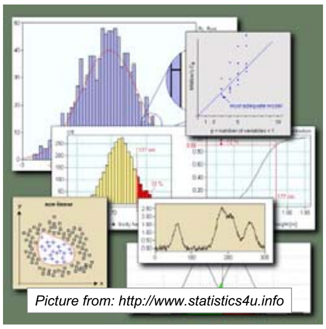
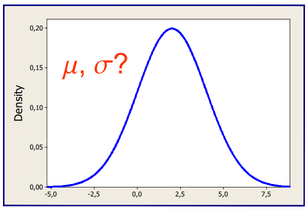
Icke-parametriska test

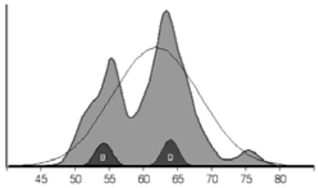

Picture from: http://www.statistics4u.info

uwe.menzel@math.uu.se

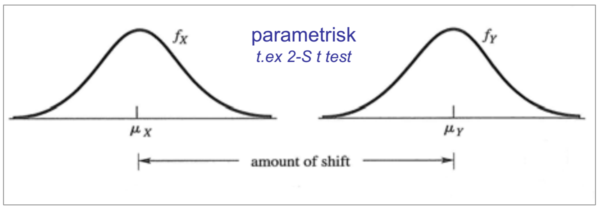
Icke-parametriska metoder

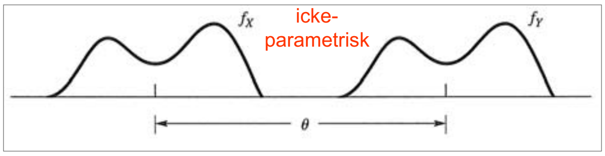
Parametriska metoder	Icke-parametriska metoder
Fördelningen för populationen som stickprovet togs ifrån är känd så nära som på ett antal parametrar, t.ex: N med okända μ och σ	Vill helst ha samma test men utan antaganden om den underliggande fördelningen!





"Location Shift"





När behövs ett icke-parametrisk test?

- Om förutsättningar för parametriska test inte är uppfyllda:
  - 2-Sample t-test: kräver normalfördelning för de underliggande populationerna
  - One-Way ANOVA: kräver normalfördelning för alla grupper, homogena varianser
- Om det är svårt att kvantifiera data
  - ordinalskalor (storleksordning finns, men differenser saknar betydelse: t.ex storlek av T-tröjor)
  - dessa data kan dock *rangordnas*

Hur pass bra fungerar icke-parametriska test?

- *Nästan* så bra som parametriska test om normalfördelning och andra förutsättningar föreligger
  - styrkan kan vara mindre jämfört med parametriska test
- Är ofta bättre om förutsättningarna (normalfördelning osv.) *inte* är uppfyllda

Vilket test "ersätts" med vilket?

population normalfördelad	population inte normalfördelad	
1-Sample t-test (1-Sample z-test)	1-Sample Sign test	allmän fördelning
	1-Sample Wilcoxon test	symmetrisk fördelning
Paired t-test	Matched Pairs Sign test	1-Sample Sign test på differenserna
	Wilcoxon-Signed Rank test	1-Sample Wilcoxon på differenserna
2-Sample t-test	Mann-Whitney U-test	
One-Way ANOVA	Kruskal-Wallis test	

A) One-Sample Sign test

- prövar om medianen för en fördelning är lika med ett hypotetiskt värde
- om nollhypotesen stämmer borde ungefär hälften av alla värden ur ett stickprov vara större än det hypotetiska värdet, andra hälften borde vara mindre
  - resultatet av ett stickrov kan skrivas så här:
    + + = + - + = + - = = + = + + - + + + = = + +
- med andra ord: sannolikheten att ett värde är större än det hypotetiska borde vara p=0.5

One-Sample Sign test

$p = P(X > \tilde{\mu}_0)$

$\boxed{H_0 : p = 0.5}$

$H_a : p \neq 0.5 \quad eller \quad p < 0.5 \quad eller \quad p > 0.5$

Testvariabel:

M = *antalet värden som är större än* $\tilde{\mu}_0$

$\boxed{M \in Bin(n, 0.5)}$ *under* H_0

Ω_{krit} : *mycket stora eller mycket små värden för* M

p – värdet enligt binomialfördelningen

$p = P(\ X \leq M\ ) + P(\ X \geq n - M\ ) \qquad 2 - sidigt$

Vill möss ha en egen spegel ?

