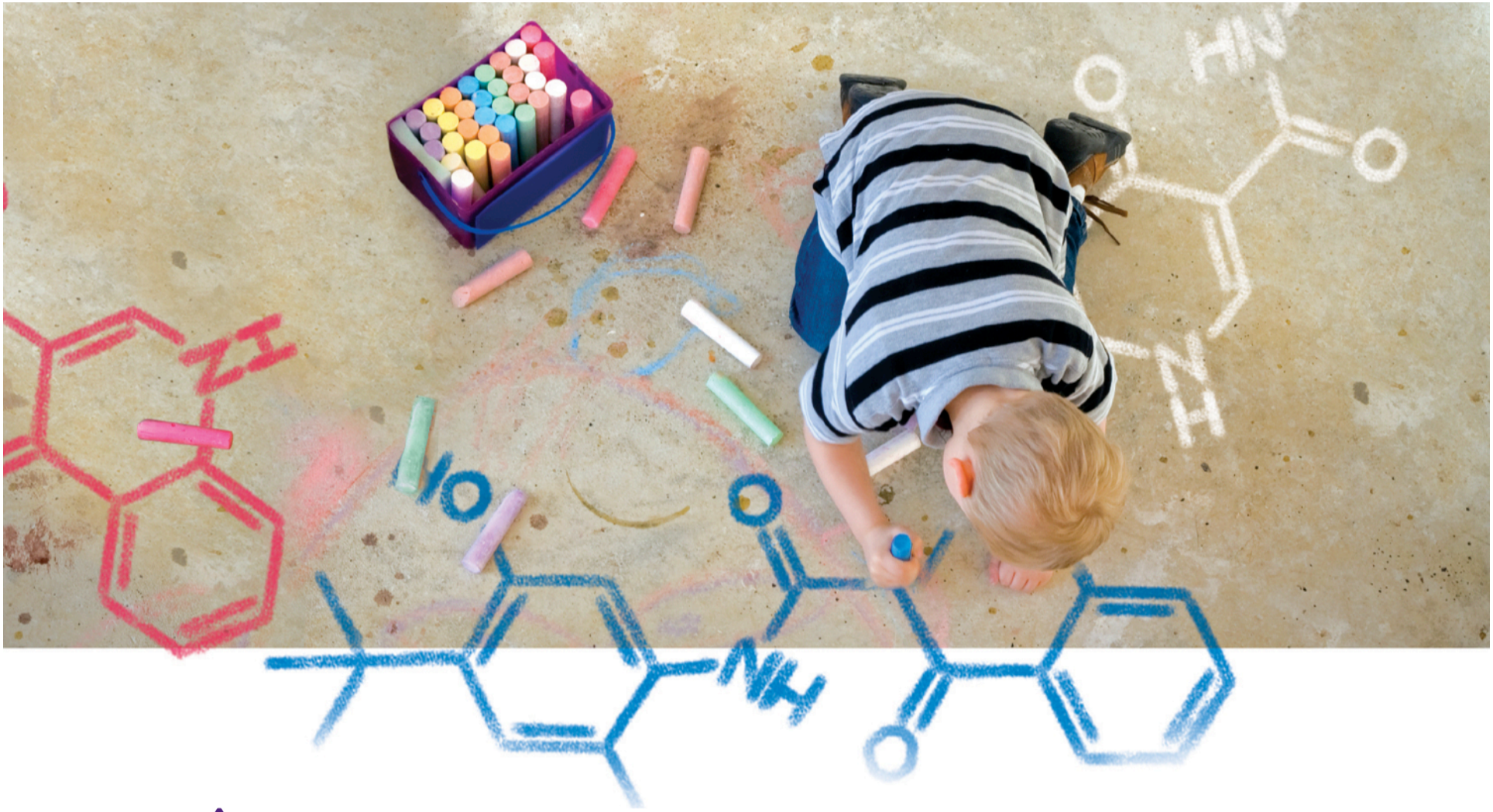


VERTEX™

THE SCIENCE *of* POSSIBILITY

Just Because You Published It Doesn't Mean It's Right

Examining a Common CADD Validation Error

Pat Walters | CADD GRC 2013 | July 22, 2013

Case Study #1

How to Improve Docking Accuracy of AutoDock4.2: A Case Study Using Different Electrostatic Potentials

Xuben Hou,[†] Jintong Du,[†] Jian Zhang,[‡] Lupei Du,[†] Hao Fang,[*,†] and Minyong Li[*,†]

[†]Department of Medicinal Chemistry, Key Laboratory of Chemical Biology (MOE), School of Pharmacy, Shandong University, Jinan, Shandong 250012, China

[‡]Department of Pathophysiology, Key Laboratory of Cell Differentiation and Apoptosis (MOE), Shanghai Jiao-Tong University School of Medicine (SJTU-SM), Shanghai, 200025, China

J. Chem. Inf. Model., 2013, 53(1):188-200. doi: 10.1021/ci300417y

VERTEX

Case Study #1

How to Improve Docking Accuracy of AutoDock4.2: A Case Study Using Different Electrostatic Potentials

Xuben Hou,[†] Jintong Du,[†] Jian Zhang,[‡] Lupei Du,[†] Hao Fang,*[,†] and Minyong Li*[,†]

[†]Department of Medicinal Chemistry, Key Laboratory of Chemical Biology (MOE), School of Pharmacy, Shandong University, Jinan, Shandong 250012, China

[‡]Department of Pathophysiology, Key Laboratory of Cell Differentiation and Apoptosis (MOE), Shanghai Jiao-Tong University School of Medicine (SJTU-SM), Shanghai, 200025, China

Compare virtual screening performance of 9 charge models

- AM1-BCC
- Del-Re
- Formal
- Gasteiger−Hückel
- Gasteiger−Marsili
- Hückel
- MMFF
- Pullman
- Hartree−Fock

VERTEX

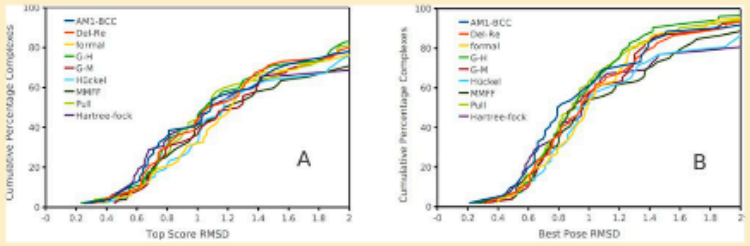
The Study

- 52 complexes from the PDB
- Assign ligand charges using each of 9 different methods
- Dock with AutoDock 4.2
- Compare RMS fit to x-ray structure

VERTEX

From the Abstract

ABSTRACT: Molecular docking, which is the indispensable emphasis in predicting binding conformations and energies of ligands to receptors, constructs the high-throughput virtual screening available. So far, increasingly numerous molecular docking programs have been released, and among them, AutoDock 4.2 is a widely used docking program with exceptional accuracy. It has heretofore been substantiated that the calculation of partial charge is very fundamental for the accurate conformation search and binding energy estimation. However, no systematic comparison of the significances of electrostatic potentials on docking accuracy of AutoDock 4.2 has been determined. In this paper, nine different charge-assigning methods, including AM1-BCC, Del-Re, formal, Gasteiger−Hückel, Gasteiger−Marsili, Hückel, Merck molecular force field (MMFF), and Pullman, as well as the ab initio Hartree−Fock charge, were sufficiently explored for their molecular docking performance by using AutoDock4.2. The results clearly demonstrated that the empirical Gasteiger−Hückel charge is the most applicable in virtual screening for large database; meanwhile, the semiempirical AM1-BCC charge is practicable in lead compound optimization as well as accurate virtual screening for small databases.



