Preliminary statistical analysis of the international eventing results 2013

Vrije Universiteit Brussel

Lausanne 28/1/14

Overview of the talk

Statistical analysis

- The data
- The statistical technique
- Analysis of the falls data (related to a fence)
- Analysis of results data

Fall statistics: Belgian situation

The major steps in a statistical analysis

- Define the interesting questions and factors, that might affect risk
- Query the FEI database and compose file with data
- Encode data for statistical analysis
- Perform the analysis
- Interpretation of the results

1. The questions addressed in this analysis

What are the factors that on their own and as a combination are correlated with good results, Refusals, EL, Horse Fall, Unseated Rider ?

Age of horse and rider, barely qualified, previous results, competing at international level since when, average number of successful / unsuccessful competitions per year at that level, number of competition as a combination, …

Are there parameters that significantly alter the probability of a fall?

Homework: questions from the audience during the seminar

2. The data file

Excel file with data on each international starter ➔ 142 000 rows and 650 columns

~~Identifications, Date, Officials, …~~

Profile of rider:
age
different horses
starts 1*, 2*, 3*, 4*
MER 1*, 2*, 3*, 4*

Profile horse:
age
different horses
starts 1*, 2*, 3*, 4*
MER 1*, 2*, 3*, 4*

Profile of competition
CCI or CIC
Star-level
SJ before XC
Speed
Distance

Result
Dressage: pen
XC: pen, EL (R, Fall) RET,WD
SJ: pen, EL, RET,WD

Profile of fall
38 columns
944 rows

Multidimensional highly correlated dataset

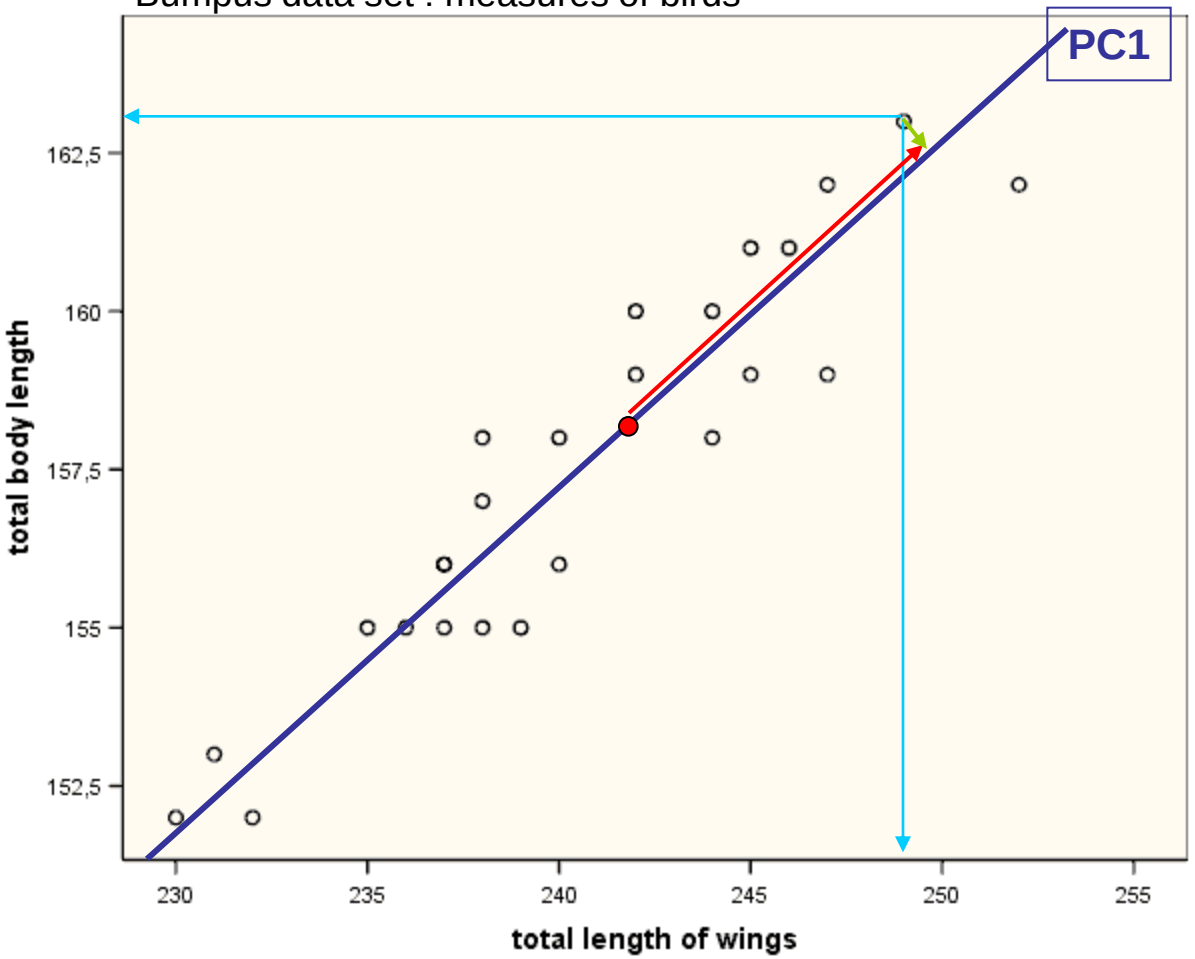
→ PCA, FA

Low dimensional uncorrelated dataset

File: 2004-2013 but focus on 2013

The principal component analysis = Data reduction due to correlations

