



AS01

Overlooked Tiny Numeric Representation Errors in SAS may Lead to Substantially Misleading p-Values

PhUSE EU 2018

Jens Klement, Ying Zhou, Lali Sandalic,
Holger Bohnemeier, David Haryanto



- Introduction
 - Computer precision
 - Everyday effects
- Examples
 - Flipping p-value
 - Wilcoxon Rank-Sum Test
 - Meaningless p-value
 - Wilcoxon Signed-Rank Test
 - Paired t-Test
- Solutions



- It is a truth universally acknowledged, that computerized systems are not indefinitely accurate.
- Computers (SAS, e.g.)...
 - operate in binary
 - only finite memory
 - for 8 byte floating point numbers: ~16 significant digits



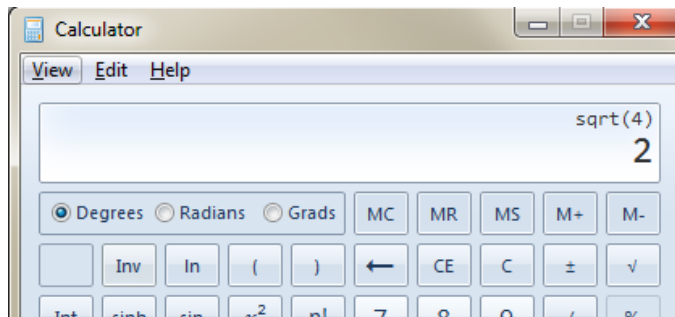
- SAS

- `diff = (0.3 - 0.2);`
`put diff; => 0.1`

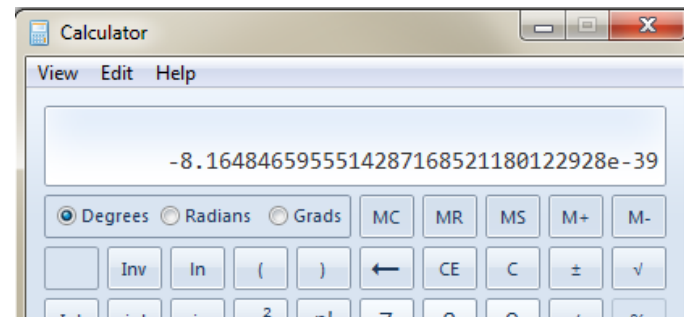
- `diff = (0.3 - 0.2 - 0.1);`
`put diff; =>  $\sim -2.78 \cdot 10^{-17}$`

- Windows calculator

- $\sqrt{4} = ?$



- $\sqrt{4} - 2 = ?$





- Discrepancies due to numeric representation errors (NRE)
 - No **ERROR**, no **WARNING**
 - value **looks** alright
 - Not due to user mistakes or bugs
 - inherent to how computers operate
- already simple operations with unremarkable numbers lead to mistakes



- **New:** NRE may affect the statistical results
- Often overlooked when calculating various statistics
- Three examples:
 - Wilcoxon Rank-Sum Test
 - Wilcoxon Signed-Rank Test
 - Paired t-Test

Mechanism: Relative magnitude of values matters



- Introduction
 - Computer precision
 - Everyday effects
- Examples
 - Flipping p-value
 - Wilcoxon Rank-Sum Test
 - Meaningless p-value
 - Wilcoxon Signed-Rank Test
 - Paired t-Test
- Solutions



Example 1 - Wilcoxon Rank-Sum Test

- Dummy study: “Double blind”
 - Comparison of products A and B
 - 20 subjects, 10 each
 - Measuring Para at t0 and t1
 - (Magnitude: ~ 0.1 g)
 - Analysing change of Para t1 – t0 (CHG) with rank-sum test
 - (Magnitude: ~ 1 mg)
 - NRE lead to flip of p-value



