Supplemental Information for

CSAR Benchmark Exercise of 2010: Selection of the protein-ligand complexes

James B. Dunbar Jr., Richard D. Smith, Chao-Yie Yang, Peter Man-Un Ung, Katrina W. Lexa, Nickolay A. Khazanov, Jeanne A. Stuckey, Shaomeng Wang, and Heather A. Carlson

The size of any data set dictates the confidence intervals.

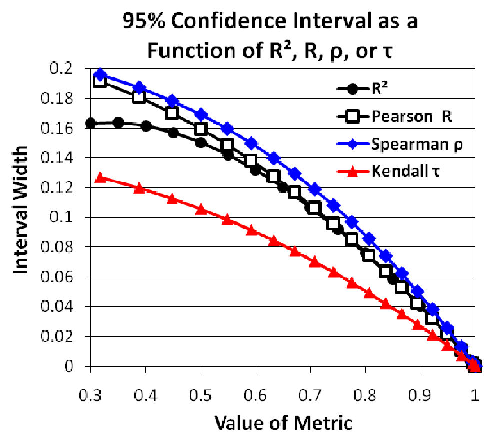
The following relationship from eqn 3 in Bonett and Wright (2000, Pyschometrika, 65: 23-28) relates Fisher confidence intervals to the sample size n:

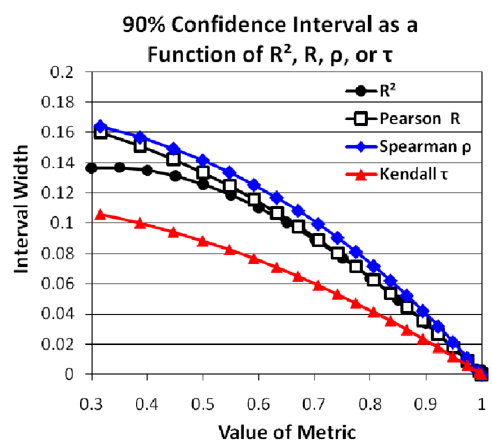
$$n = 4c^2(1-\theta^2)^2\left(\frac{z_{\alpha/2}}{w}\right)^2 + b$$

where w is the width of the interval and $z_{\alpha/2}$ dictates the confidence: 1.64 for 90%, 1.96 for 95%, and 2.58 for 99%. The values of c, b, and θ are related through the variance (σ^2) and are dictated by the fitness parameter being evaluated:

Metric	θ	c^2	b	σ^2
Pearson	R	1	3	$(1-R^2)^2/(n-3)$
Spearman	ρ	$1+\rho^2/2$	3	$(1+\rho^2/2)(1-\rho^2)^2/(n-3)$
Kendall	τ	0.437	4	$0.437\times(1-\tau^2)^2/(n-4)$

For the full dataset of 343 complexes, the metrics and their confidence intervals are:





R^2	**Pearson R**	$w_{95\%}$ **for R**	**95% interval** $(R \pm w_{95\%}/2)$	$w_{90\%}$ **for R**	**90% interval** $(R \pm w_{90\%}/2)$	**95% interval** $(R_{up}^2-R_{down}^2)$	$w_{95\%}$ **for** R^2	**90% interval** $(R_{up}^2-R_{down}^2)$	$w_{90\%}$ **for** R^2
1	1	0	(1.0-1.0)	0	1.0-1.0	(1.0-1.0)	0.0	1.0-1.0	0.0
0.99	0.995	0.002	(0.996-0.994)	0.002	0.996-0.994	(0.992-0.988)	0.004	0.992-0.988	0.004
0.95	0.975	0.011	(0.980-0.969)	0.009	0.979-0.970	(0.960-0.940)	0.021	0.959-0.941	0.017
0.9	0.949	0.021	(0.959-0.938)	0.018	0.958-0.940	(0.920-0.880)	0.040	0.917-0.883	0.034
0.8	0.894	0.043	(0.916-0.873)	0.036	0.912-0.877	(0.838-0.762)	0.076	0.832-0.768	0.064
0.7	0.837	0.064	(0.869-0.805)	0.053	0.863-0.810	(0.754-0.648)	0.107	0.745-0.656	0.089
0.6	0.775	0.085	(0.817-0.732)	0.071	0.810-0.739	(0.668-0.536)	0.132	0.656-0.546	0.110
0.5	0.707	0.106	(0.760-0.654)	0.089	0.752-0.663	(0.578-0.428)	0.150	0.565-0.439	0.126
0.4	0.632	0.128	(0.696-0.569)	0.107	0.686-0.579	(0.485-0.323)	0.161	0.470-0.335	0.135
0.3	0.548	0.149	(0.622-0.473)	0.125	0.610-0.485	(0.387-0.224)	0.163	0.372-0.236	0.136
0.2	0.447	0.170	(0.532-0.362)	0.142	0.518-0.376	(0.283-0.131)	0.152	0.269-0.141	0.127
0.1	0.316	0.191	(0.412-0.221)	0.16	0.396-0.236	(0.170-0.049)	0.121	0.157-0.056	0.101

Small overlaps in the confidence intervals can still lead to statistically significant differences in these metrics if one uses a Levene F-test to evaluate the residuals from the linear regression.

Spearman ρ

ρ^2	**$w_{95\%}$ ρ**	**95% interval for ρ**	**(ρ ± $w_{95\%}$/2)**	**$w_{90\%}$ for ρ**	**90% interval (ρ ± $w_{90\%}$/2)**
1	1	0	1.0-1.0	0	1.0-1.0
0.99	0.995	0.003	0.996-0.994	0.002	0.996-0.994
0.95	0.975	0.013	0.981-0.968	0.011	0.980-0.969
0.9	0.949	0.026	0.961-0.936	0.021	0.959-0.938
0.8	0.894	0.050	0.920-0.869	0.042	0.915-0.873
0.7	0.837	0.074	0.874-0.800	0.062	0.868-0.806
0.6	0.775	0.097	0.823-0.726	0.081	0.815-0.734
0.5	0.707	0.119	0.767-0.648	0.099	0.757-0.657
0.4	0.632	0.140	0.702-0.563	0.117	0.691-0.574
0.3	0.548	0.160	0.628-0.468	0.134	0.614-0.481
0.2	0.447	0.178	0.536-0.358	0.149	0.522-0.373
0.1	0.316	0.196	0.414-0.218	0.164	0.398-0.234

Kendall τ

τ^2	**$w_{95\%}$ τ**	**95% interval for τ**	**(τ ± $w_{95\%}$/2)**	**$w_{90\%}$ for τ**	**90% interval (τ ± $w_{90\%}$/2)**
1	1	0	1.0-1.0	0	1.0-1.0
0.99	0.995	0.001	0.996-0.994	0.001	0.996-0.994
0.95	0.975	0.007	0.978-0.971	0.006	0.978-0.972
0.9	0.949	0.014	0.956-0.942	0.012	0.955-0.943
0.8	0.894	0.028	0.909-0.880	0.024	0.906-0.883
0.7	0.837	0.042	0.858-0.816	0.035	0.854-0.819
0.6	0.775	0.056	0.803-0.746	0.047	0.798-0.751
0.5	0.707	0.070	0.742-0.672	0.059	0.737-0.678
0.4	0.632	0.084	0.675-0.590	0.071	0.668-0.597
0.3	0.548	0.099	0.597-0.498	0.082	0.589-0.507
0.2	0.447	0.113	0.504-0.391	0.094	0.494-0.400
0.1	0.316	0.127	0.380-0.253	0.106	0.369-0.263