The results clearly demonstrated that the empirical Gasteiger−Hückel charge is the most applicable in virtual screening for large database; meanwhile, the semiempirical AM1-BCC charge is practicable in lead compound optimization as well as accurate virtual screening for small databases.

Let's grab the data and take a look

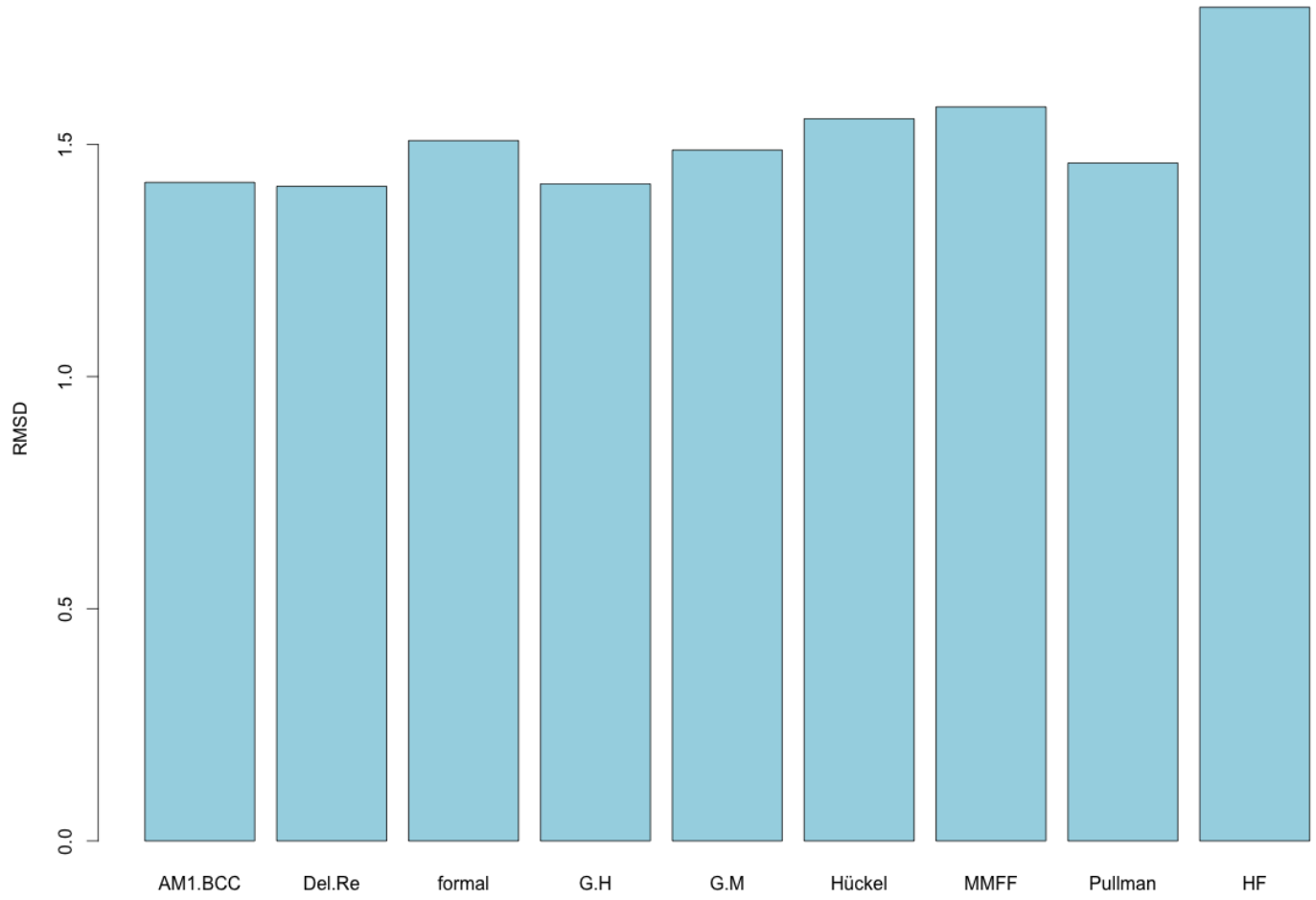
Table S2. The "Top Score RMSD" in molecular docking results of AutoDock4.2 using nine different charge-assigning methods.

PDB code	Top Score RMSD (Å)								
	AM1-BCC	Del-Re	formal	G-H	G-M	Hückel	MMFF	Pullman	HF
1S39	0.25	0.39	1.39	0.23	1.39	0.38	0.26	0.32	0.29
1ULB	0.67	0.63	0.72	0.72	0.67	0.72	0.67	0.73	0.65
2CPP	2.10	2.12	2.17	1.00	2.12	2.16	2.11	1.92	2.10
1DWB	0.68	1.92	1.93	1.92	1.93	1.93	1.92	1.94	0.52
3PTB	0.42	0.32	0.39	0.37	0.45	0.47	0.54	0.31	3.60
1AI5	2.57	2.48	2.59	2.47	2.49	2.57	2.58	2.61	1.01
1FLR	1.09	1.15	1.14	1.14	1.14	1.15	1.09	1.09	0.78
2ACK	4.70	4.70	4.69	4.71	4.69	4.69	4.70	4.69	4.70
2YPI	1.03	1.22	1.25	1.29	1.13	1.02	0.96	1.12	1.18
1AJP	3.68	3.59	3.56	3.70	3.61	3.65	4.10	3.62	2.56
1BGQ	0.64	1.04	1.06	1.05	1.03	1.06	0.94	1.05	0.67
1K4G	1.39	1.30	1.30	1.36	1.30	1.38	1.38	1.37	1.36
2IWX	0.98	3.14	0.93	3.14	3.14	0.95	3.14	3.15	0.60
7ABP	0.64	0.69	0.56	0.66	0.59	0.56	0.66	0.57	0.65
1CBR	1.03	1.00	1.09	0.96	1.01	0.93	1.01	1.04	1.01
1KV1	0.81	0.98	1.23	0.88	1.00	1.01	0.89	0.94	0.88
1M0Q	0.46	0.64	0.41	0.55	0.67	0.81	0.72	0.78	5.14
1Q8T	1.82	1.88	1.83	1.44	1.41	1.85	1.83	1.85	1.42