- 16 möss med varsin bur som har ett rum med och ett rum utan spegel
- i vilket rum uppehöll sig musen mest ?

mus	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
med		x					x				x					
utan	x		x	x	x	x		x	x	x		x	x	x	x	x

Sherwin, C.M. 2004*. Mirrors as potential environmental enrichment for individually housed laboratory mice. Appl. Anim. Behav. Sci. 87: 95-103.*

Vill möss ha en egen spegel ?

$H_0 : p = 0.5 \qquad H_a : p \neq 0.5 \qquad tvåsidigt\ test$

$stickprov: \quad M = 3$

$M \in Bin(16, 0.5) \quad under\ H_0$

$p = P(M \leq 3) + P(M \geq 13) \qquad tvåsidigt\ test$

$= P(M \leq 3) + 1 - P(M \leq 12)$

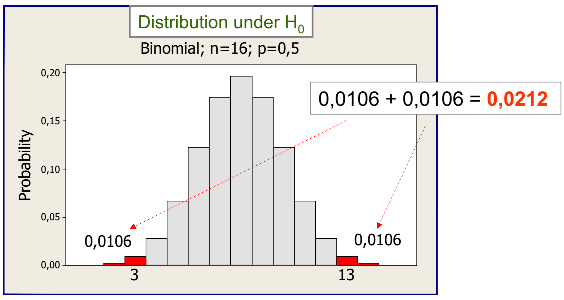
$= 0.0106 + 1 - 0.9894$

$= 0.0212$

Nollhypotesen förkastas pga. litet p-värde. Mus föredrar ... att *inte* ha någon spegel.

genetiskt förändrad mus

Tvåsidigt test

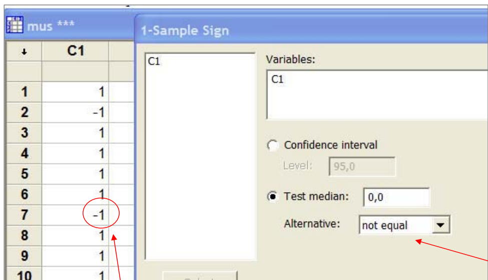


Det värdet fi fick är dock mycket osannolikt under H_0

P-värdet är sannolikheten för det värdet vi fick genom stickprovet eller mera extrema (mera osannolika) värden.

Minitab

Stat / Nonparametrics / 1-Sample Sign



tvåsidigt

negativ siffer för de som föredrade spegeln

Resultat

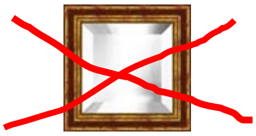
räknar bara hur många som är större än 0 och hur många som är mindre än 0

Sign Test for Median: C1

Sign test of median = 0,00000 versus not = 0,00000

	N	Below	Equal	Above	P	Median
C1	16	3	0	13	0,0213	1,000

Nollhypotesen förkastas pga. litet p-värde. Mus föredrar ... att *inte* ha någon spegel.



Om antalet värden är stort (n>25)

- Binomialfördelningen kan approximeras med normalfördelningen

$$M \in Bin(n,\, p) \qquad under\ H_0$$

$$Bin(n,\, p) \sim AsN\left(n \cdot p,\, \sqrt{n \cdot p \cdot (1-p)}\right)$$

$$\boxed{M \in N\left(\frac{n}{2}, \frac{1}{2}\sqrt{n}\right)}$$

under H_0 är M så här fördelad (p=0.5)

$$z = \frac{M - \frac{n}{2}}{\frac{1}{2}\sqrt{n}} \in N(0,1)$$

z, som är konstruerad av M, måste därför vara standard-normalfördelad.

$$\boxed{\Omega_{krit} = \left\{ |z| > \lambda_{\alpha/2} \right\} \qquad tvåsidigt}$$

vi förkastar H_0 om z verkar inte vara N(0,1)-fördelad.