Example 1 - Wilcoxon Rank-Sum Test

Subject	Product	Para [g] (t0)	Para [g] (t1)	CHG [mg]	Rank
1	A	0.1226	0.1200	-2.6	5
2	A	0.1226	0.1194	-3.2	3.5
...	...	...	...	...	...
16	B	0.1227	0.1195	-3.2	3.5
17	B	0.1222	0.1210	-1.2	13
18	B	0.1203	0.1195	-0.8	15
19	B	0.1231	0.1224	-0.7	17
20	B	0.1211	0.1203	-0.8	15

Sum of ranks A: 78

$p = 0.0445$



Example 1 - Wilcoxon Rank-Sum Test

Expected

Subject	Product	Para [g] (t0)	Para [g] (t1)	CHG [mg]	Rank
1	A	0.1226	0.1200	-2.6	5
2	A	0.1226	0.1194	-3.2	3.5
...	...	...	...	...	...
16	B	0.1227	0.1195	-3.2	3.5
17	B	0.1222	0.1210	-1.2	13
18	B	0.1203	0.1195	-0.8	15
19	B	0.1231	0.1224	-0.7	17
20	B	0.1211	0.1203	-0.8	15

Sum of ranks A: 78

$p = 0.0445$



Example 1 - Wilcoxon Rank-Sum Test

Actual result

Subject	Product	Para [g] (t0)	Para [g] (t1)	CHG [mg]	Rank
1	A	0.1226	0.1200	-2.6	5
2	A	0.1226	0.1194	-3.2	4
...	...	...	...	...	...
16	B	0.1227	0.1195	-3.2	3
17	B	0.1222	0.1210	-1.2	13
18	B	0.1203	0.1195	-0.8	14
19	B	0.1231	0.1224	-0.7	17
20	B	0.1211	0.1203	-0.8	15.5

Sum of ranks A: **80**

p = **0.0638**



Example 1 - Wilcoxon Rank-Sum Test

Reason

Subject	Product	CHG [mg] (visible NRE)	CHG [mg] (expected)	Rank (NRE)
1	A	-2.6000000000000000	-2.6	5
2	A	-3.1999999999999990	-3.2	4
...	...	...	...	...
16	B	-3.2000000000000000	-3.2	3
17	B	-1.2000000000000000	-1.2	13
18	B	-0.8000000000000000	-0.8	14
19	B	-0.7000000000000000	-0.7	17
20	B	-0.7999999999999990	-0.8	15.5

$p(\text{expected}) = 0.0445$ $p(\text{NRE}) = \mathbf{0.0638}$



- Dummy study: “Sauna”
 - Comparison of placebo A and antiperspirant B
 - 10 subjects, each receive A and B
 - Weighing pad before and after sauna
 - (Magnitude: ~100 mg)
 - Analysing weight of sweat (CHG)
 - (Magnitude: ~ 10 mg)
 - Made up data: **No** effect within subjects
 - NRE lead to meaningless p-value



Example 2 - Wilcoxon Signed-Rank Test & t-Test

Subj.	Product	Tare [mg] (m0)	Gross [mg] (m1)	CHG [mg]	Paired diff. B - A
1	A	119.4	130.2	10.79999999999999900	1.4E-14
	B	119.2	130.0	10.79999999999999900	
2	A	123.3	137.1	13.80000000000000000	1.4E-14
	B	119.5	133.3	13.80000000000000000	
3	A	121.0	132.6	11.59999999999999900	1.4E-14
	B	120.3	131.9	11.60000000000000000	
...	...	...	...	...	...
	...	...	...	...	
8	A	127.9	138.7	10.79999999999999900	2.8E-14
	B	126.5	137.3	10.80000000000000000	
9	A	135.8	147.0	11.19999999999999900	2.8E-14
	B	167.2	178.4	11.20000000000000000	
10	A	126.3	140.2	13.89999999999999900	1.4E-14
	B	119.6	133.5	13.90000000000000000	