Bumpus data set : measures of birds



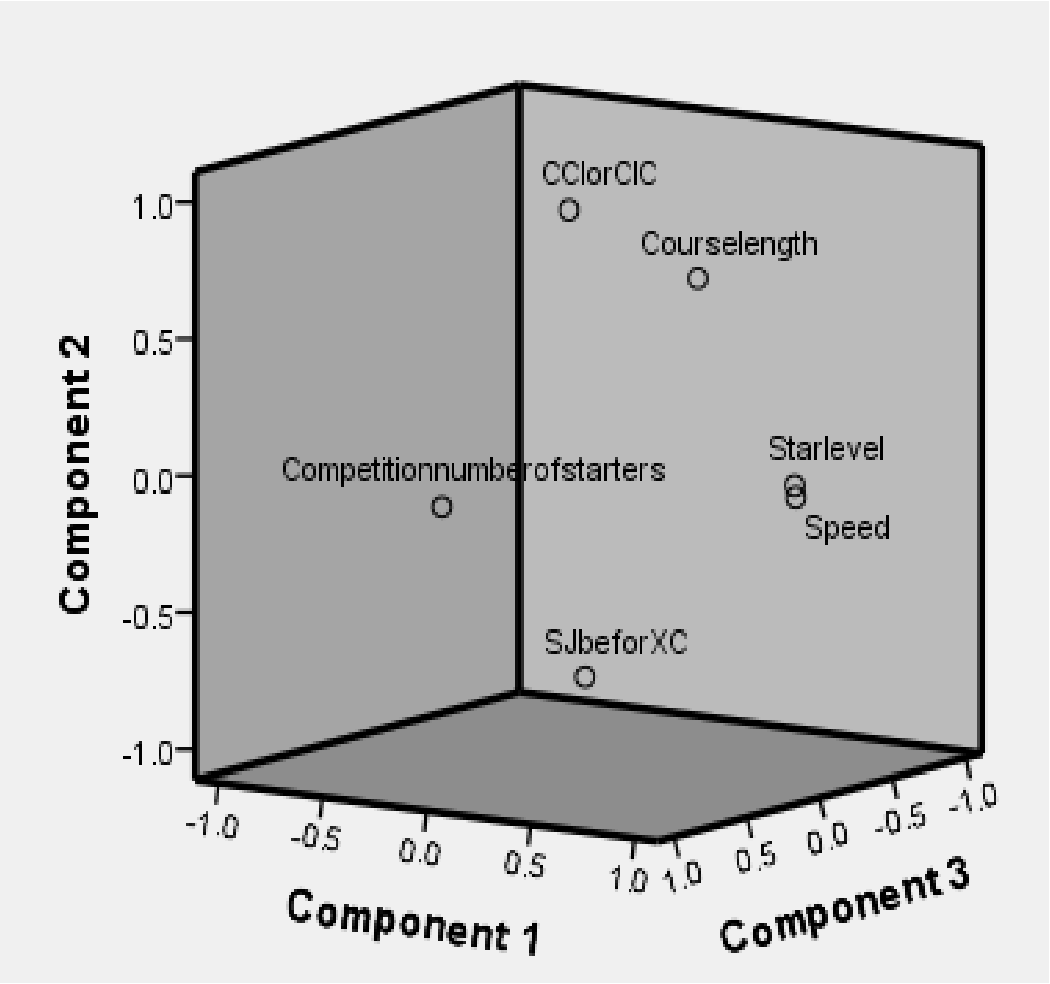
97% of variation in data is along PC1

Reduction of 2-dim data point (body length, length of wings)
To
1-dim data point=
Projection on PC1

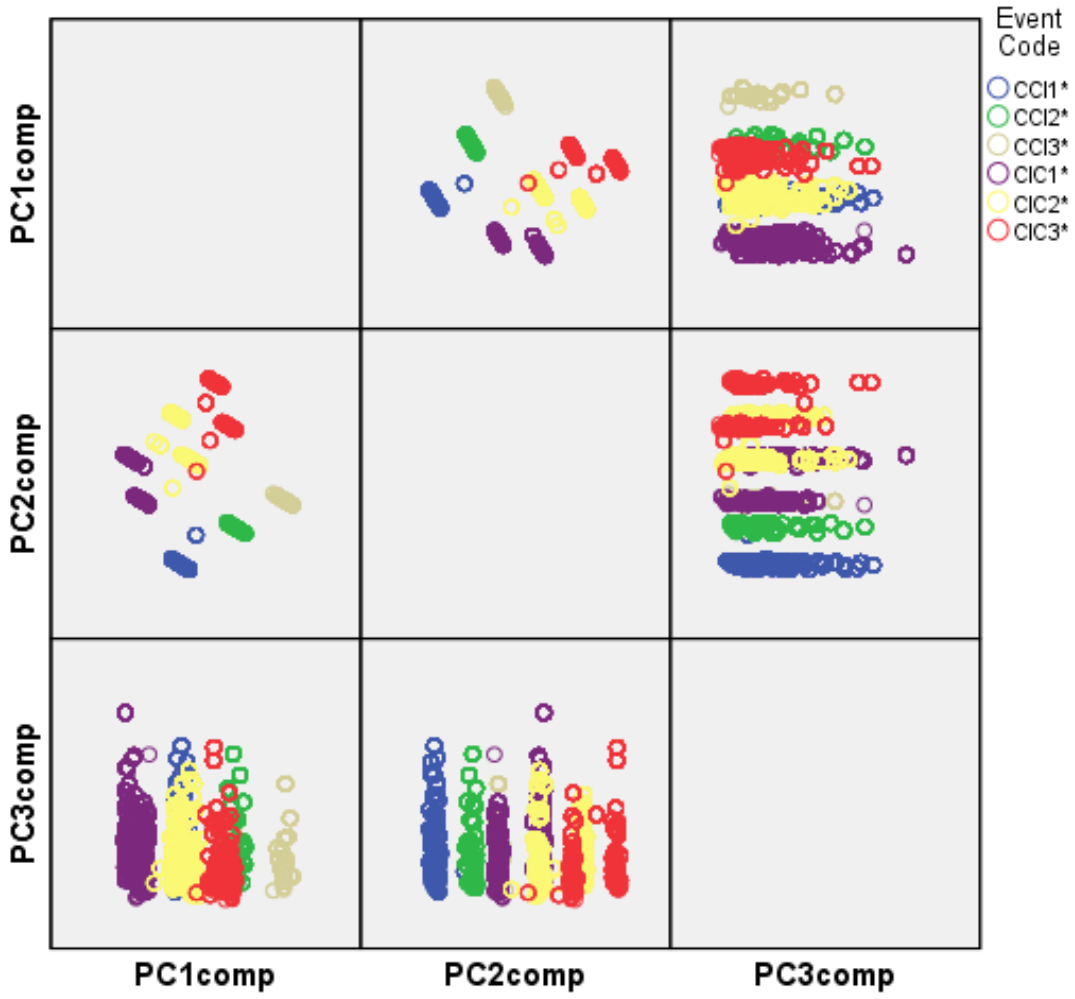
Only 3% loss of variation
! Noise !

Birds with a large body length do have a large length of wings + negation.

“Competition profile” as illustration of PCA- technique



Component Matrix[a]			
	Component		
	1	2	3
Starlevel	,834		
CCIorCIC		-,802	
SJbeforXC		,667	
Course length	,890		
Speed	,803		
Competition number of starters			,994



Component Matrix[a]			
	Component		
	1	2	3
Starlevel	,834		
CCIorCIC		-,802	
SJbeforXC		,667	
Course length	,890		
Speed	,803		
Competition number of starters			,994

Fall Analysis

Fall characteristics

	FALSE	TRUE	valid	%FALSE	%TRUE
Combination	380	478	858	0,4	0,6
Fence Type					
Option	4	854	858	1,0	0,0
Frangible	806	52	858	0,9	0,1
Portable	467	391	858	54,4	45,6
Related to water	739	205	944	78,3	21,7
Terrain	406	538	944	43,0	57,0
Ground Line	208	650	858	24,2	75,8
Off a Bend	633	311	944	67,1	32,9
Air Jacket worn	286	494	780	36,7	63,3
Air Jacket activated	510	434	944	54,0	46,0
Frangible break		7			
Horse fell or tread on Rider	900	44	944	95,3	4,7
Accident Type	653	291	944	69,2	30,8
Horse slipped?	859	85	944	91	9,0
Wind	824	120	944	87,3	12,7
Poor Visibility	943	1	944	99,9	0,1
Horse refused?					
Horse hit Fence?	463	481	944	49,0	51,0
Horse tipped portable?		5			
Horse broke fence?	828	30	858	96,5	3,5
Horse somersaulted?	818	40	858	95,3	4,7
Inexperience	548	396	944	58,1	41,9
Rider fatigue	940	4	944	99,6	0,4
Horse out of control	913	31	944	96,7	3,3
Horse too fast	831	113	944	88	12
Horse too slow	877	67	944	92,9	7,1
Light situation	905	39	944	95,9	4,1
Horse impaired by health	942	2	944	99,8	0,2
Horse fatigued	922	22	944	97,7	2,3
Horse distracted	932	12	944	98,7	1,3

