Fördelningsfria test

Vi ska nu betrakta två test där normalfördelning inte behöver förutsättas.

1.1 Teckentestet

Vi betraktar situationen med parvisa observationer, dvs vi har n par av värden $(x_1, y_1),\ (x_2, y_2), \dots, (x_n, y_n)$, där ett par uppstått t.ex. därigenom att ett visst objekt undersökts enligt två olika metoder A och B. Värdena x_i, y_i betraktas som oberoende observationer av de (kontinuerliga) s.v. X_i, Y_i. Man önskar avgöra om det, bortsett från slumpvariationer, finns någon skillnad mellan komponenterna i paren, dvs korrektare uttryckt: mellan fördelningarna för X_i och Y_i.

Teckentestet utföres så här:
Man ser efter i hur många fall x-värdet är större än y-värdet. Detta antal betecknas med z. Om det inte finns någon systematisk skillnad mellan komponenterna bör det bli ungefär lika många fall med x större än y som med x mindre än y.

Vi preciserar detta till:

$$\begin{aligned}
H_0 &: X_i \text{ och } Y_i \text{ har samma fördelning } (i = 1, \dots, n);\\
H_1 &: X_i\text{:s fördelning är förskjuten i förhållande till } Y_i\text{:s}\\
&\quad\ \text{fördelning åt } \textit{samma håll} \text{ för alla } i.
\end{aligned}$$

Om H_0 är sann är z en observation från $\mathrm{Bin}(n, 1/2)$. Orsaken härtill är att, om H_0 är sann, $P(X_i > Y_i) = P(X_i < Y_i) = 1/2$. H_0 testas lämpligen med P-värdesmetoden. Om $n > 40$ så kan normalfördelningsapproximation användas.

1.2 Wilcoxons rangsummetest

Vi övergår nu till fallet med *två oberoende stickprov* utan normalfördelningsantaganden. Låt $x_1, \ldots, x_{n_1}$ och $y_1, \ldots, y_{n_2}$ vara oberoende observationer av de (kontinuerliga) s.v. X och Y. De kan t.ex. ha uppstått genom att n_1 respektive n_2 mätningar utförts med två olika metoder A och B. Finns det någon skillnad mellan metoderna? Mer preciserat betrakta

$$H_0 : X \text{ och } Y \text{ har samma fördelning;}$$

$$H_1 : X\text{:s fördelning är förskjuten i förhållande till } Y\text{:s.}$$

Wilcoxons rangsummetest utförs så här:
Storleksordna alla n_1+n_2 observationerna och ersätt den minsta observationen med rangen 1, den näst minsta med 2,..., den största med n_1+n_2. Därmed har x-observationerna ersatts med rangerna $r_1, \ldots, r_{n_1}$. Som teststorhet använder vi rangsumman $r = r_1 + \cdots + r_{n_1}$ och förkastar H_0 om den är tillräckligt liten eller tillräckligt stor. Vad menas ned tillräckligt?

Observera följande: Om X och Y har samma kontinuerliga fördelning så uppkommer $r_1, \ldots, r_{n_1}$ precis som om man drog slumpmässigt utan återläggning n_1 av talen $1, 2, \ldots, n_1 + n_2$. Under H_0 blir rangsumman r en observation av en s.v. R med en fördelning som i princip kan härledas genom att skriva upp alla möjligheter; tabeller finns för små n_1 och n_2. Vi har

$$E(R) = n_1 \cdot \frac{n_1 + n_2 + 1}{2} \quad \text{och} \quad V(R) = \frac{n_1 n_2 (n_1 + n_2 + 1)}{12}.$$

Utom för (helt) små n_1 och n_2 så kan normalapproximation av R göras.

Eftersom medelrangen under H_0 är

$$\frac{1 + 2 + \cdots + (n_1 + n_2)}{n_1 + n_2} = \frac{n_1 + n_2 + 1}{2},$$

följer $E(R)$. $V(R)$ är svårberäknad, och knepet är att betrakta antalet gånger som händelserna $\{X_i < Y_j\},\ i = 1, \ldots, n_1,\ j = 1, \ldots, n_2$ inträffar.