Mer om One-Sample Sign testet

- Förutsättningar:
  - slumpmässigt stickprov, som vanligt
- Om fördelningen är normalfördelad är ett One-Sample t-test mera effektivt. (t-testet "känner" ju till avståndet av varje mätvärde till nollhypotesen ...)
- Har två matvärden samma storlek, alltså samma rang (s.k. "ties"), tilldelas båda medelvärdet av rangen
- Mätvärden med samma storlek som den hypotetiska medianen ignoreras ("="-värden), n förminskas motsvarande

+ + ~~=~~ + - + ~~=~~ + - ~~=~~ ~~=~~ + ~~=~~ + + - + + + ~~=~~ ~~=~~ + +

n → n - 7

B) 1-Sample Wilcoxon test

- prövar om medianen för en fördelning är lika med ett hypotetiskt värde
- gör detsamma som *Sign test*, men fördelningen av den underliggande populationen måste vara symmetrisk
- rangordning räknas ut, som för många icke-parametriska test
- Exempel: Vi har följande värden:
  - *-3,1 -6,3 1,2 -2,0 -1,0 -7,2 5,6 2,2 -12,0 -12,3 -5,3 -0,1 -23,4*
  - **Är medianen lika med noll?**

Absolutvärdena rangordnas:

Om vi t.ex prövar om $\mu_0=0$ summeras alla rang dess tillhörande värden är större än 0 och alla rang dess tillhörande värden är mindre än 0:

$$W^+ = \sum_i R_i^+$$

$$W^- = \sum_i R_i^-$$

$$W^+ = 3+8+5 = 16$$

$$W^- = 6+9+4+2+10+11+12+7+1+13 = 75$$

$$W = \min\left(W^+, W^-\right) = W^+ = 16 \qquad tvåsidigt$$

$$W = W^- \qquad lower\ tail$$

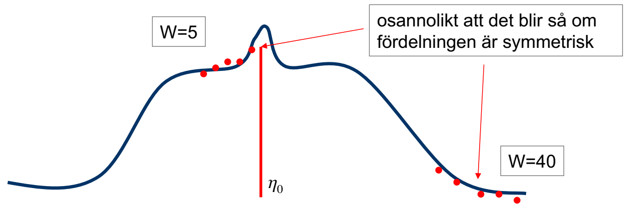
$$W = W^+ \qquad upper\ tail$$

$$\Omega_{krit} = \{ W < W_0 \} \qquad tabell$$

värde	abs	rang
-3,1	3,1	6
-6,3	6,3	9
1,2	1,2	3
-2,0	2,0	4
-1,0	1,0	2
-7,2	7,2	10
5,6	5,6	8
2,2	2,2	5
-12,0	12,0	11
-12,3	12,3	12
-5,3	5,3	7
-0,1	0,1	1
-23,4	23,4	13

Idèn bakom 1-S Wilcoxon testet

- H_0: Om medianen av en symmetrisk fördelning var 0 borde W+ och W- vara ungefär lika stora.
- H_0 förkastas när t.ex många höga rang tillhör positiva värden och många låga rang tillhör negativa värden
- OBS!: H_0 förkastas om testvariabeln har ett *lågt* värde! ... eftersom W=min(W⁺,W⁻)