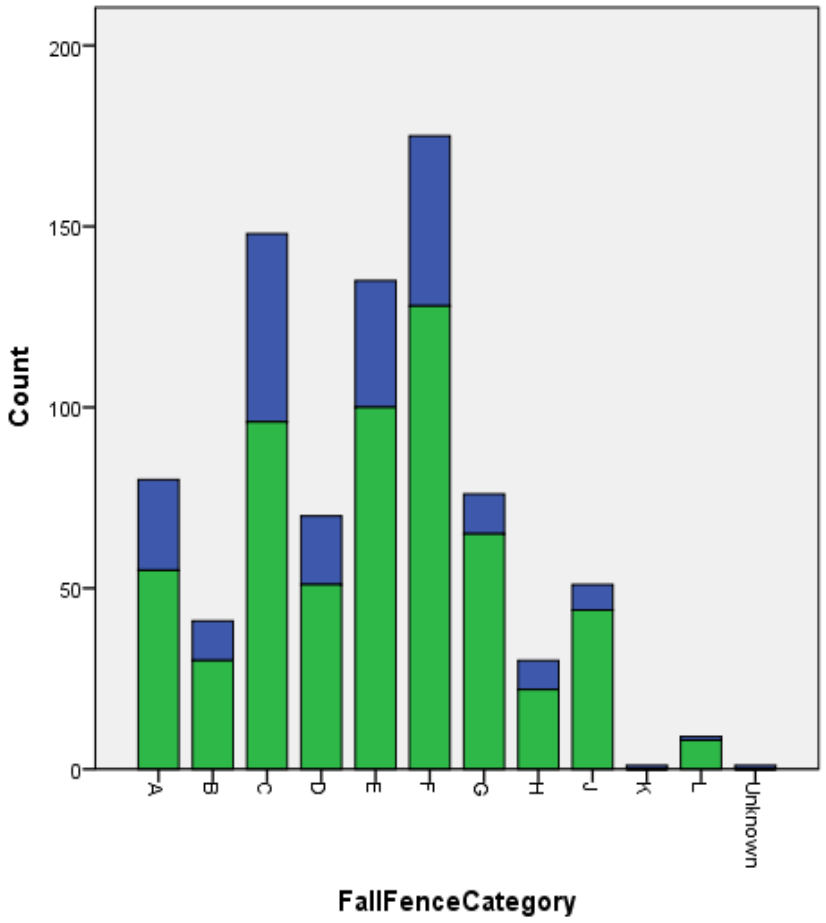
Five possible classes of falls related to a fence

	Component				
	1	2	3	4	5
Fall Fence Combination		,663			
Fall Fence Portable					,461
Fall Fence Related to water		,681			
Fall Fence Ground line					
Fall Fence Off a bend		,397			
Did horse fall on or tread on athlete	,418				
Did the horse slip					,437
Wind					
Did horse refuse	-,697				
Did horse hit the fence	,704				
Did horse break the fence					
Did horse somersault	,476				
Inexperience			,419	,407	
Athlete impaired by fatigue					
Horse out of control			,685		
Horse too fast			,730		
Horse too slow				,631	
Horse fatigued				,519	
Horse distracted					,544

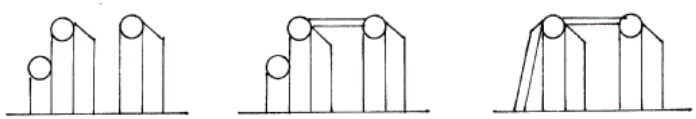
NB: 82 falls not related to fence are not included in this analysis