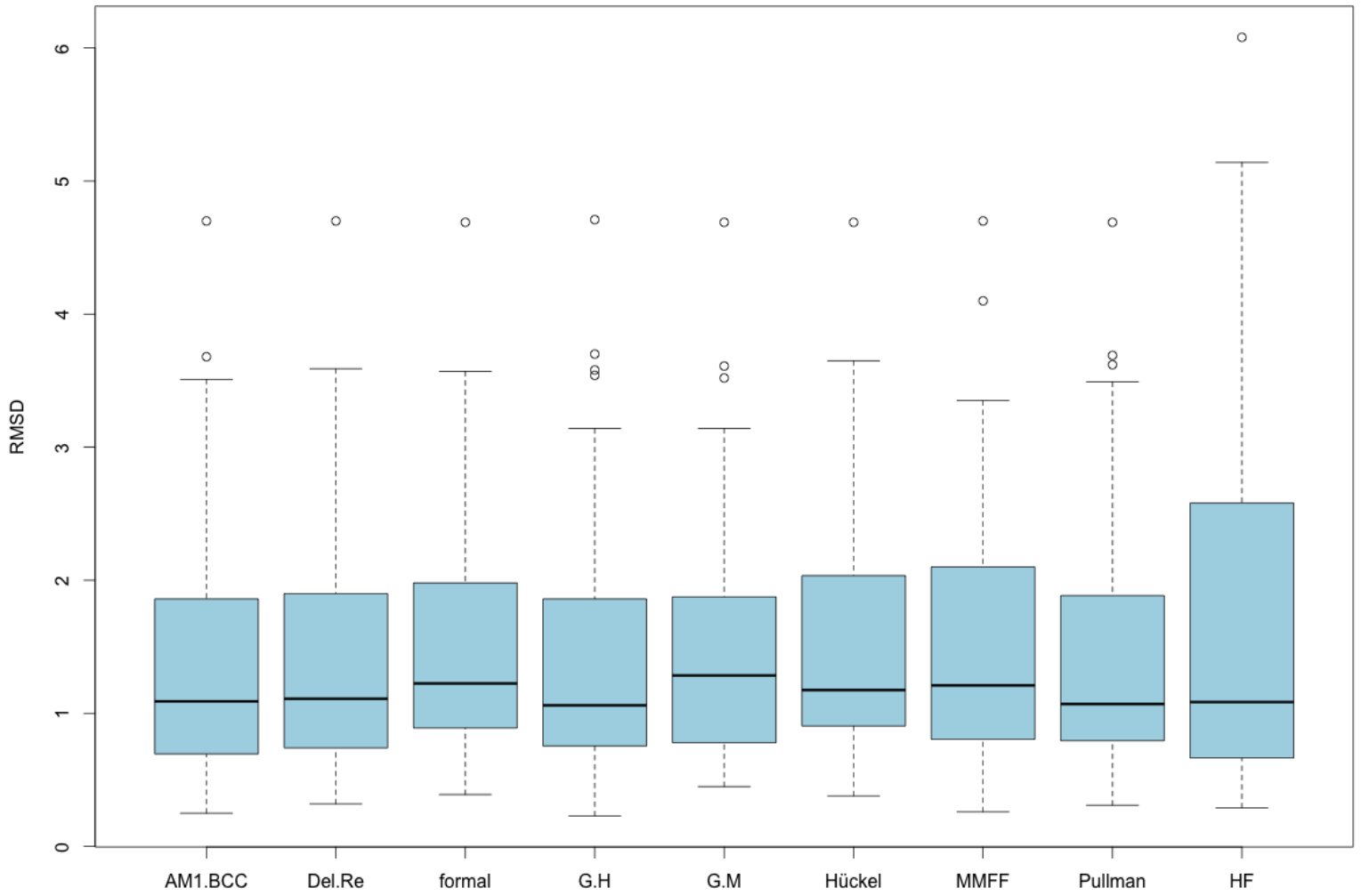
VERTEX

Mean RMSD for 52 Targets

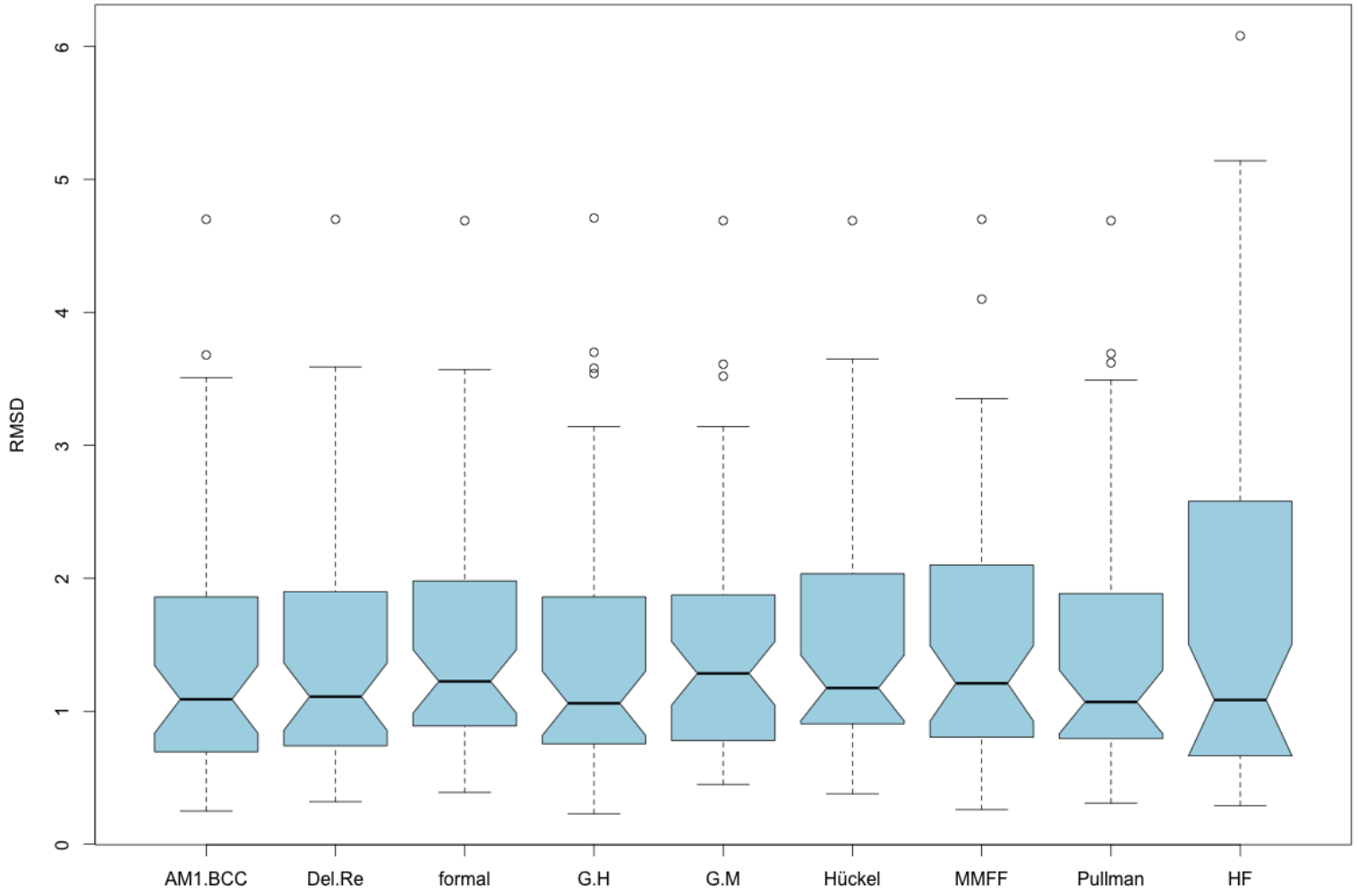
Are these really different?