Det är förvånande att man förlorar mycket lite i styrka om man använder Wilcoxson-testet i stället för t-testet med ”två stickprov” då normalfördelningsantagandet är uppfyllt. Begreppet ”asymptotisk relativ effektivitet”, som nämns i boken, är komplicerat.

Planering av statistiska undersökningar

1.3 Icke-jämförande undersökningar

Vi begränsar oss till frågan om *stratifiering*. Därmed menas undersökning av en population som avsiktligt har uppdelats i k *delpopulationer* eller *strata*. Ett stickprov tas ut från varje delpopulation.

Antag att det i varje stratum finns en okänd parameter μ_i i i:te stratum ($i = 1, \ldots, k$), där k är antalet strata. Antag vidare att varje värde x_{ij} i stickprovet $x_{i1}, x_{i2}, \ldots, x_{in_i}$ från detta stratum är observation av en s.v. X_{ij} med väntevärdet lika med parametern μ_i samt standardavvikelsen σ_i. Dessa parametrar har dock bara sekundärt intresse, eftersom det är totalpopulationens egenskaper vi vill åt. Vi antar nämligen att det som skall skattas är ett vägt medelvärde

$$\mu = \sum_1^k c_i \mu_i, \quad \text{där} \quad \sum_1^k c_i = 1$$

av parametrarna μ_i (med kända vikter c_i som ofta är relaterade till stratas storlekar, t.ex. $c_i = N_i/N$).

Som skattning av μ tar vi naturligt nog det vägda medelvärdet

$$\mu^* = \sum_1^k c_i \bar{x}_i$$

av stickprovsmedelvärdena i de olika stickproven.

Vi har

$$E(\mu^*) = \sum_{i=1}^k c_i E\big(\overline{X}_i\big) = \sum_{i=1}^k c_i \mu_i = \mu$$
$$V(\mu^*) = \sum_{i=1}^k c_i^2 V\big(\overline{X}_i\big) = \sum_{i=1}^k c_i^2 \sigma_i^2 / n_i.$$

Första relationen utsäger att μ^* är väntevärdesriktig, den andra att dess varians beror av standardavvikelserna inom strata men inte av eventuella skillnader mellan μ_i-värdena. Stratifiering är alltså fördelaktig från precisionssynpunkt, då den population som skall undersökas kan indelas i homogena delar, dvs i strata inom vilka standardavvikelserna är små.

Vi skall något beröra frågan hur man vid planering bör välja antalen n_i. En möjlighet är att begagna *proportionella* antal. Därvid insamlas totalt n värden, vilka fördelas mellan delpopulationerna i samma proportioner som vägningstalen c_i, dvs man väljer

$$n_i = n \cdot c_i.$$

Detta ger

$$V(\mu^*) = \frac{1}{n} \sum_{i=1}^{k} c_i \sigma_i^2.$$

Bättre är i regel att använda *optimala* antal: Totala antalet uttagna element n fixeras. Antalen n_i väljes så att variansen $V(\mu^*)$ blir så liten som möjligt. Matematiskt sett innebär detta att man minimerar

$$V(\mu^*) = \sum_{i=1}^{k} c_i^2 \sigma_i^2 / n_i.$$

med bivillkoret $\sum_{i=1}^{k} n_i = n$. Lösningen blir

$$n_i = n \cdot \frac{c_i \sigma_i}{\sum\limits_{i=1}^{k} c_i \sigma_i}$$

där svaret justeras så att n_i blir heltal. Detta ger

$$V(\mu^*) = \frac{1}{n} \left(\sum_{i=1}^{k} c_i \sigma_i \right)^2.$$

Man ser att delpopulationer med stor spridning tilldelas proportionsvis fler värden än de som har liten spridning, vilket är mycket rimligt. En svårighet är att standardavvikelserna σ_i inte alltid är kända på förhand. Man är då hänvisad till att utföra förundersökningar eller begagna redan tillgängliga undersökningsresultat eller ”gissningar”.