Falls and type of fences

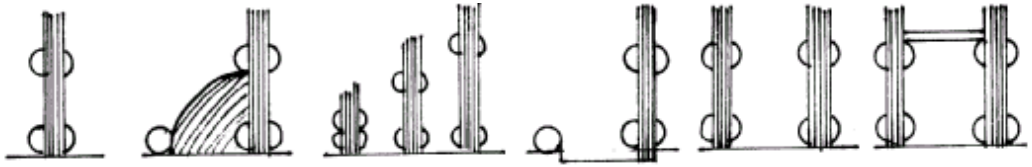
In total 817 falls related to fences



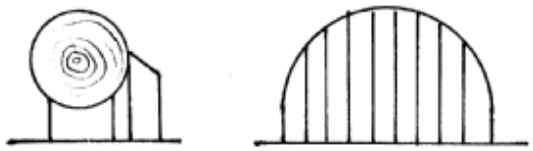
C. square spread



E. brush

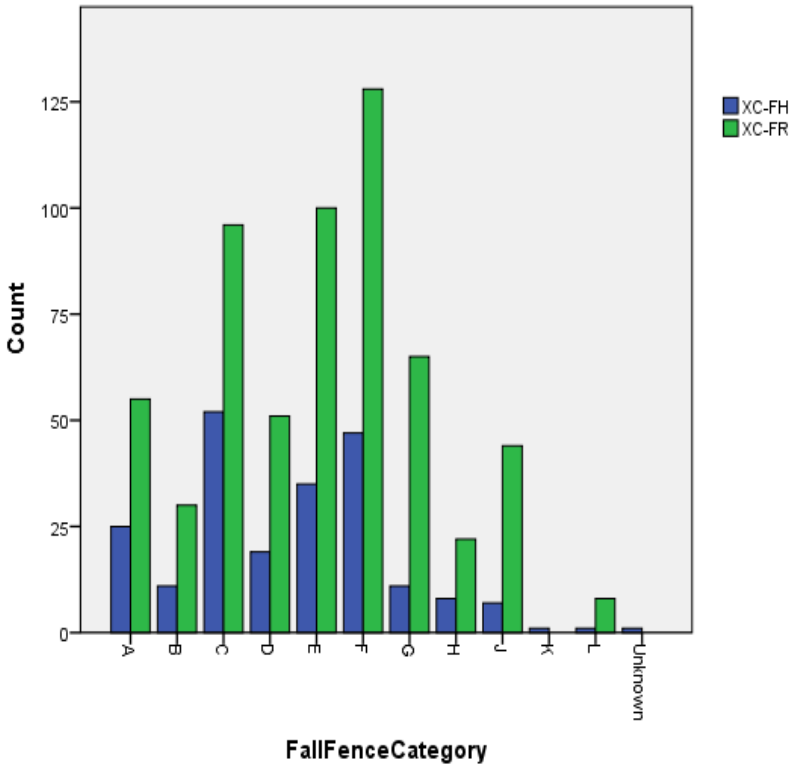


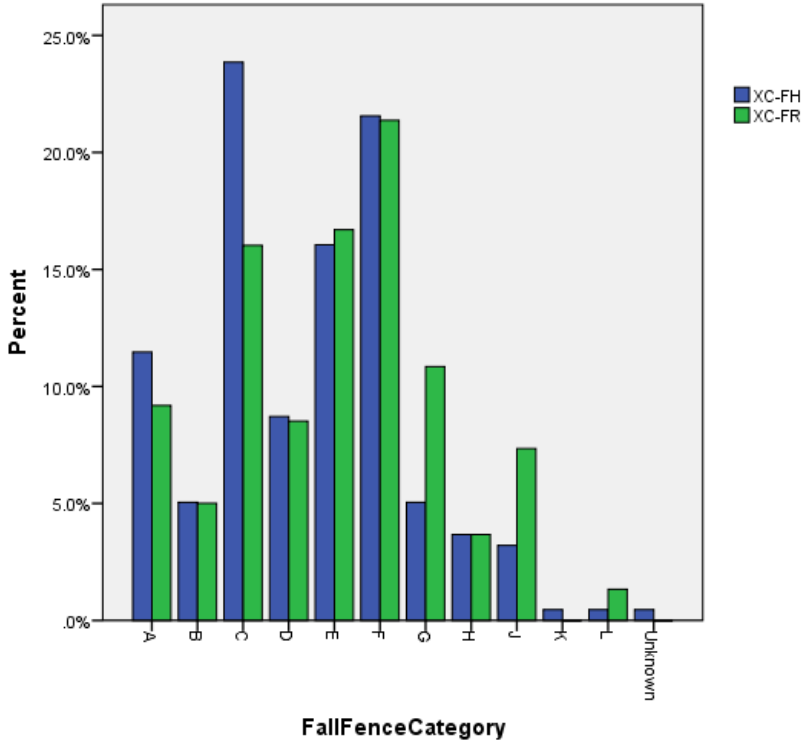
F. round



No normalization by the number of times the type of fence is jumped!

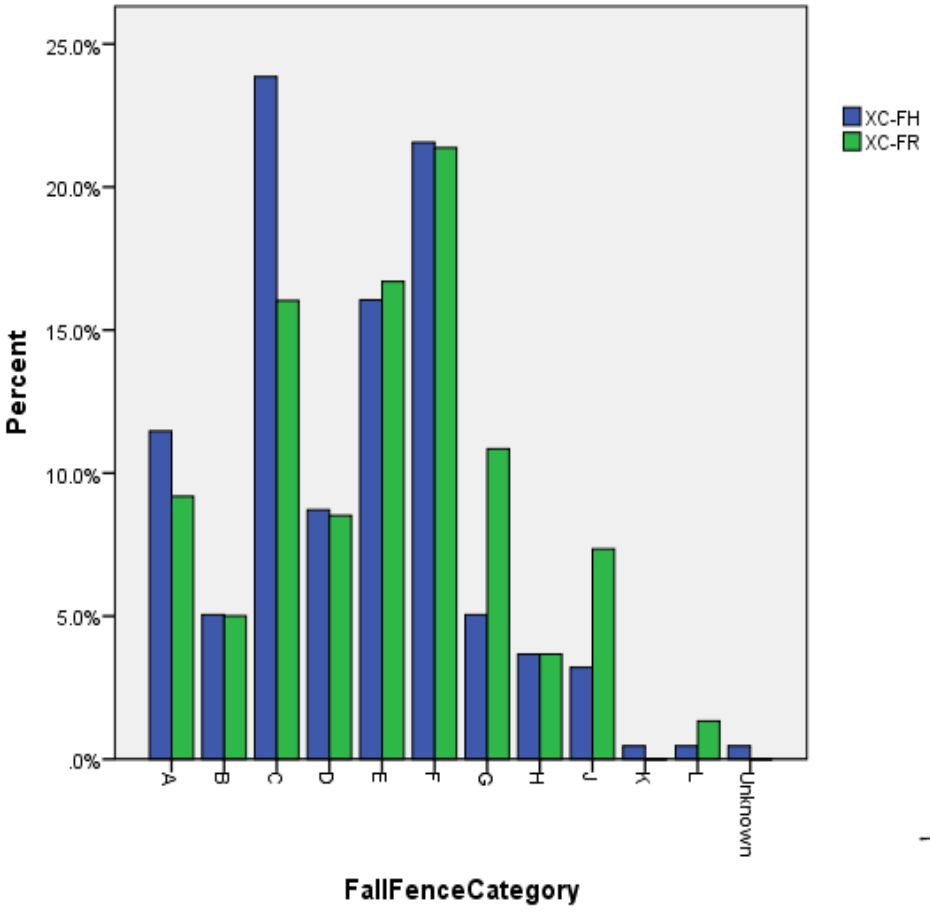
Unseated Riders: 599 (73%) ⟷ Horse Falls: 218 (27%)



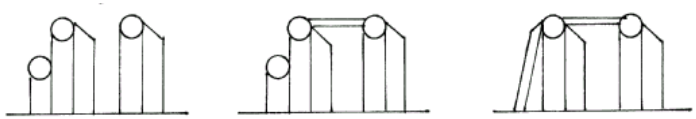


Significant difference between the distribution of Unseated riders and horse falls (chi-square p=0.017)

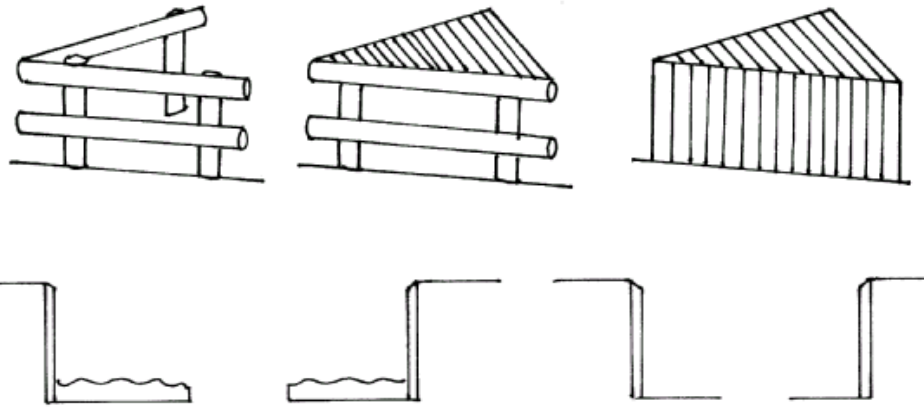
Distribution Unseated riders and horse falls over the fence types