Wilcoxon test - tabell

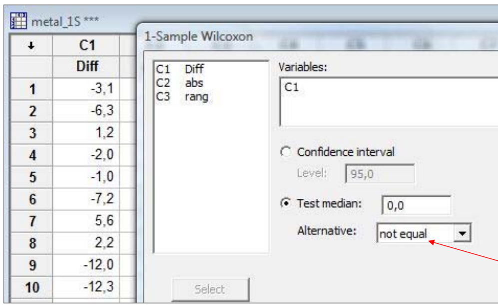
Critical values **Wilcoxon**

One-sided	Two-sided	$n = 5$	$n = 6$	$n = 7$	$n = 8$	$n = 9$	$n = 10$
$P = .05$	$P = .10$	1	2	4	6	8	11
$P = .025$	$P = .05$		1	2	4	6	8
$P = .01$	$P = .02$			0	2	3	5
$P = .005$	$P = .01$				0	2	3
One-sided	**Two-sided**	$n = 11$	$n = 12$	$n = 13$	$n = 14$	$n = 15$	$n = 16$
$P = .05$	$P = .10$	14	17	21	26	30	36
$P = .025$	$P = .05$	11	14	17	21	25	30
$P = .01$	$P = .02$	7	10	13	16	20	24
$P = .005$	$P = .01$	5	7	10	13	16	19
One-sided	**Two-sided**	$n = 17$	$n = 18$	$n = 19$	$n = 20$	$n = 21$	[illegible]
$P = .05$	$P = .10$	41	47	54	60	68	[illegible]
$P = .025$	$P = .05$	35	40	46	52	[illegible]	[illegible]
$P = .01$	$P = .02$	28	33	38	43	[illegible]	56
$P = .005$	$P = .01$	23	28	32	37	[illegible]	49
One-sided	**Two-sided**	$n = 23$	$n = 24$	$n = 25$	[illegible]	$n = 27$	$n = 28$
$P = .05$	$P = .10$	83	92	[illegible]	110	120	130
$P = .025$	$P = .05$	73	81	90	98	107	117
$P = .01$	$P = .02$	62	69	77	85	93	102
$P = .005$	$P = .01$	55	68	68	76	84	92

Beräknas med hjälp av kombinatorik

Minitab

Stat / Nonparametrics / 1-Sample Wilcoxon



tvåsidigt

Resultat

Results for: metal_1S

Wilcoxon Signed Rank Test: Diff

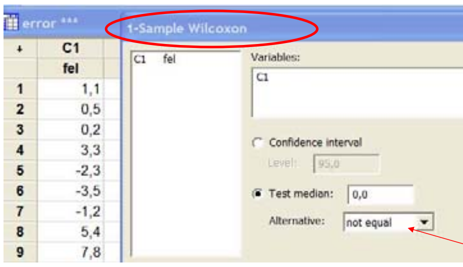
Test of median = 0,000000 versus median not = 0,000000

	N	N for Test	Wilcoxon Statistic	P	Estimated Median
Diff	13	13	16,0	0,043	-4,100

Nollhypotesen förkastas.
Medianen skiljer sig signifikant från 0.

Minitab

Stat / Nonparametrics / 1-Sample Wilcoxon



tvåsidigt

Jag vill veta om medianen av ett mätfel är noll (eller om det finns ett systematiskt fel)

Resultat

Wilcoxon Signed Rank Test: fel

Test of median = 0,000000 versus median not = 0,000000

	N	N for Test	Wilcoxon Statistic	P	Estimated Median
fel	10	10	39,0	0,262	1,750

H_0 förkastas inte (P>0.05). Medianen kan vara noll. Ett systematiskt fel kan inte bevisas.

Om antalet värden är stort (n>25)

Under H_0: hälften av den totala rangsumman

$$W^+ \in AsN\left[E(W^+), \sqrt{V(W^+)}\right]$$

$$E(W^+) = \frac{1}{2} \cdot \frac{n \cdot (n+1)}{2} \qquad V(W^+) = \frac{n \cdot (n+1) \cdot (2n+1)}{24}$$

$$z = \frac{W^+ - E(W^+)}{\sqrt{V(W^+)}} = \frac{W^+ - \frac{n \cdot (n+1)}{4}}{\sqrt{\frac{n \cdot (n+1) \cdot (2n+1)}{24}}}$$

$$\Omega_{krit} = \left\{ |z| > \lambda_{\alpha/2} \right\} \qquad \textit{tvåsidigt}$$

"större än" stämmer, vi testar ju W^+

C) Wilcoxon-Signed Rank test för parade stickprov

- två parade stickprov (t. ex före/efter)
- Finns en förskjutning (”location shift”) mellan båda fördelningar eller är de identiska (H_0)
- praktiskt taget detsamma som One-Sample Wilcoxon test, se nere
- Exempel: aluminiumhalt av träd i en förorenad areal i august månad och november månad →

Laureysens, I., R. Blust, L. De Temmerman, C. Lemmens and R. Ceulemans. 2004. *Clonal variation in heavy metal accumulation and biomass production in a poplar coppice culture. I. Seasonal variation in leaf, wood and bark concentrations. Environ. Pollution 131: 485-494.*

Rangordna de absoluta differenserna!

