Activity Characterization of Triazines Analogues: Statistical Parameters for Models Assessment

Sorana Daniela BOLBOACĂ & Lorentz JÄNTSCHI

"Iuliu Hatieganu" & Technical Universities, Cluj-Napoca, Romania

International Symposium on Organic Chemistry
December 09-12 , 2006 Sofia, Bulgaria

Purpose

- To analyzed three previously reported **Molecular Descriptors Family (MDF)** on **Structure-Activity Relationships (SAR)** models by the use of correlation coefficients:
  - Pearson (r_{Prs})
  - Spearman (ρ_{Spm})
  - Semi-quantitative (r_{sQ})
  - Kendall's tau ($\tau_{Ken,a}$, $\tau_{Ken,b}$, $\tau_{Ken,c}$)
  - Gamma (Γ)

} correlation coefficients

- 1,3,5-substituted-triazines: thirty compounds
- inhibition activity on Chorella: pI_{50} = concentration required for 50% inhibition of Hill reaction

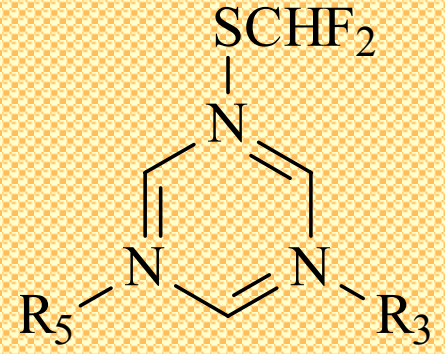
[*Morita at all., 1987*]

Previously reported QSAR

$$\text{Est } pI_{50} = 9.614 - 0.153 \cdot X_5 - 58.888 \cdot 1/V_5 - 2.430 \cdot 1/N_3$$

[*Diudea at all., 2002*]

- n = 30
- r^2 = 0.9694
- F = 274.3
- r_{loo} = 0.9778



where

X_5 = topological descriptor for substituent number 5

V_5 = fragmental volumes of the substituent in the position 5 (cm^3/mol)

N_3 = total number of hydrogen's in the substituent 3

MDF SAR Methodology

- Preparing chemical compounds for molecular modeling
- Generating the molecular descriptors family
- Finding
- Validating
- Comparing

} the MDF SAR models

Methods

- Correlation coefficients, Squared correlation coefficient
- Statistical test and associated significance

  - **Pearson**
  - **Spearman**
  - **Semi-quantitative**

  } correlation coefficients

$$r_{Prs} = \frac{\sum (Y_{m-i} - \overline{Y}_m)(Y_{est-i} - \overline{Y}_{est})}{\sqrt{\left(\sum (Y_{m-i} - \overline{Y}_m)^2\right)\left(\sum (Y_{est-i} - \overline{Y}_{est})^2\right)}} \qquad r_{Spm} = \frac{\sum (R_{Y_{m-i}} - \overline{R}_{Y_m})(R_{Y_{est-i}} - \overline{R}_{Y_{est}})}{\sqrt{\left(\sum (R_{Y_{m-i}} - \overline{R}_{Y_m})^2\right)\left(\sum (R_{Y_{est-i}} - \overline{R}_{Y_{est}})^2\right)}}$$

$$r_{sQ} = \sqrt{\frac{\sum (Y_{m-i} - \overline{Y}_m)(Y_{est-i} - \overline{Y}_{est})}{\sqrt{\left(\sum (Y_{m-i} - \overline{Y}_m)^2\right)\left(\sum (Y_{est-i} - \overline{Y}_{est})^2\right)}} \cdot \frac{\sum (R_{Y_{m-i}} - \overline{R}_{Y_m})(R_{Y_{est-i}} - \overline{R}_{Y_{est}})}{\sqrt{\left(\sum (R_{Y_{m-i}} - \overline{R}_{Y_m})^2\right)\left(\sum (R_{Y_{est-i}} - \overline{R}_{Y_{est}})^2\right)}}}$$