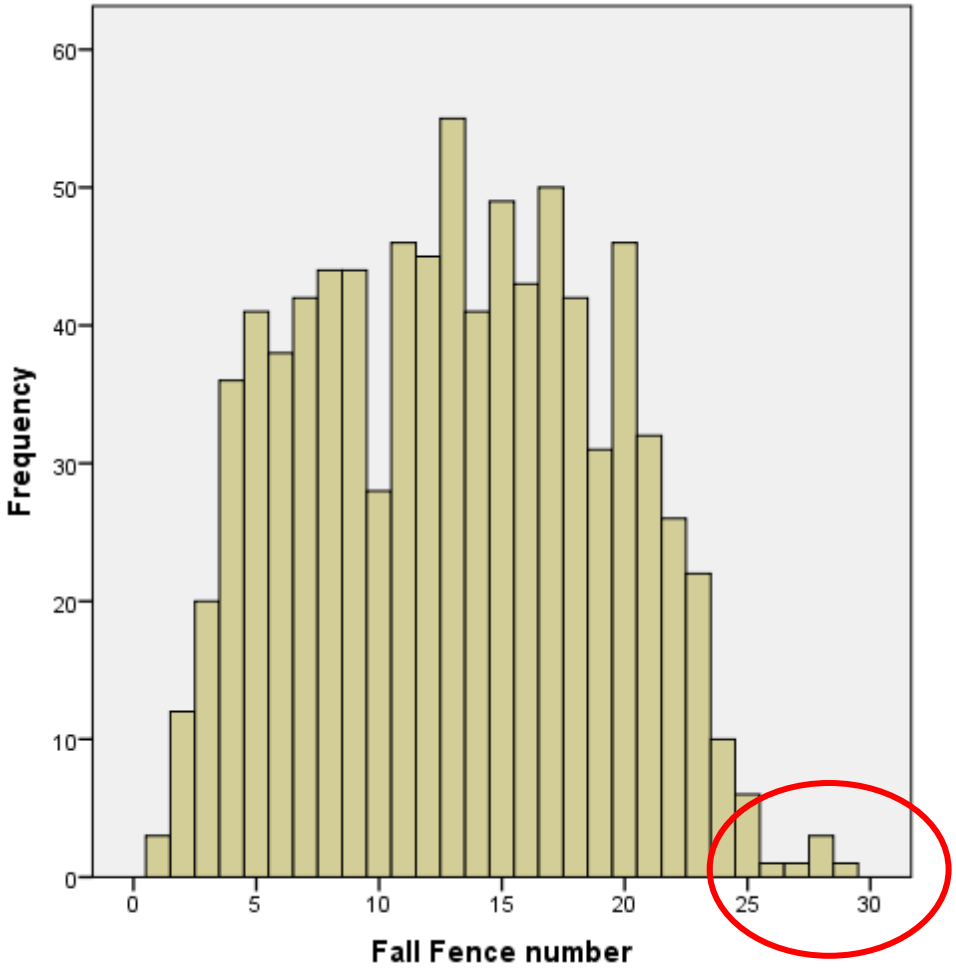
Relative more Horse Falls than Unseated Riders on type C



Relative more Unseated Riders than Horse Falls on type G (corner) and J(step)