Boxplots Can Depict the Distribution

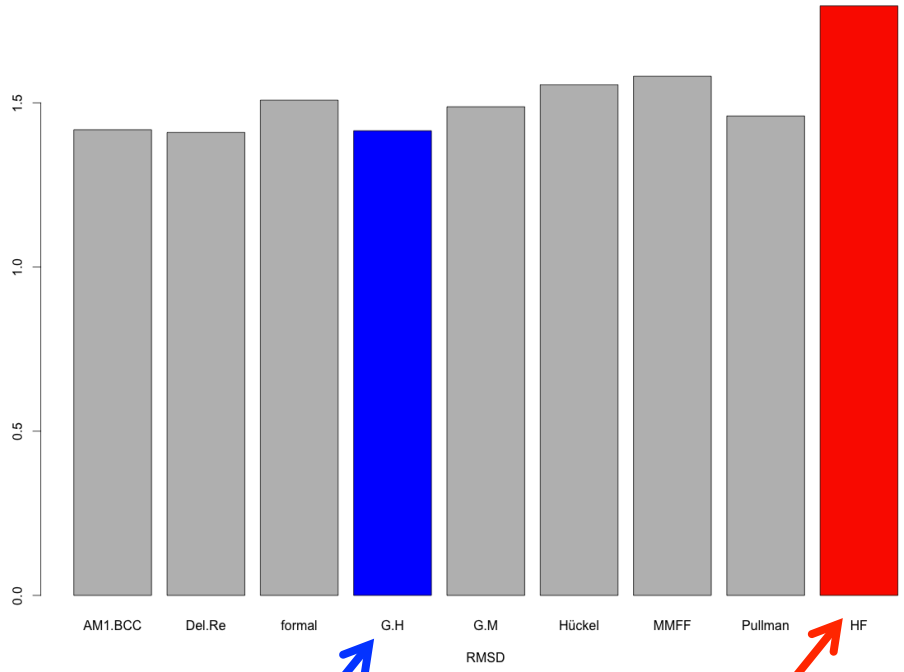


Boxplot Notches Show Standard Deviation



VERTEX

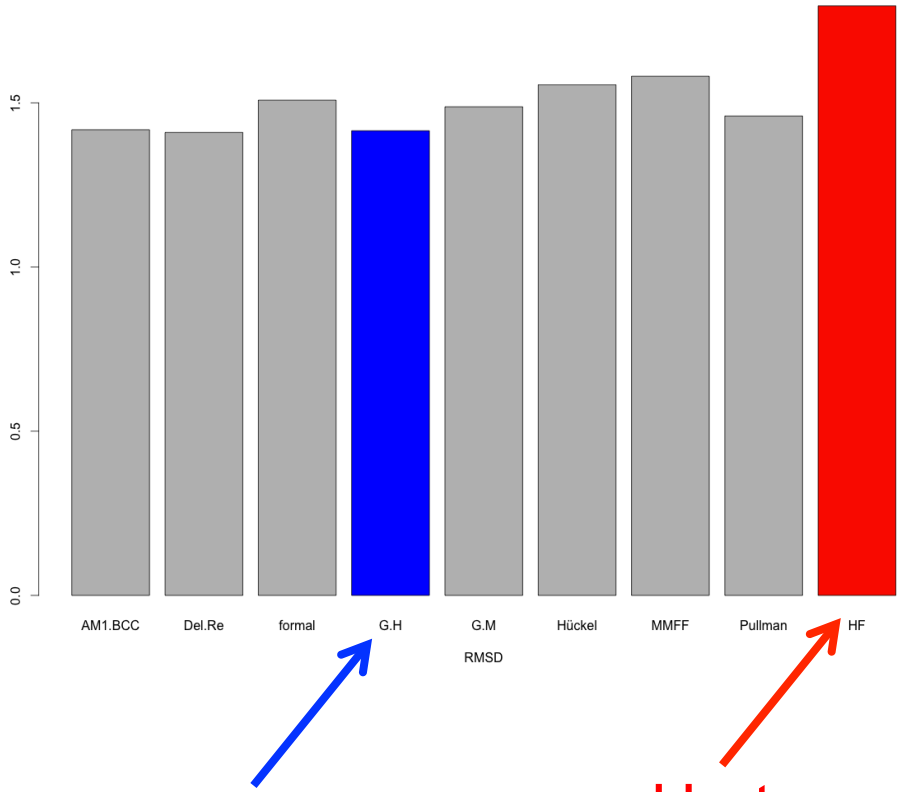
Let's Put Some Numbers to This

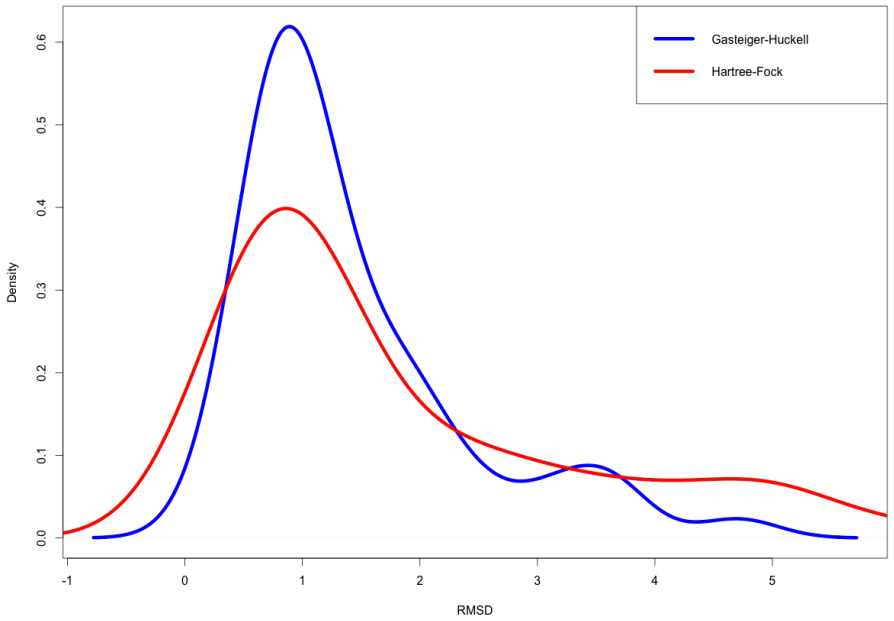


Gasteiger−Hückel

Hartree−Fock

Let's Put Some Numbers to This





Gasteiger−Hückel

Hartree−Fock

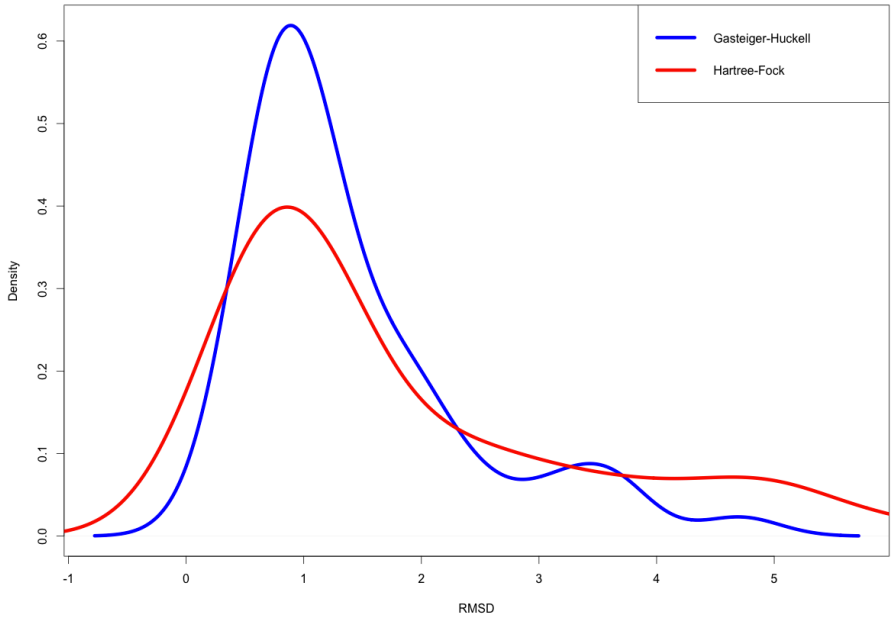