Aug	Nov	Diff	abs	rang
8,1	11,2	-3,1	3,1	6
10,0	16,3	-6,3	6,3	9
16,5	15,3	1,2	1,2	3
13,6	15,6	-2,0	2,0	4
9,5	10,5	-1,0	1,0	2
8,3	15,5	-7,2	7,2	10
18,3	12,7	5,6	5,6	8
13,3	11,1	2,2	2,2	5
7,9	19,9	-12,0	12,0	11
8,1	20,4	-12,3	12,3	12
8,9	14,2	-5,3	5,3	7
12,6	12,7	-0,1	0,1	1
13,4	36,8	-23,4	23,4	13

Idèn bakom testet

- Om båda fördelningar var identiska borde ungefär hälften av de parvisa differenserna vara positiva och den andra hälften vara negativa
- Dessutom borde positiva och negativa differenser av samma storlek vara lika sannolika
- Vi förväntar oss alltså, under H_0, att alla positiva differenser får samma rangsumma som alla negativa differenser
- Vi genomför alltså ett 1-S Wilcoxon test för de parvisa differenserna (ganska lika Paired t-test)

Testvariabel: samma som för 1-S Wilcoxon

$$W^+ = \sum_i R_i^+$$

$$W^- = \sum_i R_i^-$$

$$W^+ = 3+8+5 = 16$$

$$W^- = 6+9+4+2+10+11+12+7+1+13 = 75$$

$$W = \min\left(W^+, W^-\right) = W^+ = 16 \quad \textit{tvåsidigt}$$

$$W = W^- \quad \textit{lower tail}$$

$$W = W^+ \quad \textit{upper tail}$$

$$\Omega_{krit} = \left\{ W < W_0 \right\} \quad \textit{tabell}$$

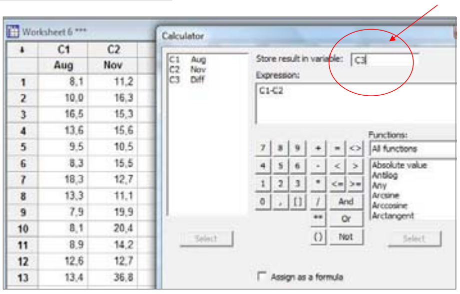
rang
6
9
3
4
2
10
8
5
11
12
7
1
13

Minitab

- Kör One-Sample Wilcoxon på differenserna!

Calc / Calculator

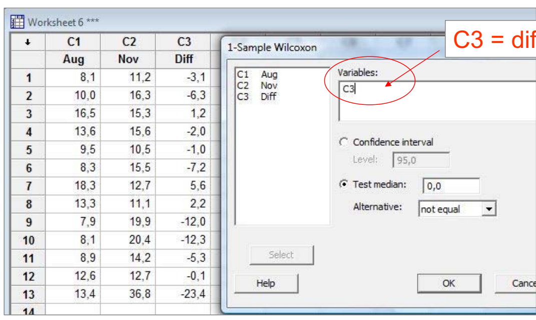
Differens till C3



Mintab

Stat / Nonparametrics / 1-Sample Wilcoxon

C3 = differens



Resultat

***Wilcoxon Signed Rank Test**: Diff*

Test of median = 0,000000 versus median not = 0,000000

	N	N for Test	Wilcoxon Statistic	P	Estimated Median
Diff	13	13	16,0	0,043	-4,100

W

D) Mann-Whitney U-test

- H_0: kommer två *oberoende* (icke-parade) stickprov från identiska fördelningar?
- H_a: det finns en förskjutning ("location shift")
- motsvarar Student's t-test för icke-normala PDF
- Exempel: två stickprov A och B
  - A: 25 26 27 31
  - B: 28 29 32 35
- Man ordnar alla värden enligt deras storlek och räknar hur många värden av A kommer före varje B-värdet →