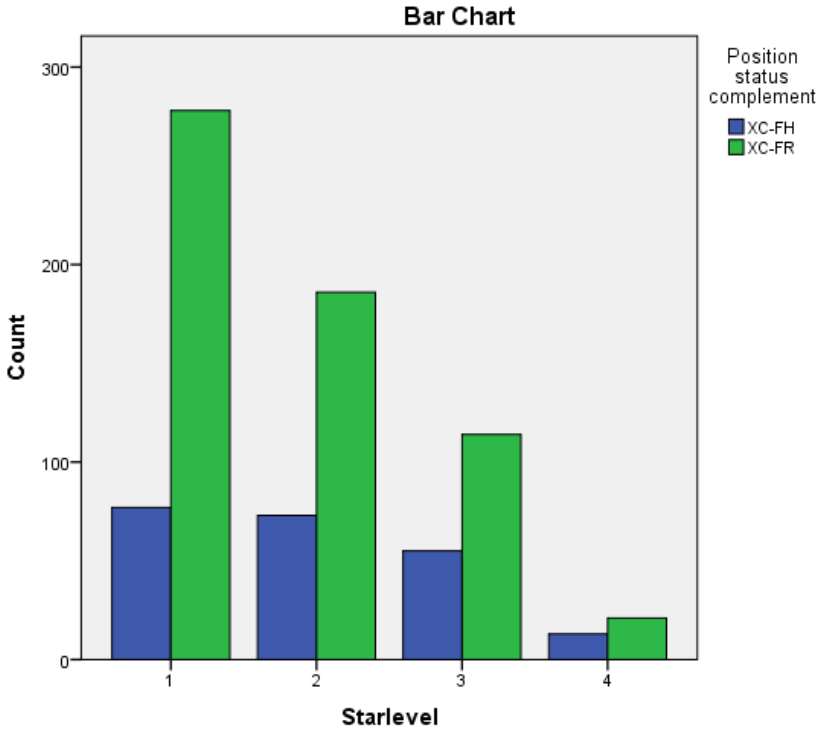
Number of fence where the fall occurs

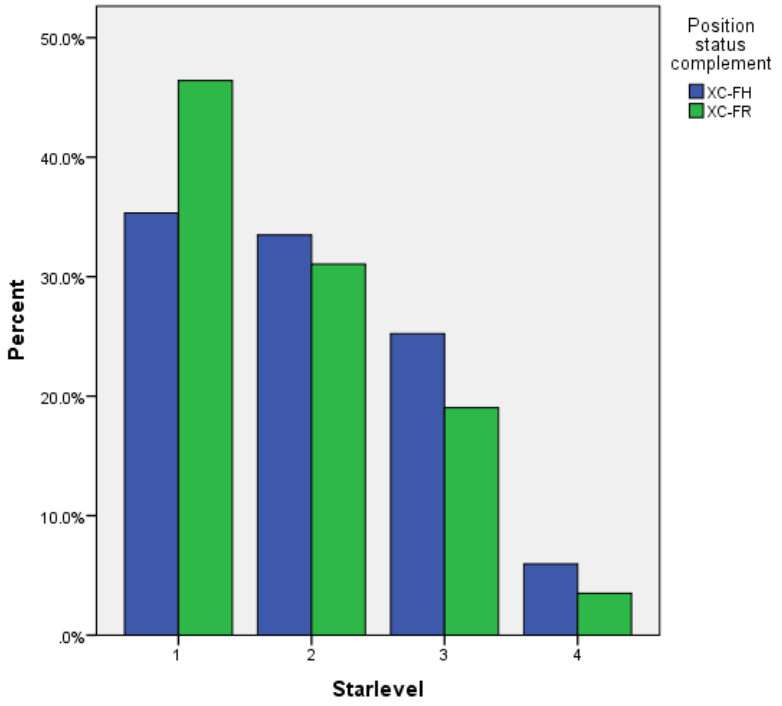


Total number of fences

No significant tendency
Fatigue of horse/rider

Falls and star-level

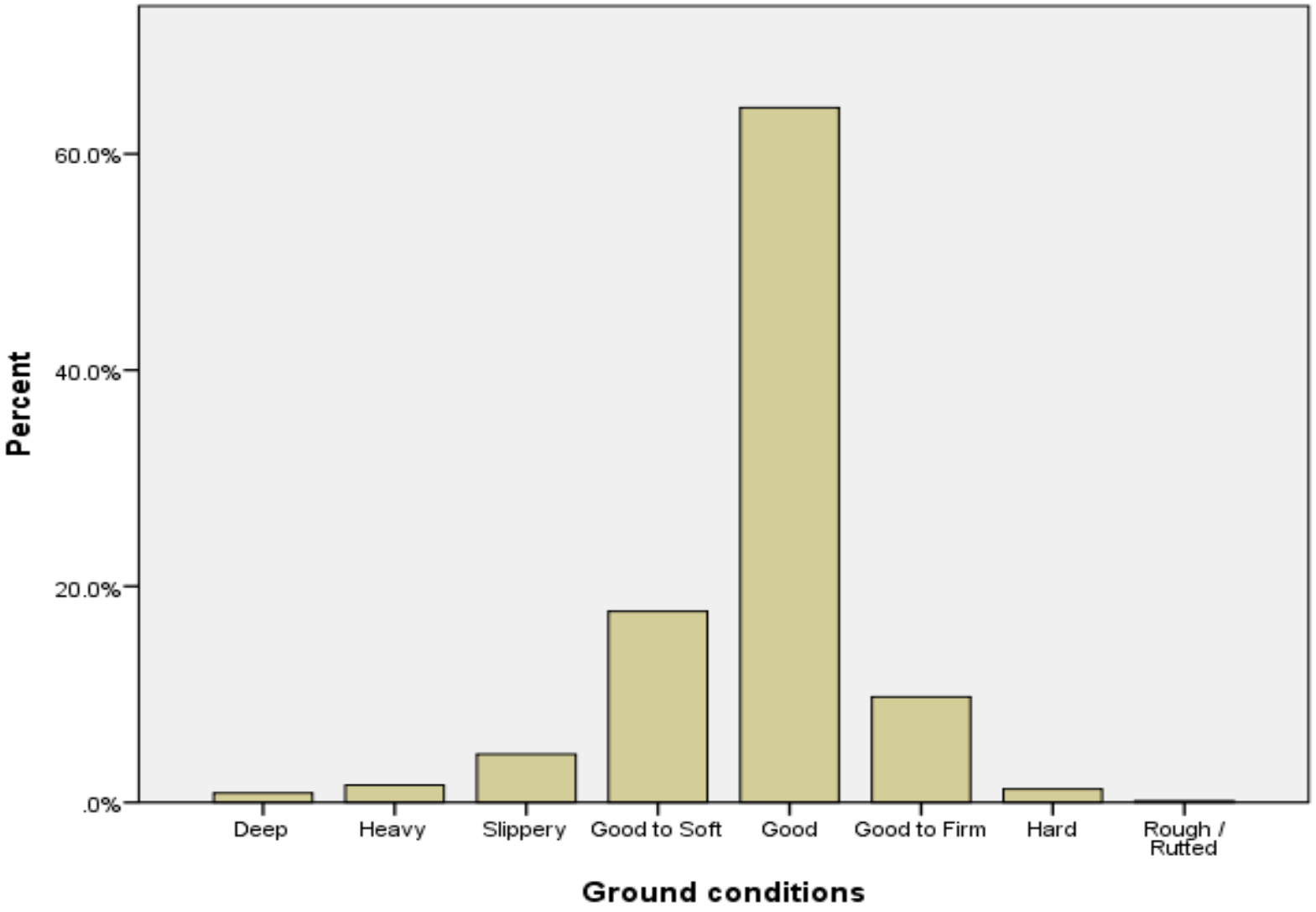




Significantly higher relative number of Unseated Riders in the 1* level

To do: Normalization with the number of starters per star level

Falls and ground conditions

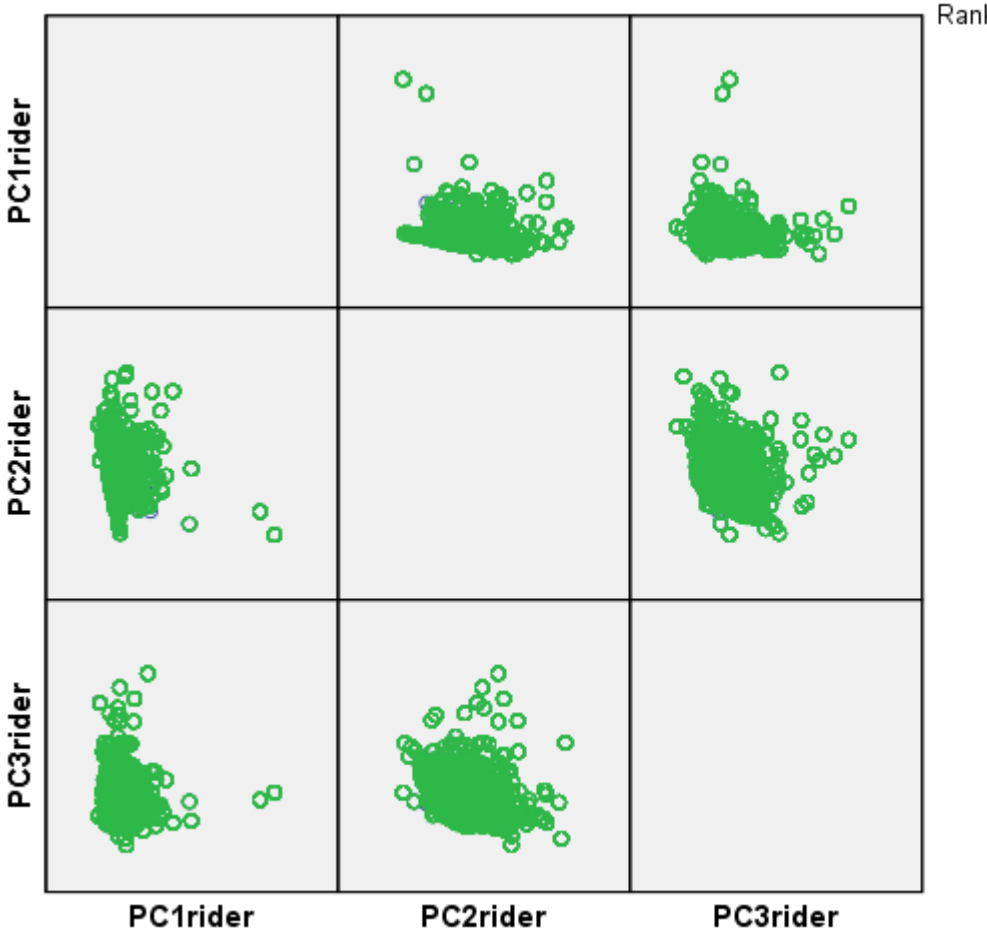


Results Analysis

Rider profile ⟷ Did Finish YES or NO

79% variation explained with 3 PC's

	Component		
	1	2	3
Athlete age at time of competition			
atstarts104to12			,654
Athlete number of starts 1* 2013			,928
atstarts204to12		,726	
Athlete number of starts 2* 2013		,798	
atstarts304to12	,765		
Athlete number of starts 3* 2013	,616	,613	
atstarts404to12	,922		
Athlete number of starts 4* 2013	,877		
atnumhor0412	,631	,629	
Athlete number of different horses ridden 2013		,604	
atMER10412		,625	,634
Athlete MER 1* 2013			,903
atMER20412		,693	
Athlete MER 2* 2013		,809	
atMER30412	,786		
Athlete MER 3* 2013	,618	,601	
atMER40412	,927		
Athlete MER 4* 2013	,880		