Methods

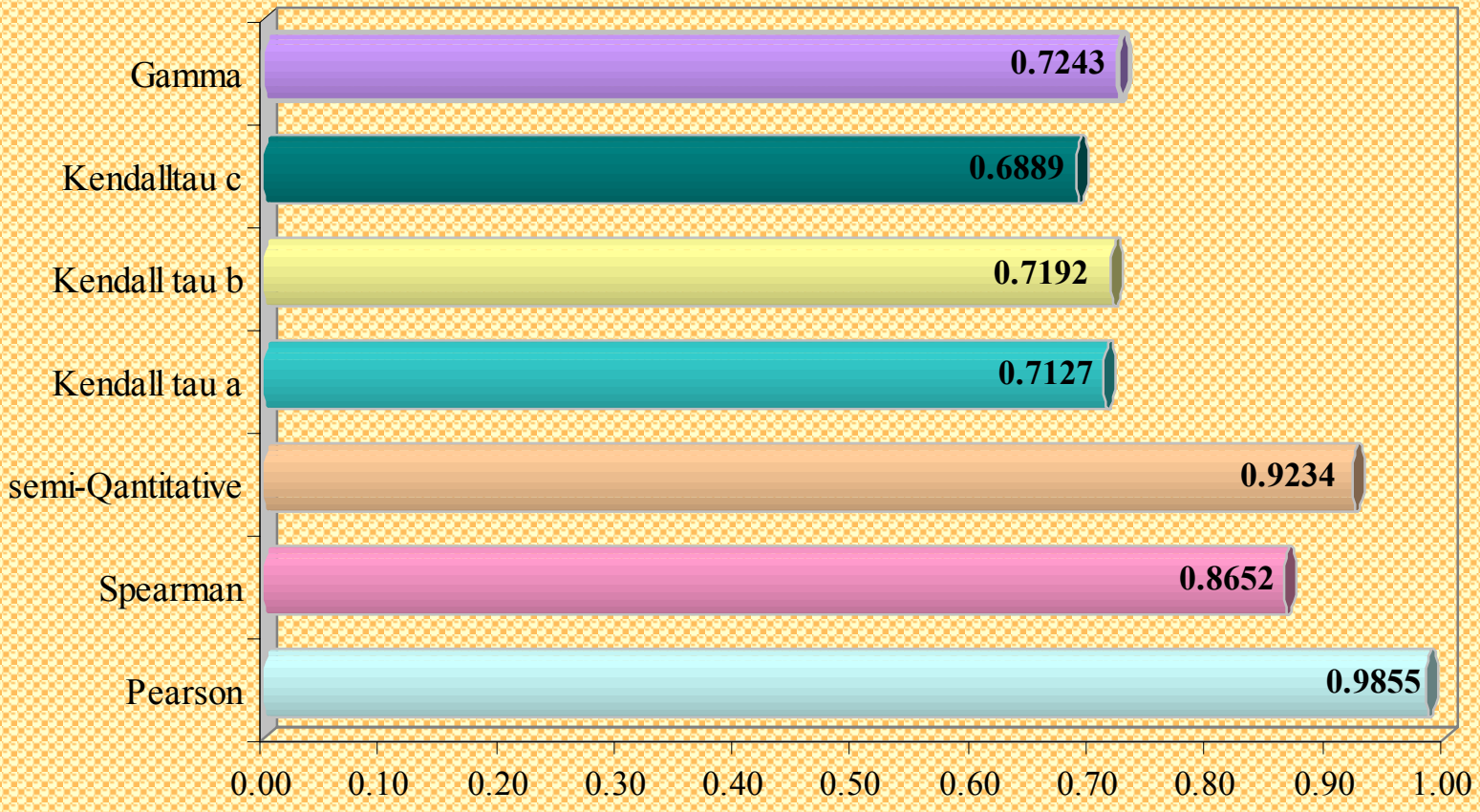
- **More correlation coefficients (!)**:

  n(n-1)/2 pairs, *t* = number of tied Y_m values and *u* = number of tied Y_{est} values, C = concordances, D = discordances

  - **Kendall**:
    - tau-a ($\tau_{Ken,a}$): $\tau_{Ken,a} = (C-D)/[n(n-1)/2]$
    - tau-b ($\tau_{Ken,b}$): $\tau_{Ken,b} = (C-D)/\sqrt{[(n(n-1)/2-t)(n(n-1)/2-u)]}$
    - tau-c ($\tau_{Ken,c}$): $\tau_{Ken,c} = 2(C-D)/n^2$
  - **Gamma**:
    - $\Gamma = (C-D)/(C+D)$

Results for MDF SARs

for 5.52 - 8112.2·*iSMMWHg* + 194.35·*iSMmEQt (2v structure vs. pI_{50}, from* [*Bolboacă & Jäntschi, LEJPT (DOAJ), 2006*])



