



Universiteit
Leiden
The Netherlands

Statistical integration of diverse omics data

Bouhaddani, S. el

Citation

Bouhaddani, S. el. (2020, June 2). *Statistical integration of diverse omics data*. Retrieved from <https://hdl.handle.net/1887/92366>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/92366>

Note: To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/92366> holds various files of this Leiden University dissertation.

Author: Bouhaddani, S.

Title: Statistical integration of diverse omics data

Issue Date: 2020-06-02

**Stellingen behorend bij het proefschrift getiteld
“Statistical integration of diverse omics data”**

1. The maximum likelihood estimator for Probabilistic PLS is less prone to overfitting, in terms of the prediction error, than the PLS estimator, since the former is a full-information maximum likelihood, and the latter is a limited information maximum likelihood. (this thesis)
2. When applying probabilistic data integration models, specific parts are needed to obtain unbiased estimates of the joint parts. (this thesis)
3. In the SEM modeling framework, when including data-specific parts, standard cross-validation of the prediction error of Y given X over a full grid is intrinsically wrong. (this thesis)
4. Probabilistic data integration models such as PO2PLS are highly robust against non-normality of the data. (this thesis)
5. Hair (2014) claims that “PLS-SEM has several advantages over ML-SEM in many situations commonly encountered in social sciences research, for example, when sample sizes are small, or when complex models with many indicators are estimated.” With respect to small sample size and high dimensionality, maximum likelihood SEM is better in terms of estimation bias and prediction. (this thesis)
6. Structural Equation Models are a natural choice to describe biological pathways, they are not too specific, nor too general.
7. Integration of omics data is preferred over integration of results from omics data analyses.
8. Without knowledge of methodology, software, and biology, one cannot be a good biostatistician.
9. Integration of multiple disciplines greatly improves integration of multiple omics data.
10. While in chemometrics and social sciences, “significance testing is less important in SEM than in other multivariate techniques” (Westland, 2015), in life sciences, statistical inference is crucial to draw conclusions about a population.
11. Interpreting output of statistical software packages should not be reduced to “sterrenkijken” (stargazing).
12. Researchers should invest more time for online visibility and social media.
13. Good scientific conduct can make the difference between life and death.