Mann-Whitney U-test

värde	25	26	27	28	29	31	32	35
sample	A	A	A	B	B	A	B	B
bidrag till U				3	3		4	4

3 värden från stickprov A kommer före detta B

4 värden från stickprov A kommer före detta B

$$U_B = u_1 + u_2 + u_3 + u_4 = 3+3+4+4 = 14 \qquad eller$$
$$U_A = u_1 + u_2 + u_3 + u_4 = 0+0+0+2 = 2$$

Testvariabel för U-testet

$$U_B = u_1 + u_2 + u_3 + u_4 = 3+3+4+4 = 14 \qquad eller$$
$$U_A = u_1 + u_2 + u_3 + u_4 = 0+0+0+2 = 2$$

Mycket höga eller mycket små värden för U tyder på en skillnad mellan fördelningarna för population A respektive population B, som här:

värde	25	26	27	28	29	31	32	35
sample	A	A	A	A	B	B	B	B
bidrag till U					4	4	4	4

Kritiska området

- Mycket höga eller mycket små värden för U tyder på en skillnad mellan fördelningarna för population A respektive population B

$$two-tailed:$$
$$U = \min(U_A, U_B)$$
$$\Omega_{krit} = \left\{ U < U_0 \quad med \quad P(U \le U_o) = \alpha/2 \right\} \quad tabell$$

$$one-tailed:$$
$$\Omega_{krit} = \left\{ U_A < U_0 \quad med \quad P(U_A \le U_o) = \alpha \right\} \quad tabell$$

Mann-Whitney's U hänger ihop med Wilcoxon's rank sum W

$$U_A = n_1 \cdot n_2 + \frac{n_1 \cdot (n_1+1)}{2} - W_A$$
$$U_B = n_1 \cdot n_2 + \frac{n_2 \cdot (n_2+1)}{2} - W_B$$
$$U_A + U_B = n_1 \cdot n_2$$

I praktiken används ofta denna formel.

U hänger ihop med W

värde	25	26	27	28	29	31	32	35
sample	A	A	A	B	B	A	B	B
bidrag till U				3	3		4	4
rang	1	2	3	4	5	6	7	8

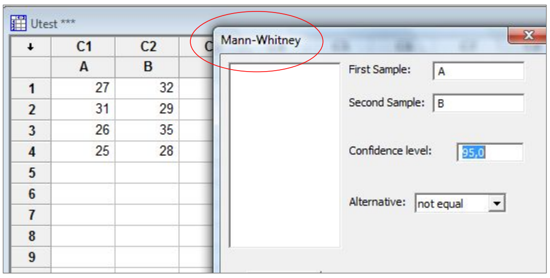
$$\left.\begin{aligned} W_A &= 1+2+3+6=12 \\ W_B &= 4+5+7+8=24 \end{aligned}\right\} \text{ vi utgår från Wilcoxons rangsumma}$$

$$\left.\begin{aligned} U_A &= n_1 \cdot n_2 + \frac{n_1 \cdot (n_1+1)}{2} - W_A = 16 + \frac{20}{2} - 12 = 14 \\ U_B &= n_1 \cdot n_2 + \frac{n_2 \cdot (n_2+1)}{2} - W_B = 16 + \frac{20}{2} - 24 = 2 \end{aligned}\right\} \text{ som förut}$$

$$U_A + U_B = n_1 \cdot n_2 = 16$$

Minitab

Stat / Nonparametrics / Mann-Whitney



Resultat

Mann-Whitney Test and CI: A; B

N Median
A 4 26,500
B 4 30,500

Point estimate for ETA1-ETA2 is -3,500
97,0 Percent CI for ETA1-ETA2 is (-9,998;2,998)
W = 12,0
Test of ETA1 = ETA2 vs ETA1 not = ETA2 is significant at ***0,1124***

median

H_0 förkastas inte, p för stort

Ingen signifikant skillnad, fördelningarna kan vara identiska.