Results for MDF SARs

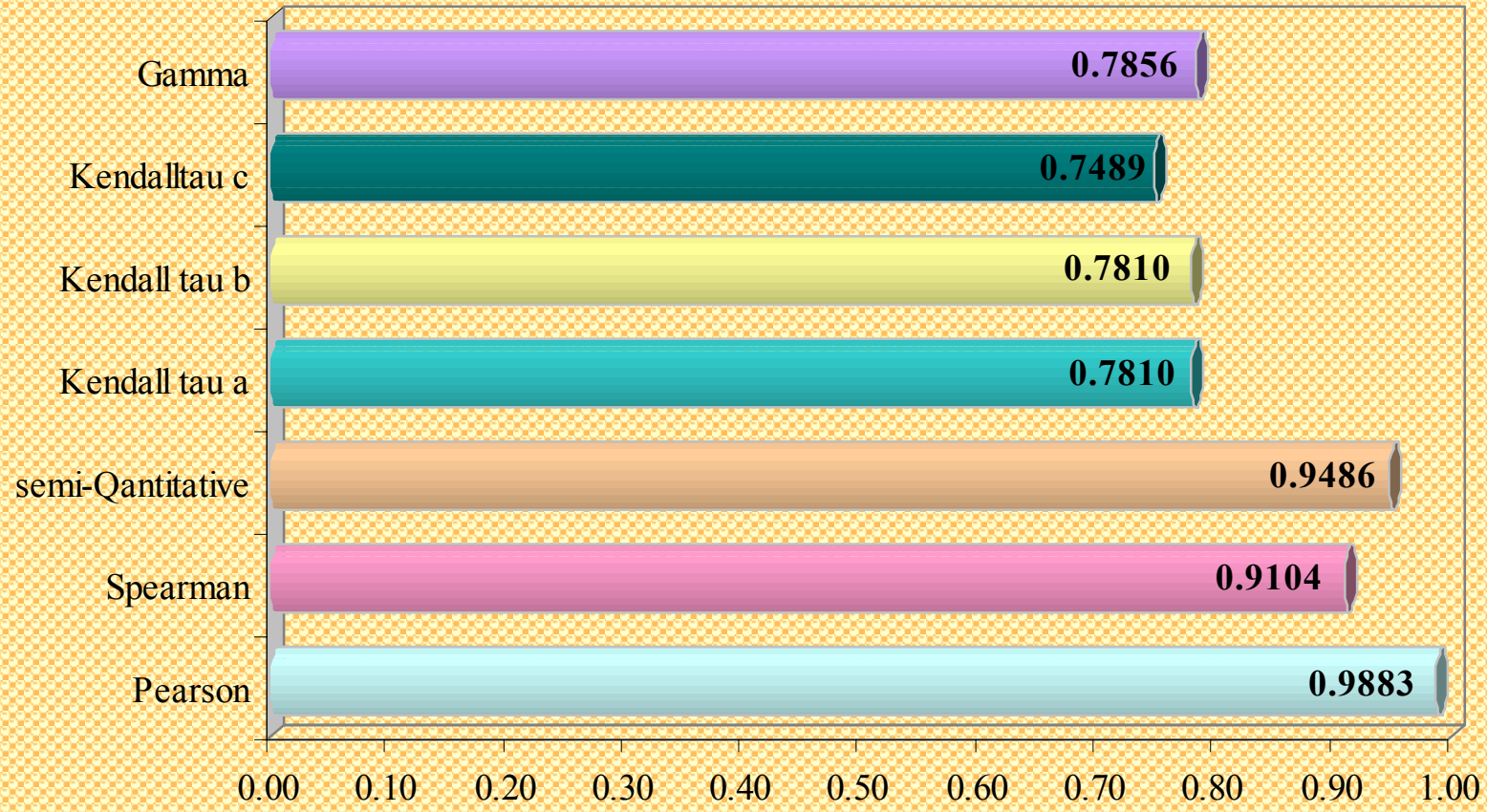
Pearson	95%CI r_{Prs} [0.9694-0.9931]	r_{Prs}^2 = 0.9712	$t_{Prs,1}$ = 30.71†
Spearman	95%CI ρ_{Spm} [0.7335-0.9342]	ρ_{Spm}^2 = 0.7485	$t_{Spm,1}$ = 9.13†
semi-Qantitative	95%CI r_{sQ} [0.8438-0.9632]	r_{sQ}^2 = **0.8526**	t_{sQ} = 12.73†
Kendall a	95%CI $\tau_{Ken,a}$ [0.4741-0.8537]	$\tau_{Ken,a}^2$ = 0.5079	$Z_{Ken,\tau a}$ = 5.53†
Kendall b	95%CI $\tau_{Ken,b}$ [0.4844-0.8572]	$\tau_{Ken,b}^2$ = 0.5173	$Z_{Ken,\tau b}$ = 5.54†
Kendall c	95%CI $\tau_{Ken,c}$ [0.4370-0.8405]	$\tau_{Ken,c}^2$ = 0.4746	$Z_{Ken,\tau c}$ = 5.35†
Gamma	95%CI $_{\Gamma}$[0.4925-0.8600]	Γ^2 = 0.5246	Z_{Γ} = 4.07†