VERTEX

Lets Just Do a T-test

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\left(s_1^2\right)/n + \left(s_2^2\right)/n}}$$

t = -1.52

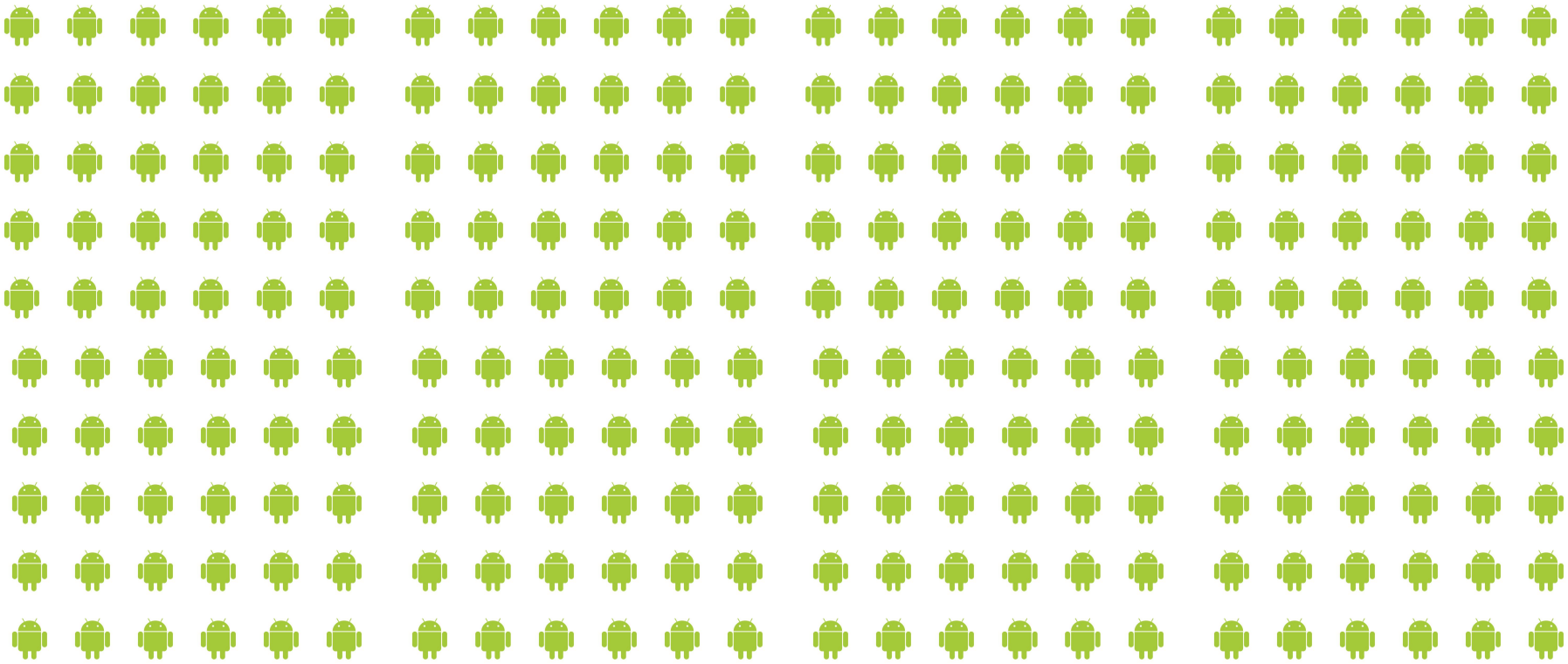
Student's t-test
p-value = 0.13



However

- The t-test is for *independent* samples
- What do we mean by this?

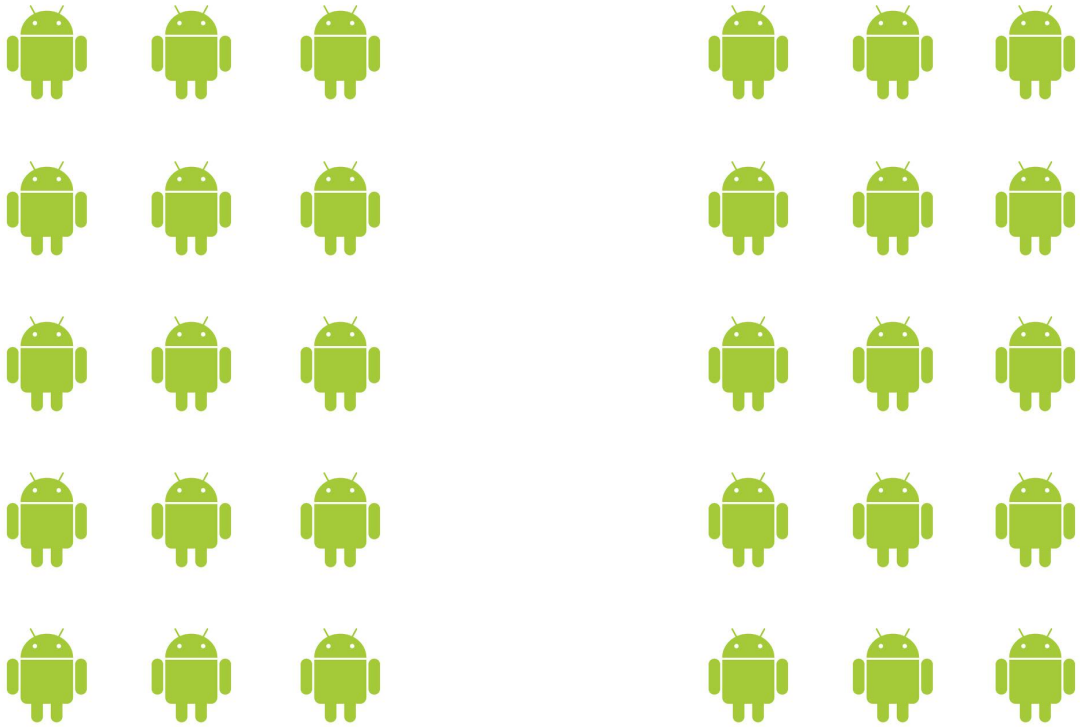
Let's Start with a Population of 240 Martians