Om antalet värden är stort (n>25)

$$z = \frac{U - m_U}{\sigma_U} \in N(0,1)$$

$$m_U = \frac{n_1 \cdot n_2}{2} \qquad \sigma_U = \sqrt{\frac{n_1 \cdot n_2 \cdot (n_1 + n_2 + 1)}{12}} \quad (utan\ "ties")$$

$$\Omega_{krit} = \left\{ |z| > \lambda_{\alpha/2} \right\} \qquad tvåsidigt$$

Förutsättningar och styrka

- Oberoende slumpmässiga stickprov
- (Lite) mindre styrka jämfört med Student's t ifall populationerna är normalfördelade
- Större styrka än Student's t för många andra fördelningar

E) Kruskal-Wallis test

- Ekvivalent till One-Way ANOVA
- Antalet populationer (treatments) = k
- H_0: alla k populationer är identiska
- H_a: minst två populationer har olika positioner ("location shift")
- Kräver inte normalfördelade populationer
- Recept: Blanda alla $n=n_1+n_2+ \ldots n_k$ observationer och rangordna från 1 till n

Kruskal-Wallis test

- Blanda alla $n=n_1+n_2+\ldots n_k$ observationer och rangordna från 1 till n

Summa över alla rang

R_i = *summa av alla rang för stickprov i*

$\bar{R}_i = \frac{R_i}{n_i}$ *medelvärdet över alla rang för grupp i*

$\bar{R} = \frac{1}{n}\cdot\frac{n\cdot(n+1)}{2} = \frac{1}{2}(n+1)$ *medelvärdet över alla rang*

$V = \sum_{i=1}^{k} n_i \cdot (\bar{R}_i - \bar{R})^2$ *motsvarar SST i ANOVA*

H₀ gäller
→ alla R_i ungefär lika
→ V liten

testvariabel liknar V:

$$H = \frac{12\cdot V}{n\cdot(n+1)} = \frac{12}{n\cdot(n+1)}\cdot\sum_{i=1}^{k}\frac{R_i^2}{n_i} - 3(n+1)$$

Kritiska området

$V = \sum_{i=1}^{k} n_i \cdot (\bar{R}_i - \bar{R})^2$ *motsvarar SST i ANOVA*

$H = \frac{12\cdot V}{n\cdot(n+1)} = \frac{12}{n\cdot(n+1)}\cdot\sum_{i=1}^{k}\frac{R_i^2}{n_i} - 3(n+1)$ *testvariabel*

H_0 förkastas om H (alltså även V) är stor, i så fall avviker ju gruppmedelvärdena från det gemensamma medelvärdet.

Om H_0 gäller är testvariabeln H ungefär χ^2-fördelad, om alla n_i är tillräckligt stora (≥5). Nollhypotesen förkastas om H är stor:

$$\Omega_{krit} = \{H > \chi^2_\alpha(f)\} \qquad f = k-1 \qquad \textit{upper tail}$$

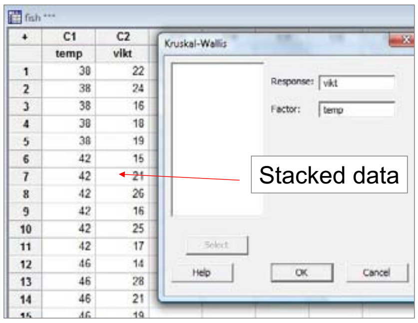
Vikten av fisk

Experiment: Vi vill veta om temperaturförhöjningen av havsvattnet i närheten av ett kärnkraftverk har ett inflytande på fiskarnas vikt:

Vikten av fisk			
38°F	42°F	46°F	50°F
22	15	14	17
24	21	28	18
16	26	21	13
18	16	19	20
19	25	24	21
	17	23	

Minitab

Stat / Nonparametrics / Kruskal-Wallis



Resultat

Kruskal-Wallis Test: vikt versus temp

temp	N	Median	Ave Rank	Z
38	5	19,00	11,6	0,04
42	6	19,00	11,5	0,00
46	6	22,00	14,0	1,11
50	5	18,00	8,4	-1,21
Overall	22		11,5	

df=k-1

H = 2,03 DF = 3 P = 0,566
H = 2,04 DF = 3 P = 0,564 (adjusted for ties)

Nollhypotesen (ingen förskjutning) kan inte förkastas. Ingen signifikant effekt av vattnets temperatur på vikten kan bevisas.