute efter att kunna dra slutsatsen ”H_0 bevisad” av att ha misslyckats med att förkasta H_0 skulle ju detta premiera ett uselt och eländigt upplagt försök med få observationer och stora mätfel! Tyvärr är det inte ovanligt i framför allt medicinska sammanhang att man på detta sätt försöker dra slutsatsen att om man inte lyckats påvisa en skillnad mellan två behandlingar så har man därmed ”bevisat” att de är lika bra! I rättegångssituationen är inte uppgiften för den anklagade att bevisa sin oskuld utan det är åklagarens uppgift att bevisa hans skuld! Att åklagaren misslyckats att få den anklagade dömd innebär ju inte att den anklagade har <u>bevisats</u> vara oskyldig (åklagaren hade kanske inte de rätta bevisen!). Dock brukar man uttrycka det som att personen är att betrakta som oskyldig tills han/hon är dömd.

Vi kan här göra två slags felbeslut:

1) Fel av första slaget (typ I-felet eller α-felet brukar det kallas) nämligen att förkasta H_0 trots att H_0 är sann. Detta motsvarar i rättegångssituationen ”att döma en oskyldig”. Detta felbeslut vill vi ha kontroll över och man har här satt denna sannolikhet till $\alpha = 1\% = 0.01$. Detta betyder alltså att risken att döma en oskyldig skall vara maximalt 1%! Man har här en kvantifiering av det som på juristspråk brukar benämnas ”bortom rimligt tvivel”! Möjligen skulle ingen jurist vilja kännas vid detta, men orsaken att de inte fattat vad det är frågan beror antagligen på dåliga kunskaper i statistik!

2) Fel av andra slaget (typ II-felet eller β-felet brukar det ibland kallas) nämligen att ej förkasta H_0 till förmån för H_1 trots att H_1 är sann. Detta motsvarar i rättegångstermer ”att frikänna en skyldig”.

Det är här en fråga om att väga typ I-felet mot typ II-felet och här är det framför allt typ I-felet som är av intresse. Valet av gränsen 1.13 (observera att det lite överstiger 1.0 som är gränsen för "skyldig") svarar mot en viss specificerad avvägning mellan typ I- och typ II-felen. Det fiffiga med denna avvägning är att den kan uttryckas som ett villkor på typ I-felet, som vi är mest oroliga för! Notera att vi kan få ner typ I-felet hur mycket som helst genom att öka på förkastelsegränsen från 1.13 till något högre värde. Det pris vi då får betala är att typ II-felet ökar, dvs vi kan öka säkerheten för de oskyldiga att bli frikända genom att kräva att man har högt medelvärde av sina mätningar för att bli dömd, men detta sker till priset av att man kommer att frikänna fler skyldiga!

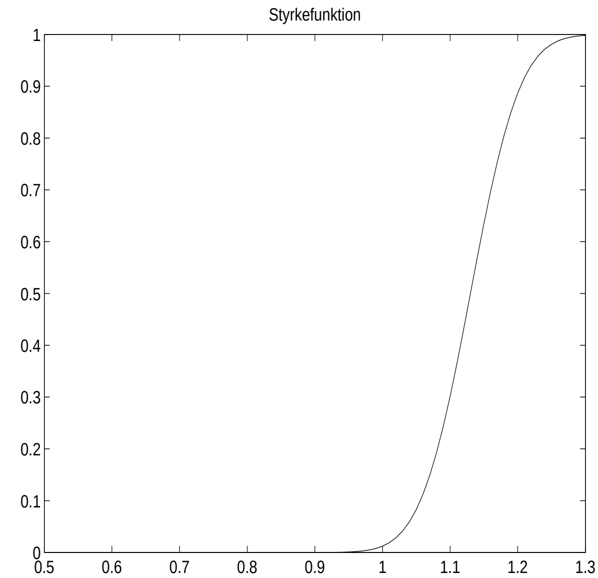
I hypotesprövningsterminologi är $\bar{x}$ en testvariabel och vi har ett förkastelseområde $C = [1.13, \infty)$ dvs om $\bar{x} \in C$ så förkastas H_0 till förmån för H_1 - annars ej.

Principen är alltså (den rimliga): Om man har "högt" medelvärde skall man dömas annars ej. Vad som menas med "högt" (dvs över 1.13) har valts så att risken att en oskyldig skall dömas skall vara högst 1%. Detta betyder alltså att $P(\text{förkasta } H_0 \text{ då } H_0 \text{ är sann})$ skall vara högst 1%. Nedan finns en härledning av hur man kommit fram till detta. Vi glömmer därför tillfälligtvis bort att gränsen för "högt" medelvärde är framräknad till 1.13 och kallar den för k och försöker ta fram ett värde på k.

Här kan det vara lämpligt att införa den s k styrkefunktionen $h(u) = P(\text{förkasta } H_0 \text{ då } \Theta = u)$, dvs sannolikheten att bli dömd då man har den verkliga promillehalten u. Detta innebär att vi skall beräkna $P(\bar{X} > k)$ då $\Theta = u$. Man noterar att $\bar{X}$ är $N(u, 0.10/\sqrt{3})$ så denna sannolikhet blir

$$h(u) = P(\bar{X} > k) = P\left(\frac{\bar{X} - u}{0.10/\sqrt{3}} \geq \frac{k - u}{0.10/\sqrt{3}}\right) = 1 - \Phi\left(\frac{k - u}{0.10/\sqrt{3}}\right).$$

Om man ritar upp denna funktion av u inser man att den är monotont växande i u, börjar i 0 och slutar i 1. Dessutom antar den värdet 1/2 för $u = k$. Att funktionen är växande i u innebär att ju högre promillehalt man har desto större är risken att dömas - onekligen en rimlig princip. Dessutom inser man lätt att om man ändrar k så innebär det en stel förskjutning av hela styrkefunktionen.