† p-value less than 0.0001

$$\min(Y) \le HM(Y) \le GM(Y) \le AGM(Y) \le AM(Y) \le EM(Y) \le \max(Y)$$

Results for MDF SARs

for 1.74 - 9261·*iSMMWHg* + 10.34·*iAMdEHg* + 3.89·*INDRLQg*
(3v structure vs. pI_{50}, from [idem])



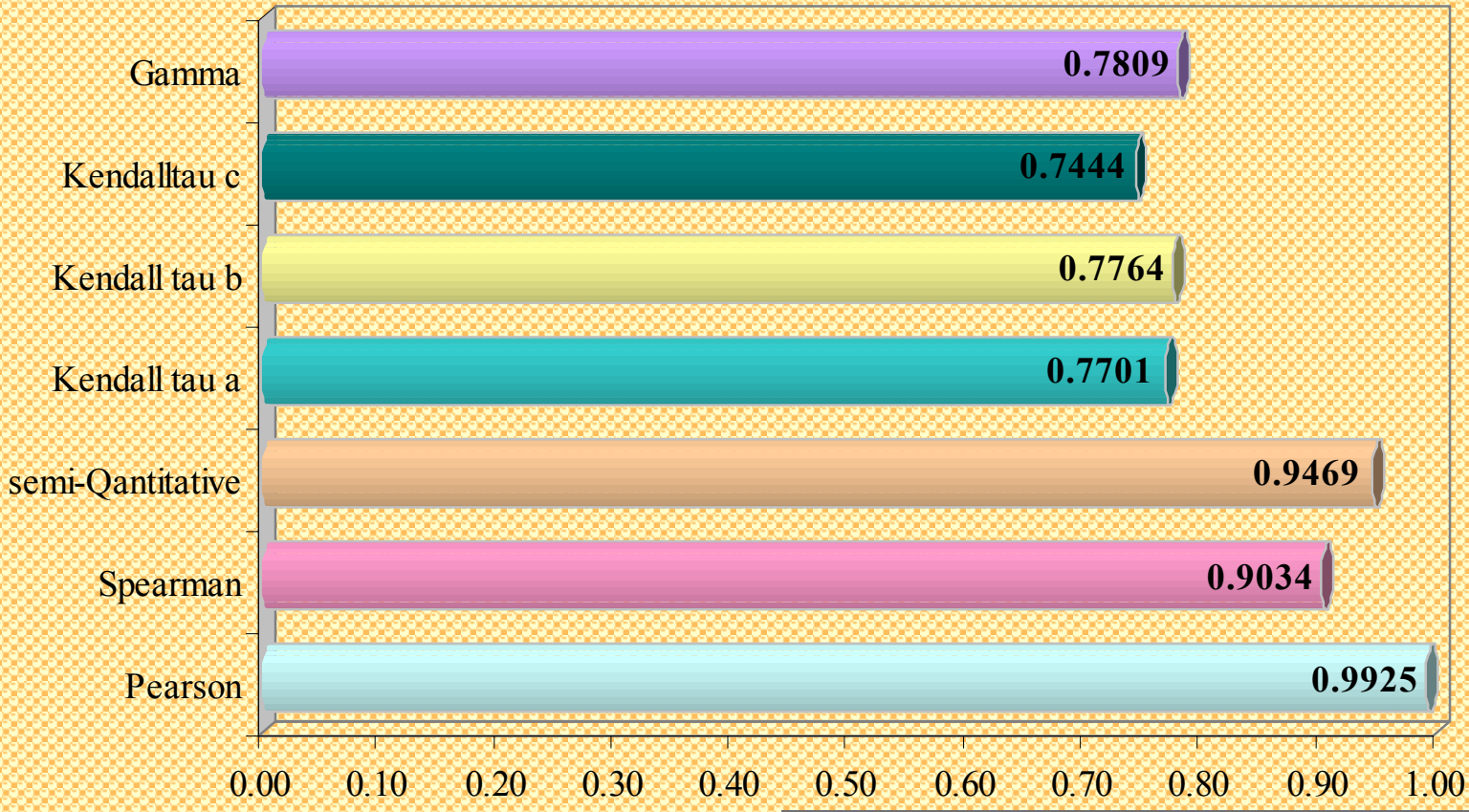
Results for MDF SARs

Pearson	95%CI r_{Prs} [0.9752-0.9944]	r_{Prs}^2 = 0.9768	$t_{Prs,1}$ = 34.35†
Spearman	95%CI ρ_{Spm}[0.8185-0.9568]	ρ_{Spm}^2 = 0.8288	$t_{Spm,1}$ = 11.64†
semi-Qantitative	95%CI r_{sQ} [0.8937-0.9754]	r_{sQ}^2 = **0.8998**	t_{sQ} = 15.85†
Kendall a	95%CI $\tau_{Ken,a}$ [0.5854-0.8906]	$\tau_{Ken,a}^2$ = 0.6002	$Z_{Ken,\tau a}$ = 6.01†
Kendall b	95%CI $\tau_{Ken,b}$ [0.5854-0.8906]	$\tau_{Ken,b}^2$ = 0.6099	$Z_{Ken,\tau b}$ = 6.02†
Kendall c	95%CI $\tau_{Ken,c}$ [0.5321-0.8734]	$\tau_{Ken,c}^2$ = 0.5608	$Z_{Ken,\tau c}$ = 5.82†
Gamma	95%CI $_{\Gamma}$ [0.5931-0.8930]	Γ^2 = 0.6171	Z_{Γ} = 4.79†

† p-value less than 0.0001

Results for MDF SARs

for 5.75 + 199·*iSMmEQt* - 9010·*iSMMWHg* - 0.071·*LADmkQt* + 2.86·*INPRJQg (2v structure vs. pI_{50}, from* [*ibidem*])