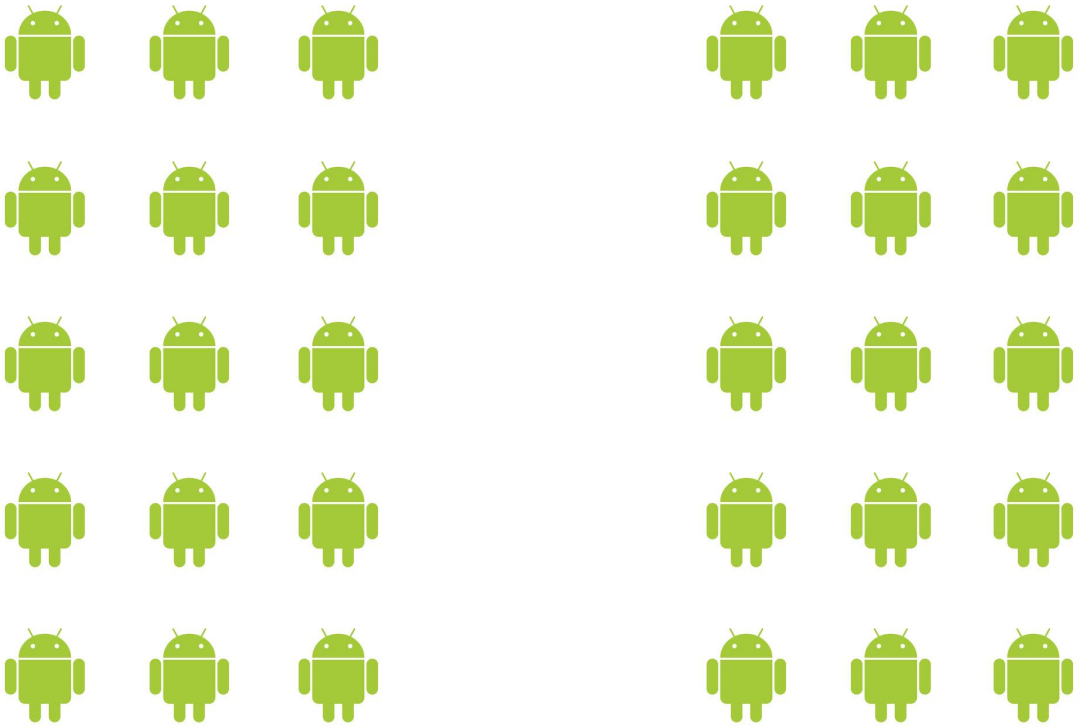
With apologies to Stanton A Glantz – Primer of Biostatistics