PC1: 3* and 4*
PC2: 2*, 3* and a lot of horses
PC3: 1*

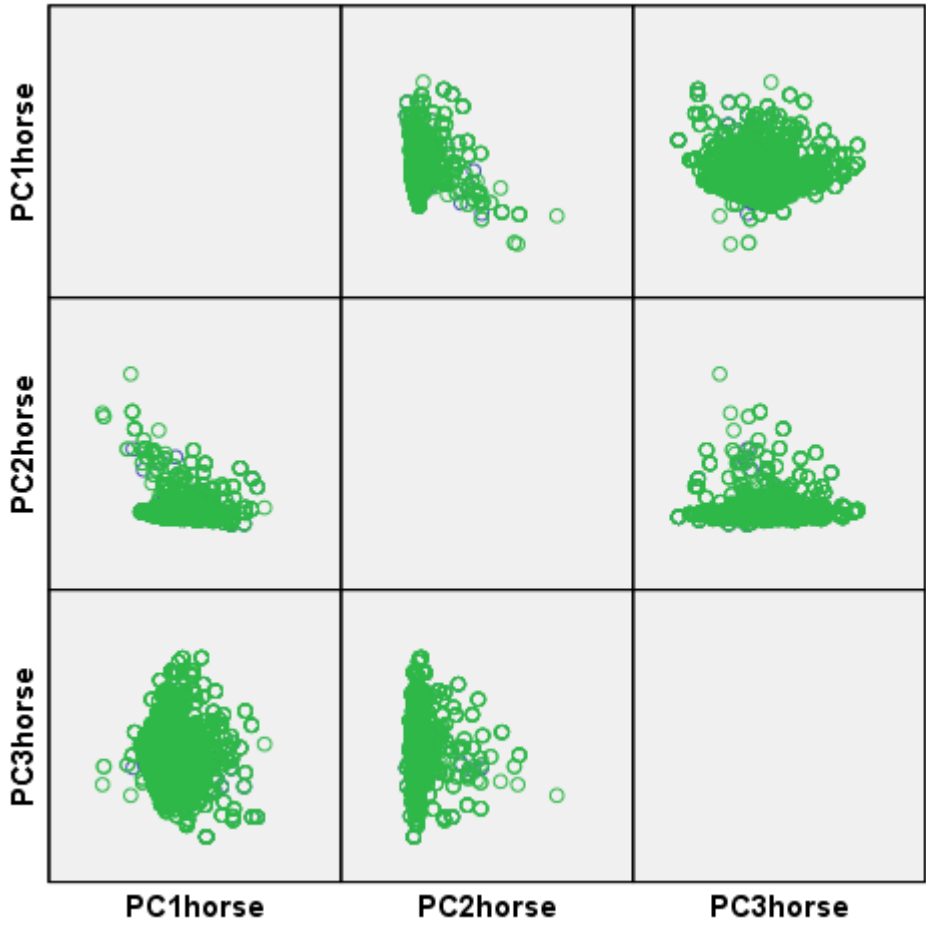
+ correlated with finish

NB: success rate, code 1-7,9-11

Horse profile ⟷ Did Finish YES or NO

Only 55% variation explained with 3 PC's

	Component		
	1	2	3
Horse age at time of competition			
hostart10412	,850		
Horse number of starts 1* 2013			-,757
hostart20412	,758		
Horse number of starts 2* 2013			,522
hostart30412		,798	
Horse number of starts 3* 2013			,592
hostart40412		,827	
Horse number of starts 4* 2013			
hoMER10412	,884		
Horse MER 1* 2013			-,611
hoMER20412	,714		
Horse MER 2* 2013			,557
hoMER30412		,821	
Horse MER 3* 2013			,565
hoMER40412		,720	
Horse MER 4* 2013			



PC1: "old" 1* and 2*
PC2: "old" 3*, 4*
PC3: "new" 2*, 3* and no 1*

Neg. correlated with finish

Preliminary summary

ResultCode		PC1rider	PC2rider	PC3rider	PC1horse	PC2horse	PC3horse
XC-clear	Mean						
	SE						
XC-20pen	Mean				-0.133		
	SE				0.025		
XC-40pen	Mean						
	SE						
XC-60pen	Mean		-0.318		-0.298		
	SE		0.060		0.084		
XC-EL-R	Mean		-0.351		-0.256		
	SE		0.023		0.035		
XC-EL-FR	Mean		-0.227	-0.193			
	SE		0.030	0.032			
XC-EL-FH	Mean			-0.188			0.220
	SE			0.061			0.083
XC-RET	Mean						
	SE						
XC-WD	Mean		0.180				0.176
	SE		0.039				0.055
Total	Mean	0.001	0.000	-0.001	-0.001	0.001	0.001
	SE	0.008	0.008	0.008	0.009	0.009	0.009

No experienced 1 and 2* horses

No 2 and 3* rider , Not many horses
No experienced 1 and 2* horses

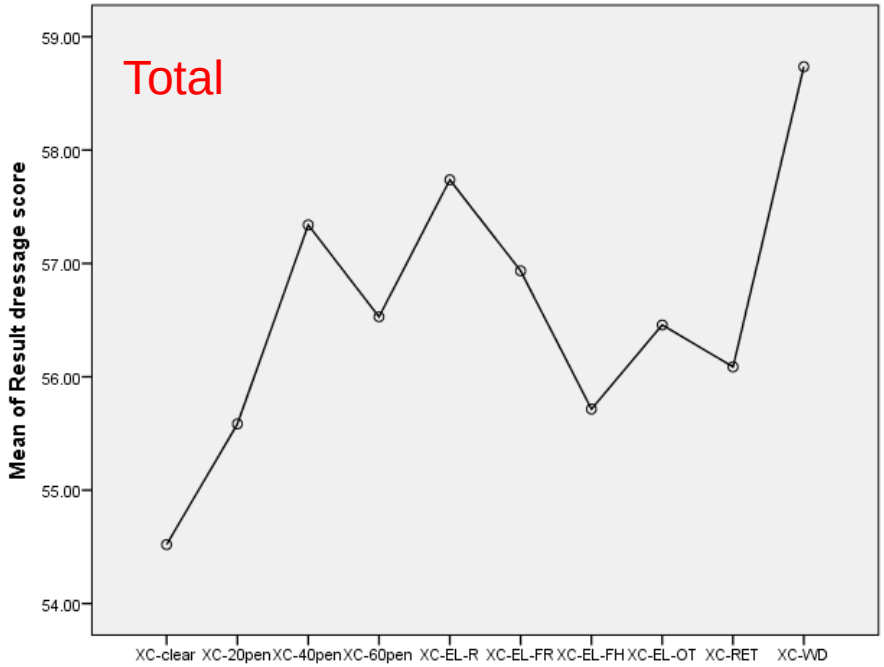
No 2 and 3* rider , Not many horses
No experienced 1 and 2* horses