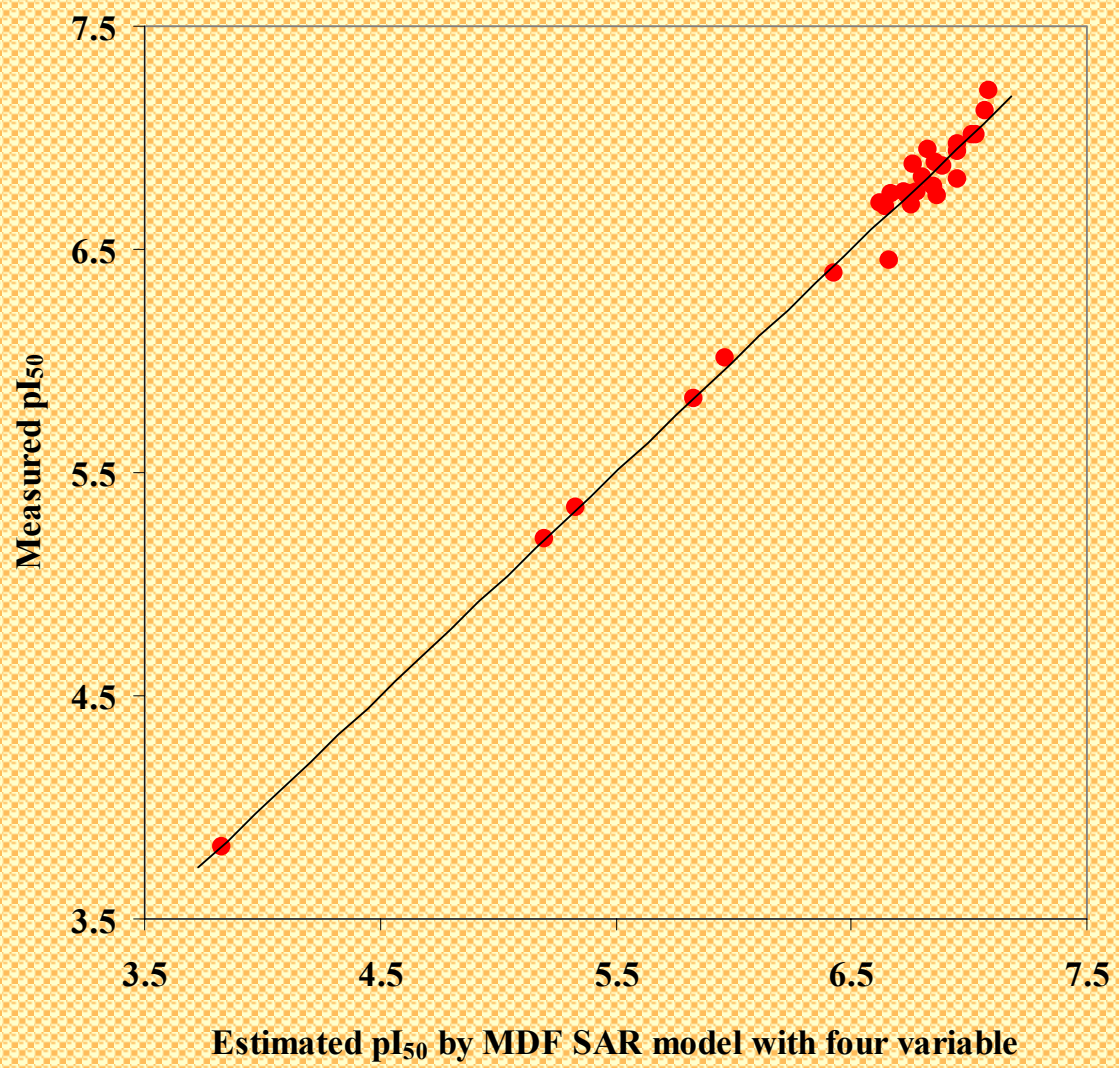
Results for MDF SARs

Plot:

- Cross-validation leave-one-out correlation score (more correlation coefficient !!):

  $r^2_{cv\text{-}loo} = 0.9849$

- Steiger's Z test (vs. previously reported QSAR) = 2.828 (p-value < 0.05)



Results for MDF SARs

Pearson	95%CI r_{Prs} [0.9841-0.9964]	r_{Prs}^2 = 0.9850	$t_{Prs,1}$ = 42.85
Spearman	95%CI ρ_{Spm} [0.8051-0.9533]	ρ_{Spm}^2 = 0.8162	$t_{Spm,1}$ = 11.15
semi-Qantitative	95%CI r_{sQ} [0.8903-0.9746]	r_{sQ}^2 = **0.8967**	t_{sQ} = 15.59
Kendall a	95%CI $\tau_{Ken,a}$[0.5671-0.8848]	$\tau_{Ken,a}^2$ = 0.5931	$Z_{Ken,\tau a}$ = 5.98
Kendall b	95%CI $\tau_{Ken,b}$[0.5776-0.8881]	$\tau_{Ken,b}^2$ = 0.6028	$Z_{Ken,\tau b}$ = 5.98
Kendall c	95%CI $\tau_{Ken,c}$ [0.5248-0.8710]	$\tau_{Ken,c}^2$ = 0.5542	$Z_{Ken,\tau c}$ = 5.78
Gamma	95%CI Γ [0.5852-0.8905]	Γ^2 = 0.6098	Z_{Γ} = 4.73

† p-value less than 0.0001

Concluding Remarks

- All seven methods for appreciating of the correlation between measured and estimated herbicidal activity of studied triazines analogues had statistical significance (see tables) → any of can serve for biological activity estimators evaluating (!)

Concluding Remarks

- **If** … is considered as being a quantitative variable, **then** ***Pearson*** correlation coefficients is the statistical parameter that can be use in evaluation of relationship between the measured activity an the activity estimated by models.
- **If** … is considering as being a qualitative variable (only the relative ordering can be assumed to be reproductible) **then** Spearman (or Kendall …)
- **If** … is considered as being a semi-quantitative variable, **then** proposed semi-Quantitative r is the best SAR evaluator

Acknowledgements:

- The research was partly supported by UEFISCSU Romania through projects ET36/2005 & ET108/2006.
- A. v. Humboldt Foundation
- Organizers of I.S.O.C.