VERTEX

Randomly Select 2 Groups of 15



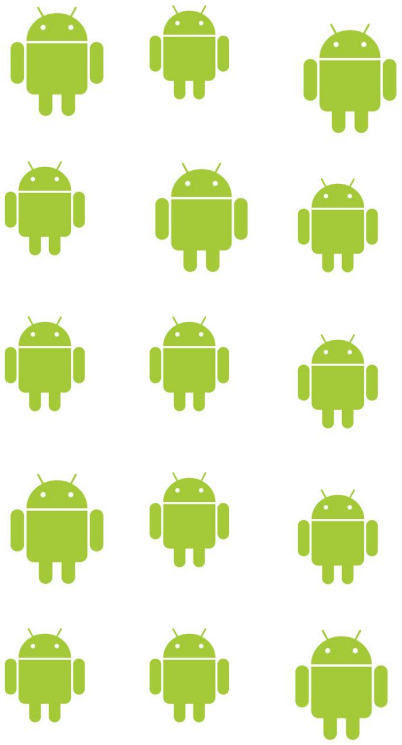
Conduct the Experiment



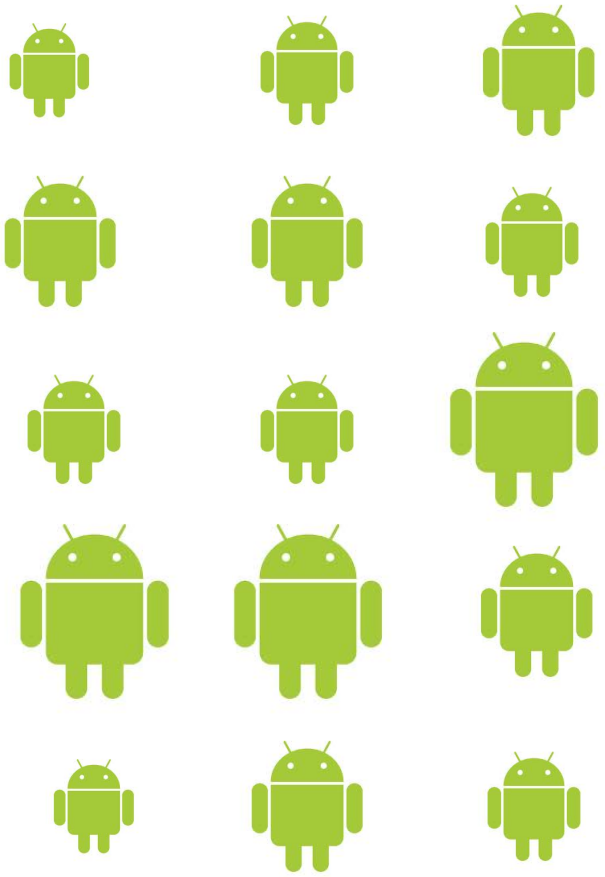
Placebo

Testosterone

Examine Results

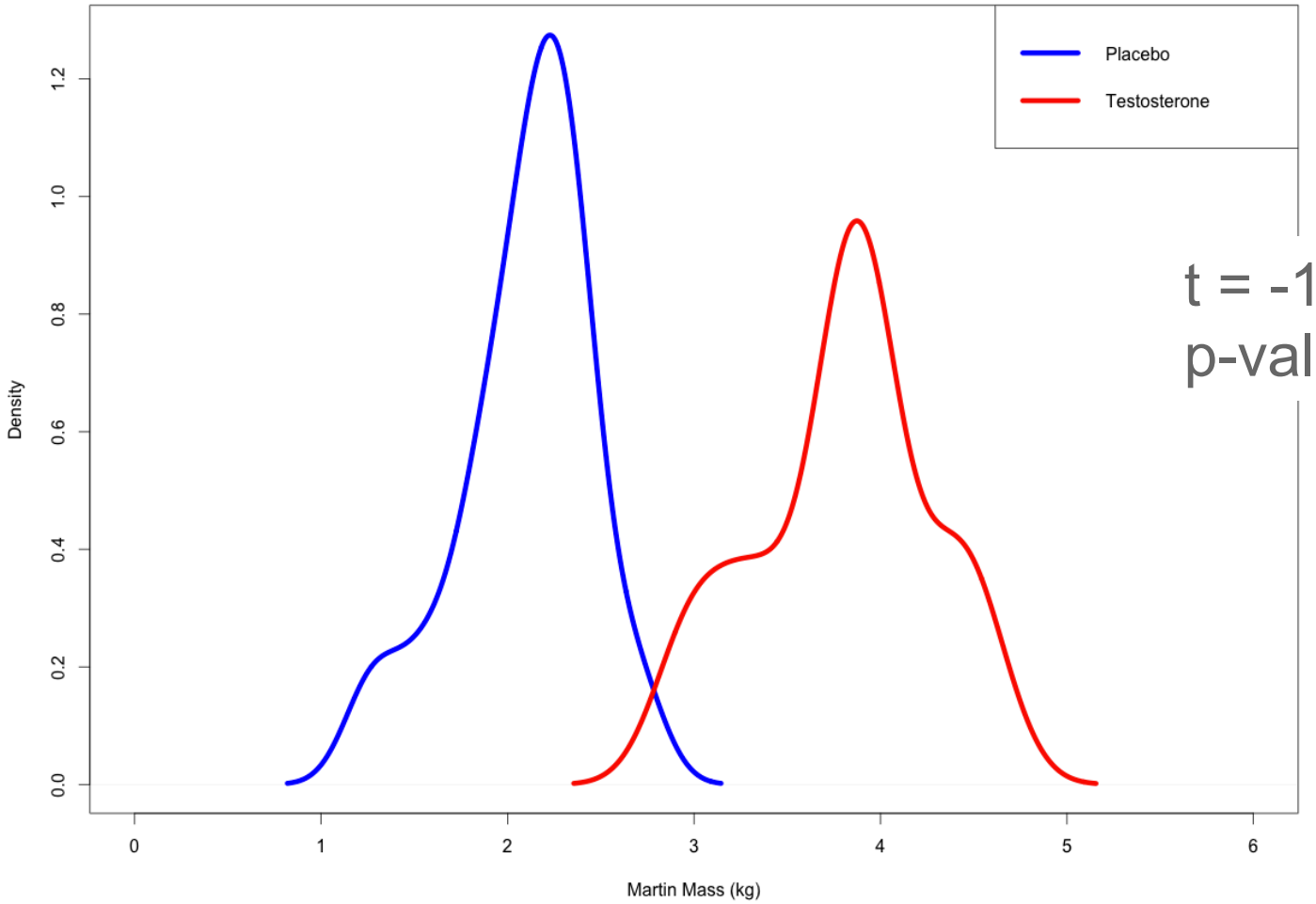


Placebo



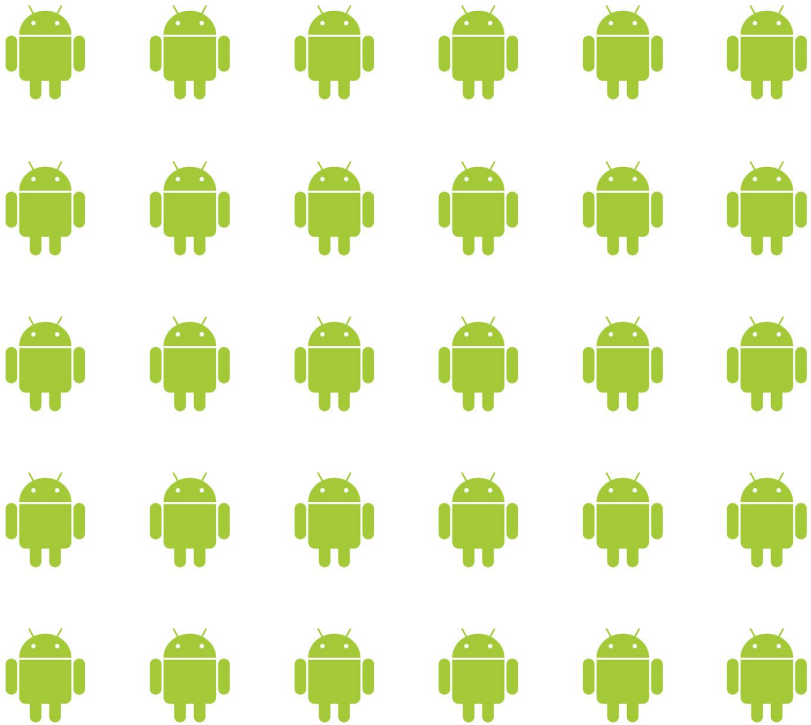
Testosterone

Carry Out the Analysis

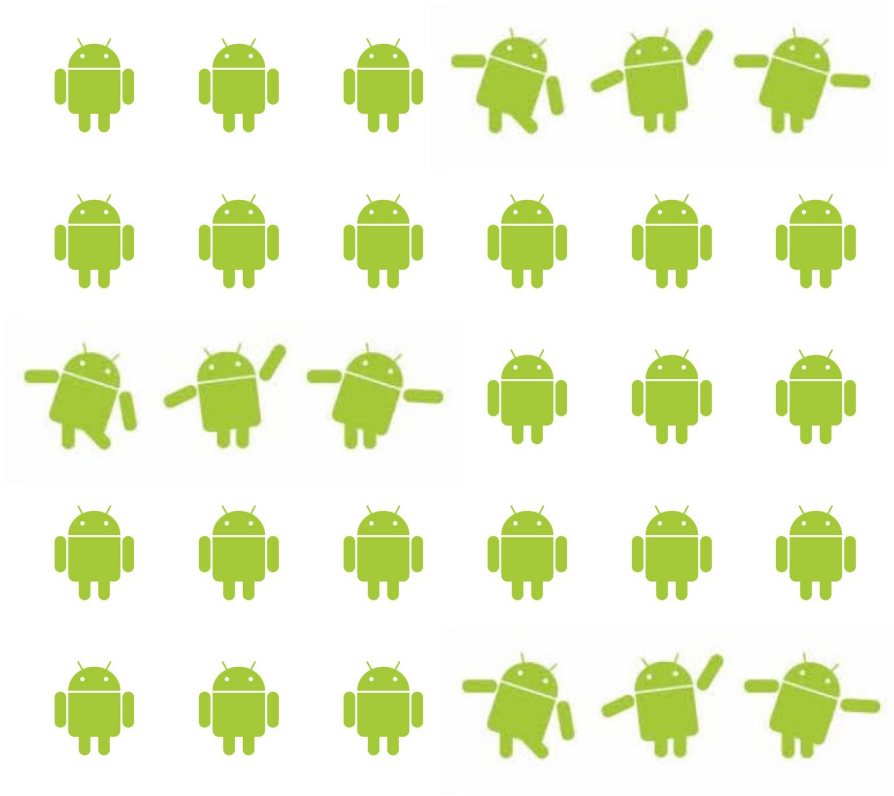


t = -10.9
p-value = 3.72e-10

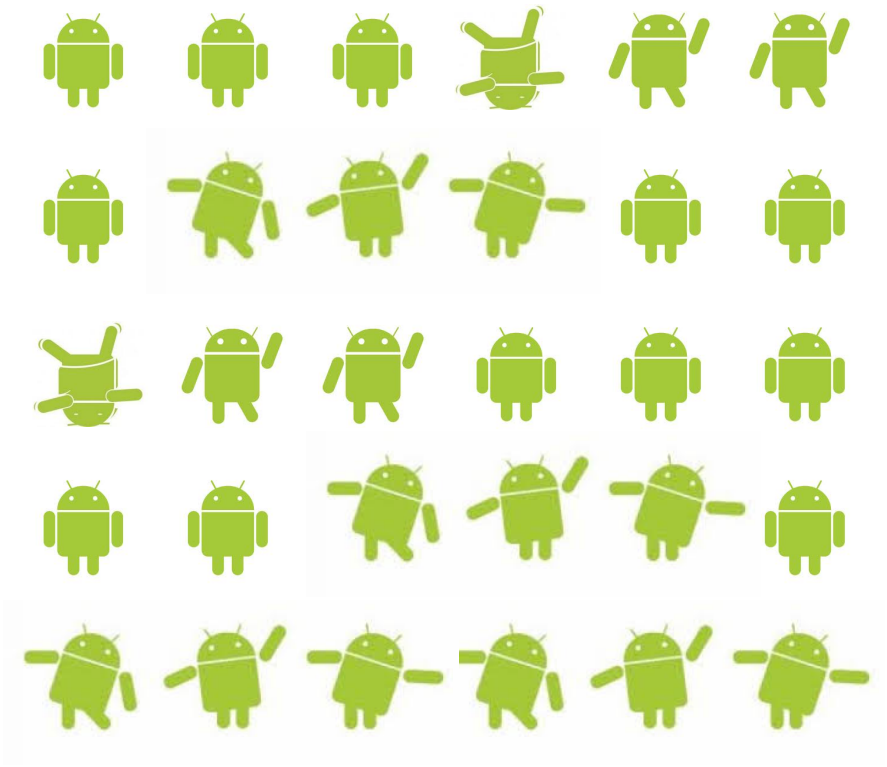
But Our Data is not Independent



Multiple Tests on the Same Population



Multiple Tests on the Same Population



So what do we do?

- Comparing two methods: Paired t-test
- Comparing multiple methods: Friedman's test
  - Also works with rankings

Student's Paired t-test

Name		
hgudyd		
dhsishd		
dfkdjdhd		
dkdshs		
vkfhsdksd		

Student's Paired t-test

$$t = \frac{d - D}{SE}$$

- d – mean of difference
- D – hypothesized mean difference (D = 0 for null)
- SE – standard error of the mean differences

p-value is taken from the t-distribution

Paired t-test on the Charge/Docking Dataset

	AM1.BCC	Del.Re	formal	G.H	G.M	HF	Hückel	MMFF
Del.Re	0.941	-	-	-	-	-	-	-
formal	0.271	0.254	-	-	-	-	-	-
G.H	0.976	0.927	0.279	-	-	-	-	-
G.M	0.526	0.106	0.813	0.181	-	-	-	-
HF	0.105	0.060	0.184	0.072	0.149	-	-	-
Hückel	0.130	0.091	0.396	0.142	0.440	0.253	-	-
MMFF	0.022	0.085	0.428	0.060	0.357	0.333	0.803	-
Pullman	0.710	0.333	0.544	0.476	0.655	0.121	0.229	0.211

The Importance of Multiple Comparisons

- *P* (at least one significant result) = 1 – *P* (no significant results)
- *P* (at least one significant result) = $1 – (1-0.05)^{36}$
- *P* (at least one significant result) = 0.84

84% chance of identifying at least one chance correlation

http://www.aaos.org/news/aaosnow/apr12/research7.asp

VERTEX

Correcting for Multiple Testing

	AM1.BCC	Del.Re	formal	G.H	G.M	HF	Hückel	MMFF
Del.Re	1.00	-	-	-	-	-	-	-
formal	1.00	1.00	-	-	-	-	-	-
G.H	1.00	1.00	1.00	-	-	-	-	-
G.M	1.00	1.00	1.00	1.00	-	-	-	-
HF	1.00	1.00	1.00	1.00	1.00	-	-	-
Hückel	1.00	1.00	1.00	1.00	1.00	1.00	-	-