No 1, 2 and 3* rider , Not many horses

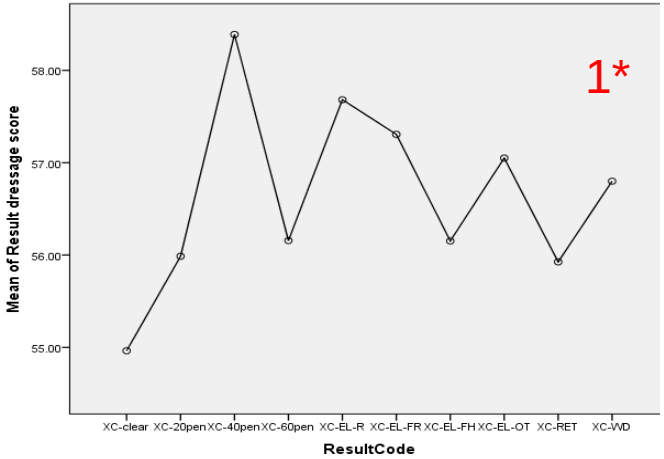
No 1* rider
Horse start in 2 and 3*

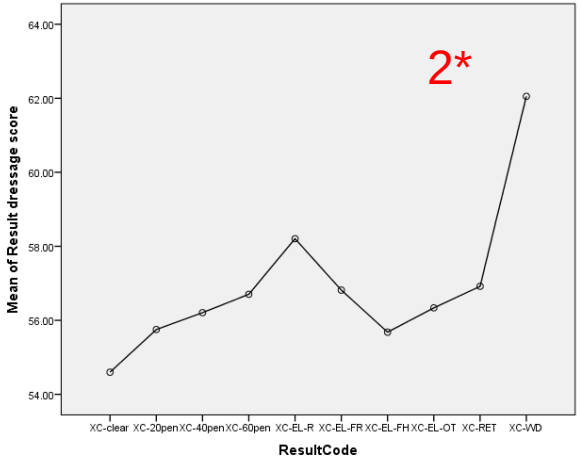
2and 3* rider with one horse
Horse start in 2 and 3*

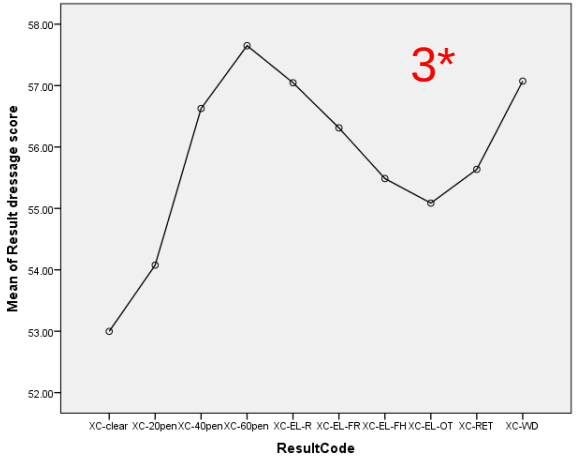
Dressage score ⟷ Result code, Starlevel



ResultCode	N	Mean	Std. Deviation
XC-clear	11549	54.52	7.62
XC-20pen	2087	55.58	7.64
XC-40pen	348	57.34	8.07
XC-60pen	150	56.53	8.31
XC-EL-R	825	57.74	7.33
XC-EL-FR	652	56.93	7.70
XC-EL-FH	233	55.72	8.11
XC-EL-OT	180	56.46	8.32
XC-RET	783	56.09	8.33
XC-WD	620	58.74	38.86
Total	17427	55.22	10.57

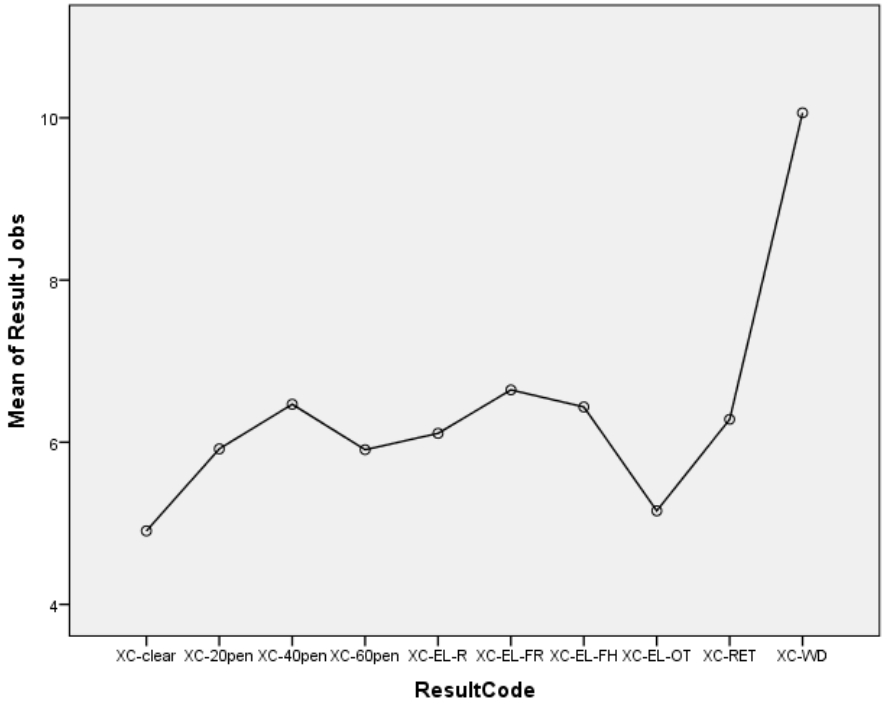






ANOVA: the ResultCode has an influence; no difference for different Star-levels

SJ score ⟷ **Result Code, Starlevel**



ResultCode	N	Mean	Std. Deviation
XC-clear	3847	4.91	5.446
XC-20pen	659	5.92	6.319
XC-40pen	94	6.47	6.676
XC-60pen	54	5.91	7.080
XC-EL-R	192	6.11	6.312
XC-EL-FR	127	6.65	6.279
XC-EL-FH	46	6.43	6.449
XC-EL-OT	52	5.15	5.507
XC-RET	191	6.28	5.618
XC-WD	227	10.06	8.803
Total	5489	5.42	5.935

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	8183.160[a]	29	282.178	8.321	.000
Intercept	31230.661	1	31230.661	920.918	.000
ResultCode	6578.092	9	730.899	21.552	.000
Starlevel	5.670	2	2.835	.084	.920
ResultCode * Starlevel	1115.353	18	61.964	1.827	.017
Error	185128.415	5459	33.913		
Total	354728.000	5489			
Corrected Total	193311.575	5488			

ANOVA: the ResultCode has an influence; no difference for different Star-levels

Age of Rider, Horse ⟷ Result Code, Star-level

Mean		
Starlevel	Horse age at time of competition	Athlete age at time of competition
1	9,67	28,68
2	10,37	31,01
3	11,64	34,10
4	12,62	36,52
Total	10,27	30,44

Age of horse and age of the athlete are significantly different in the different star-levels

Mean		
ResultCode	Horse age at time of competition	Athlete age at time of competition
XC-clear	10,19	30,48
XC-20pen	10,24	29,84
XC-40pen	10,36	29,92
XC-60pen	10,03	28,41
XC-EL-R	10,39	29,03
XC-EL-FR	10,60	29,33
XC-EL-FH	10,49	29,91
XC-EL-OT	10,64	33,04
XC-RET	10,66	32,15
XC-WD	10,62	32,91
Total	10,27	30,44

The mean age of the athletes is significantly In the different results classes; not the age of the horse.

Acknowledgements

FEI-eventing department: Giuseppe, Catrin, Nathalie

FEI-IT department: Jérôme Begey and Gaspard Dufour