All are online (!!):

http://vl.academicdirect.org/molecular_topology

/mdf_findings/

MDF (Demo) Calculator
MDF SAR Predictor
Leave One Out Analysis
MDF Investigator
Training vs. Test Experiment

/mdf_findings/rank/

[presented methods]

Thank you for your attention!

http://sorana.academicdirect.ro
&
http://lori.academicdirect.org

Activity Characterization of Triazines Analogues: Statistical Parameters for Models Assessment

Sorana Daniela BOLBOACĂ[1], Lorentz JÄNTSCHI[2]

[1] *„Iuliu Haţieganu" University of Medicine and Pharmacy, 13 Emil Isac Street, 400023 Cluj-Napoca, Romania. E-mail: sbolboaca@umfcluj.ro*

[2] *Technical University of Cluj-Napoca, Romania, 15 Constantin Daicoviciu Street, 400020 Cluj-Napoca, Romania*

Abstract

Correlation coefficients and associated squared values are used as parameters in validation of structure-activity relationship models. By using of the molecular descriptors family on structure-activity relationship method [1] the herbicidal activity of a sample of triazines analogues was modelled [2]. A number of three multivariate models proved to have estimation and prediction abilities [2].

Starting from the hypothesis that the measured activity of triazines analogue is a semi-quantitative variable, the aim of the research was to analyzed the three previously reported models by using the Pearson, Spearman, Kendall's and Gamma correlation coefficients.

The structure-activity relationship models were previously reported [2]. The measured herbicidal activity of triazines analogues and the value estimated by the previously reported models were investigated by using the Pearson, Spearman, Kendall's τ_a, τ_b, τ_c and Gamma correlation coefficients (r_{Prs}, ρ_{Spm}, $\tau_{Ken,a}$, $\tau_{Ken,b}$, $\tau_{Ken,c}$, Γ) and associated squared correlation coefficient (r_{Prs}^2, ρ_{Spm}^2, $\tau_{Ken,a}^2$, $\tau_{Ken,b}^2$, $\tau_{Ken,c}^2$, Γ^2).

The results of investigation, express as correlation coefficients and associated 95% confidence intervals, squared correlation coefficient and Student's t, respectively the parameter of the Z test were calculated and analyzed.

The correlation coefficients obtained with all methods were statistical significant ($p < 0.0001$). The correlation coefficients vary according with the model as follow:

- Model with two descriptors: from r = 0.6889 ($\tau_{Ken,c}$) - 95%CI [0.4370-0.8405] to r = 0.9855 (r_{Prs}) - 95% CI [0.9694-0.9931];
- Model with three descriptors: from 0.7489 ($\tau_{Ken,c}$) - 95%CI [0.5321-0.8734] to the value equal with 0.9883 (r_{Prs}) - 95% CI [0.9752-0.9944];

- Model with four descriptors: from 0.7444 ($\tau_{Ken,c}$) - 95%CI [0.5248-0.8710] to the value equal with 0.9925 (r_{Prs}) - 95% CI [0.9841-0.9964].

If there is considered that the herbicidal activity of triazines analogues is a quantitative variable, the Pearson correlation coefficients is the statistical parameter that must be used in evaluation of relationship between the measured and estimated activity.

Considering the measured activity of triazines analogues as a semi-quantitative variable, a rank correlation coefficient is the statistical parameter able to provide more reliable estimation rather than Pearson correlation coefficient. However, which is the proper rank correlation coefficients that can be use in this circumstance? The comparisons between the rank correlation coefficients are discussed and the further plan of research is highlighted.

Keywords: Triazines Analogues, Statistical Models, Correlation Coefficients

Acknowledgements

The research was partly supported by UEFISCSU Romania through projects ET36/2005 & ET108/2006.

References

[1] Jäntschi L. Molecular Descriptors Family on Structure Activity Relationships 1. Review of the Methodology. Leonardo Electronic Journal of Practices and Technologies 2005;6:76-98.

[2] Bolboacă S., Jäntschi L. Molecular Descriptors Family on Structure-Activity Relationships: Modeling Herbicidal Activity of Substituted Triazines Class. Bulletin of University of Agricultural Sciences and Veterinary Medicine - Agriculture 2006;62:35-40.