MMFF	0.78	1.00	1.00	1.00	1.00	1.00	1.00	-
Pullman	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

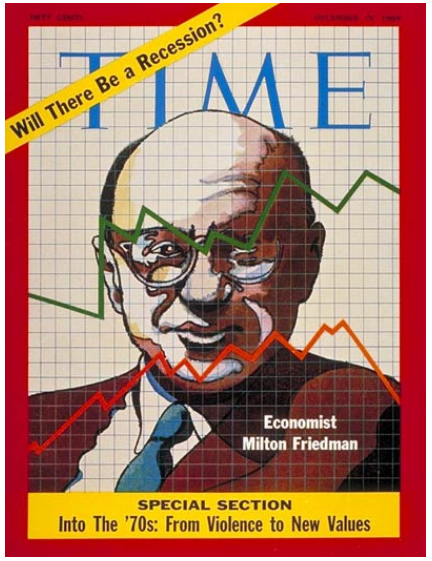
For more information look up:

Bonferroni correction

Holm's method

Friedman's Test

	AM1.BCC	Del.Re	formal
1S39	0.25	0.39	1.39
1ULB	0.67	0.63	0.72
2CPP	2.10	2.12	2.17
1DWB	0.68	1.92	1.93
3PTB	0.42	0.32	0.39



Convert to Ranks

	AM1.BCC	Del.Re	formal
1S39	1	2	3
1ULB	2	1	3
2CPP	1	2	3
1DWB	1	2	3
3PTB	3	1	2

Calculate Rank Sums

	AM1.BCC	Del.Re	formal
1S39	1	2	3
1ULB	2	1	3
2CPP	1	2	3
1DWB	1	2	3
3PTB	3	1	2
	8	8	14

Calculate the Expected Average Rank

	AM1.BCC	Del.Re	formal
1S39	1	2	3
1ULB	2	1	3
2CPP	1	2	3
1DWB	1	2	3
3PTB	3	1	2
	8	8	14

Average rank sum is $n(k+1)/2 = 5(3+1)/2=10$

n = number of complexes
k = number of methods

Calculate S - the Sum of Squared Deviates

	AM1.BCC	Del.Re	formal
1S39	1	2	3
1ULB	2	1	3
2CPP	1	2	3
1DWB	1	2	3
3PTB	3	1	2
	8	8	14

$$S = \sum \left[R_t - n(k+1)/2 \right]^2$$

$$S = (8-10)^2 + (8-10)^2 + (14-10)^2$$

$$S = 24$$

Calculate χ^2 and associate a P-value

$$\chi_r^2 = \frac{S}{nk(k+1)/12}$$

Friedman chi-squared = 4.8
p-value = 0.09

We cannot reject the null hypothesis that the methods are the same.

VERTEX

Back to the Original Dataset

Method	Rank Sums
AM1.BCC	165
G.H	178
Del.Re	187
HF	190
G.M	215
Pullman	219
formal	229
MMFF	229
Hückel	260

Friedman chi-squared = 15.2
p-value = 0.06

We cannot reject the null hypothesis that the methods are the same.

VERTEX

Recommendations

- Visualize comparisons with notched boxplots
- Use appropriate statistical tests
- Comparing 2 dependent methods
  - Paired t-test
- Beware of multiple hypothesis testing
  - Friedman's test

VERTEX

Thanks

- Anna Legedza
- Brian Goldman
- R-bloggers
- Ant

Properly validate you must

VERTEX

Code and Data are on GitHub

- All of the code and data used in this presentation are available

https://github.com/PatWalters/cadd_grc_2013

- The mechanics of the data analysis will be covered in an informal session this afternoon.

VERTEX