För att få ett typ I-fel av maximalt $\alpha = 0.01$ så skall vi bestämma k så att $\max_{u \leq 1.0} h(u) \leq \alpha = 0.01$ eftersom H_0 svarar mot $u \leq 1.0$. Detta innebär att vi måste ha $h(1.0) = \alpha = 0.01$ (eftersom maximum antas för $u = 1.0$) dvs

$$1 - \Phi\left(\frac{k - 1.0}{0.10/\sqrt{3}}\right) = 0.01.$$

Detta ger $\dfrac{k - 1.0}{0.10/\sqrt{3}} = \lambda_{0.01}$ dvs att $k = 1.0 + \lambda_{0.01} 0.1/\sqrt{3} \approx 1.13$.

Vad som hänt sen 1960-talet är att mätutrustningen förbättrats väsentligt på så sätt att mätfelsvariansen minskat betydligt från ovan angivna 0.10. Man har dock inte ändrat förkastelsegränsen utan vad som i praktiken hänt är att α-felet (den s k signifikansnivån) i hypotesprövningen minskat betydligt. Egentligen borde man ha anpassat förkastelsegränsen i enlighet med ovanstående kalkyl med hänsyn till mätfelsvariansen till $1.0 + \lambda_{0.01}\sigma_0/\sqrt{3}$ där σ_0^2 är mätfelsvariansen, men denna kan ju skilja sig mellan olika laboratorier och man skulle då få olika definition av "högt" medelvärde vilket antagligen skulle upplevas som stötande.

För övrigt kan man notera att villkoret för att dömas är $\bar{x} > \Theta_0 + \lambda_\alpha\sigma_0/\sqrt{n}$ där $\Theta_0 = 1.0, \alpha = 0.01, \sigma_0 = 0.10$ och $n = 3$. Detta kan skrivas som $\Theta_0 < \bar{x} - \lambda_\alpha\sigma_0/\sqrt{n}$ där man känner igen uttrycket som undre gränsen i ett ensidigt konfidensintervall (konfidensgrad $1-\alpha$) av typen $(\bar{x} - \lambda_\alpha\sigma_0/\sqrt{n}, \infty)$ och detta innebär att hypotesprövningen kunnat utföras med hjälp av den s k konfidensmetoden. Denna innebär ju att vi förkastar $H_0 : \Theta \leq \Theta_0$ till förmån för $H_1 : \Theta > \Theta_0$ om Θ_0 ej tillhör ett enkelsidigt konfidensintervall av ovanstående typ. I rattfyllerikretsar lär f ö 0.13 benämnas "rabatten" - man kan ju uppfatta det som man får dra av 0.13 från sitt medelvärde $\bar{x}$ och det är bara om detta "rabatterade" värde överstiger 1.0, som man döms.

Skulle vi i stället ha velat testa $H_0 : \Theta \geq \Theta_0$ mot $H_1 : \Theta < \Theta_0$ skulle vi ha valt konfidensintervallet $(-\infty, \bar{x} + \lambda_\alpha\sigma_0/\sqrt{n})$. I rättegångstermer skulle detta svara mot att bevisbördan låg på den anklagade - han/hon skulle behöva övertyga rätten om att promillehalten <u>understeg</u> 1.0.

Skulle vi slutligen vilja testa $H_0 : \Theta = \Theta_0$ mot $H_1 : \Theta \neq \Theta_0$ (ett dubbelsidigt test) skulle vi ha beräknat ett två-sidigt konfidensintervall av typen $(\bar{x} \pm \lambda_{\alpha/2}\sigma_0/\sqrt{n})$. För övrigt kan nämnas att ett dubbelsidigt test på signifikansnivån α lämpligen borde uppfattas som två stycken enkelsidiga test på nivån $\alpha/2$ vardera. Om vi låter händelsen A='ena enkelsidiga testet förkastar H_0' och B='andra enkelsidiga testet förkastar H_0' så ser vi att A och B är disjunkta eftersom $\bar{x}$ inte samtidigt kan vara ("långt") under Θ_0 och ("långt") över Θ_0. Detta ger $P(A \cup B) = P(A) + P(B)$ så om vardera testet utförs på nivån $\alpha/2$ blir totala nivån α. Detta innebär alltså att man inte bara kan säga att $\Theta \neq \Theta_0$ (som egentligen är slutsatsen av det dubbelsidiga testet som det är formulerat i läroboken) utan också göra ett uttalande om att Θ ligger under eller över Θ_0. Ytterst få läroböcker (och inte heller vår lärobok) skriver ut detta explicit, men i de genomräknade exemplen drar man (naturligtvis) slutsats om åt vilket håll Θ avviker från Θ_0 i de fall som data ger upphov till att vi förkastar $H_0 : \Theta = \Theta_0$ till förmån för $H_1 : \Theta \neq \Theta_0$.