STICKPROV I PAR

Jämförelse mellan två metoder vid parvisa observationer.

Vi betraktar mätningar gjorda parvis med två metoder X och Y.
Anta att vi har parvisa observationer (x_k,y_k) , k=1,2,...,n vars koordinater x_k och y_k hör till normalfördelade s.v. Vi antar att paren (x_1,y_1),...,(x_n,y_n) är oberoende.
Låt Z=X–Y.

X	x_1	x_2	...	x_n
Y	y_1	y_2	...	y_n
Z=X – Y	$z_1 = x_1 - y_1$	$z_2 = x_2 - y_2$	...	$z_n = x_n - y_n$

Vi betraktar **oftast** förekommande fall med **okända** standardavvikelser.

Vi beräknar $\bar{z} = \dfrac{z_1 + z_2 + \cdots + z_n}{n}$ och

$$\sigma_z = \sqrt{\text{Variansen}} = \sqrt{\frac{(z_1-\bar{z})^2 + (z_2-\bar{z})^2 + \cdots + (z_n-\bar{z})^2}{n-1}}$$

Konfidensintervall för μ_z med $(1-\alpha)$ konfidensgraden blir då

$$\left[\bar{z} - t_{\alpha/2}(n-1)\cdot\frac{\sigma_z}{\sqrt{n}},\quad \bar{z} + t_{\alpha/2}(n-1)\cdot\frac{\sigma_z}{\sqrt{n}}\right].$$

Om intervallet **inte** innehåller 0 då kan vi med konfidensgraden $(1-\alpha)$ påstå att det finns en **skillnad** mellan s. v. X och Y .

==

ÖVNINGSUPPGIFTER

Uppgift 1.

Tio vuxna män i åldrarna 35 till 50 år deltog i en studie för att se hur motion och kostvanor påverkade kolesterolvärdena i blodet. Kolesterolhalten mättes hos var och en innan de började med motion och kost med lite fett, samt tre månader efter att programmet startat. Följande resultat erhölls: (3p)

Person nr	1	2	3	4	5	6	7	8	9	10
Före	265	240	258	295	251	245	287	314	260	279
Efter	229	231	227	240	238	241	234	256	247	239

Kan man med 95% sannolikhet påstå att motion och begränsat fettintag i maten ger lägre kolesterolhalt i blodet?

Lösning:

”**Stickprov i par**” ger:

$z_i = x_i - y_i$: 36, 9, 31, 55, 13, 4, 53, 58, 13, 40

$\bar{z} = 31{,}2 \quad s^*{}_{obs} = 20{,}43$, $t_{0{,}025}(9) = 2{,}2622$,

konfidensintervall:

$$\left[31{,}2-2{,}2622\cdot\frac{20{,}43}{\sqrt{10}},\ 31{,}2+2{,}2622\cdot\frac{20{,}43}{\sqrt{10}}\right]$$

$= [16{,}58;45{,}82]$

Svar: Eftersom intervallet ej innehåller noll så kan man med 95% säkerhet påstå att kolesterolhalten blir lägre.

Uppgift 2.

Man vill jämföra två maskiner A och B med avseende på en viss kvalitetsvariabel hos de tillverkade enheterna. För båda maskinerna kan denna variabel antas vara normalfördelad med okänd standardavvikelse. Man har under 5 dagar i rad tillverkat ett antal enheter med maskinerna varvid man fått följande observationer.

A	131	136	142	149	147
B	126	131	139	141	139

a) Ange ett symmetriskt 95% konfidensintervall för $Z = A - B$.
b) Kan man med 95% sannolikhet påstå att maskinerna A och B skiljer sig åt?

Lösning:

Först beräknar vi differensen A – B

A	131	136	142	149	147
B	126	131	139	141	139
z= A – B	5	5	3	8	8

och därefter medelvärdet $\bar{z}$ =5.8 och standardavvikelsen

$$\sigma^* = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(z_i-\bar{z})^2} = 2.1679$$

Eftersom n=5 har vi n-1 = 4 frihetsgrader.
$\alpha = 5\% = 0.05 \Rightarrow \alpha/2 = 0.025$ dvs F(x)= 0.975
Från tabellen för t-fördelning med r=4 frihetsgrader får vi
$t_{\alpha/2}(n-1) = 2.776$

Härav $t_{\alpha/2}(n-1)\dfrac{\sigma^*}{\sqrt{n}} = 2.776\cdot\dfrac{2.1679}{\sqrt{5}} = 2.69$ (felmarginal)

Konfidensintervall är

$$(\bar{z}-t_{\alpha/2}(n-1)\frac{\sigma^*}{\sqrt{n}},\quad \bar{z}+t_{\alpha/2}(n-1)\frac{\sigma^*}{\sqrt{n}}) = (5.8-2.69,\ 5.8+2.69) = (3.11,\ 8.49)$$

Intervallet innehåller inte 0. Därför kan vi med 95% säkerhet påstå att det finns en skillnad mellan A och B.

Svar: a) $(3.11, 8.49)$
b) Ja, vi kan med 95% säkerhet påstå att det finns en skillnad mellan A och B.

Uppgift 3.

Man vill jämföra två metoder för mätning av en viss variabel som anses normalfördelad med okänd standardavvikelse. Man har gjort 6 mätningar och fått följande observationer.

Metod 1	31	36	36	46	47	45
Metod 2	33	31	39	48	46	41

Kan man med 95% säkerhet påstå att metoderna skiljer sig åt?

Lösning:
Först beräknar vi differensen A – B

Metod 1	31	36	36	46	47	45
Metod 2	33	31	39	48	46	41
z=M1–M2	–2	5	–3	–2	1	4

och därefter medelvärdet $\bar{z}$ =0.5 och standardavvikelsen

$$\sigma^* = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(z_i-\bar{z})^2} = 3.3912$$

Eftersom n=6 har vi n–1 = 5frihetsgrader.
$\alpha = 5\% = 0.05 \Rightarrow \alpha/2 = 0.025$ dvs F(x)= 0.975
Från tabellen för t-fördelning med r=5 frihetsgrader får vi
$t_{\alpha/2}(n-1) = 2{,}5706$

Härav $t_{\alpha/2}(n-1)\dfrac{\sigma^*}{\sqrt{n}} = 2{,}5706 \cdot \dfrac{3.3912}{\sqrt{6}} = 3.5588$ (felmarginal)

Konfidensintervall är

$$(\bar{z} - t_{\alpha/2}(n-1)\frac{\sigma^*}{\sqrt{n}},\quad \bar{z} + t_{\alpha/2}(n-1)\frac{\sigma^*}{\sqrt{n}}) = (0.5-3.5588,\ 0.5+3.5588) = (-3.06,\ 4.06)$$

Intervallet innehåller 0. Därför kan vi INTE med 95% säkerhet påstå att det finns en skillnad mellan metoderna.
Svar: Nej, vi kan INTE med 95% säkerhet påstå att metoderna skiljer sig åt.