Signed-Rank Test: S=**27.5**, p = **0.0020**

t-Test: t = **9.0**, p < **0.0001**



Agenda

- Introduction
 - Computer precision
 - Everyday effects
- Examples
 - Flipping p-value
 - Wilcoxon Rank-Sum Test
 - Meaningless p-value
 - Wilcoxon Signed-Rank Test
 - Paired t-Test
- Solutions



Know that NRE can affect the outcome of various statistical test

- When relative magnitude is considered
- Rank based tests
- Comparison with reference value

Round intelligently

- After intermediate calculations
- If NRE is $\sim 10^{-17}$ round to $\sim 10^{-10}$, e.g.
- Or round to appropriate precision of parameter

Check validity of use of statistical models

- summary stats, mean effect and CI
- p-value isn't everything: Size of effect, clinical relevance

Use integers

- From example 1: multiply measured parameter by 10,000.
- Leading to higher precision (no NRE), but cumbersome.



FIN

Questions?

klementjens@prahs.com

zhouying@prahs.com



Backup slide: Example 1 - Wilcoxon Rank-Sum Test

Table 1 Different p-values of Wilcoxon Rank-Sum Test for Product Comparison Caused by Not-Handled versus Handled Numeric Representation Errors

Subject Number	Product	Para[g] (t0)	Para[g] (t1)	Default Change from BL (no additional round function)		Expected Change from BL (with additional round function)	
				Para[mg] (t1-t0)*1000 in Format 19.15	Rank	Para[mg] (t1-t0)*1000 in Format 19.15	Rank
1	A	0.1226	0.12	-2.6000000000000000	5	-2.6000000000000000	5
2	A	0.1226	0.1194	-3.1999999999999990	4	-3.2000000000000000	3.5
...							
16	B	0.1227	0.1195	-3.2000000000000000	3	-3.2000000000000000	3.5
17	B	0.1222	0.121	-1.2000000000000000	13	-1.2000000000000000	13
18	B	0.1203	0.1195	-0.8000000000000000	14	-0.8000000000000000	15
19	B	0.1231	0.1224	-0.7000000000000000	17	-0.7000000000000000	17
20	B	0.1211	0.1203	-0.7999999999999990	15.5	-0.8000000000000000	15
					Ranks Sum for A = 80 P= 0.0638	Ranks Sum for A = 78 P= 0.0445	

Numbers in bold red show different ranks between both calculations



Backup slide: Example 2 - Wilcoxon Signed-Rank Test & t-Test

Table 3 Misleading p-Values of Wilcoxon Signed-Rank Test and t-Test for Product Comparison Caused by Not-Handled Numeric Representation Errors

Subject Number	Product	Pad Tare Weight[mg] (m0)	Pad Gross Weight[mg] (m1)	Sweat Weight: Pad Net Weight [mg] (m1-m0) in Format 19.15	Paired Difference for Sweat Weight [mg] (Product B – Product A) in Format 19.15
1	A	119.4	130.2	10.799999999999900	0.000000000000014
	B	119.2	130.0	10.799999999999900	
2	A	123.3	137.1	13.800000000000000	0.000000000000014
	B	119.5	133.3	13.800000000000000	
3	A	121.0	132.6	11.599999999999900	0.000000000000014
	B	120.3	131.9	11.600000000000000	
...					
8	A	127.9	138.7	10.799999999999900	0.000000000000028
	B	126.5	137.3	10.800000000000000	
9	A	135.8	147.0	11.199999999999900	0.000000000000028
	B	167.2	178.4	11.200000000000000	
10	A	126.3	140.2	13.899999999999900	0.000000000000014
	B	119.6	133.5	13.900000000000000	

Wilcoxon Signed-Rank Test

$$S = (R^{(+)} - R^{(-)})/2 = (55-0)/2=27.5$$

$$P = 0.0020$$

t-Test

$$t = \frac{\bar{y} - \mu_0}{s/\sqrt{n}} = \frac{1.71 \times 10^{-14} - 0}{5.99 \times 10^{-15}/\sqrt{10}} = 9.0$$

$$P < 0.0001$$