

## Quality Analysis ,Before' with Respect to Good Practice in Statistical Reporting and a Learner as the User of the Statistical Report

We consider the prospective user to be a ,learner' of statistics or data analysis wishing to learn how to check if the assumptions of (classical) multiple linear regression are fulfilled (e.g. Regression Diagnostics.)

Following criteria are used to check the quality of the output of ChatGPT (example for the chat: <https://chatgpt.com/share/e2a81986-7aed-40cd-b536-65402192dcaa>) :

1. Does it contain a brief explanatory data analysis at the beginning?
2. Does it stipulate the applied statistical model?
3. Does it explain the assumptions of (classical) multiple linear regression (resp. violations of the assumptions)?
4. Does it address all assumptions and topics to be checked?
5. For each diagnostic category:
  - a. Is there an introduction?
  - b. Are there several diagnostic methods included (e.g. at least a plot and at least a statistical test) per category?
  - c. For each diagnostic method (e.g. plot or statistical test):
    - i. Is the method correct?
    - ii. Is the method explained to the learner (e.g. axes and lines for a plot, statistical test function, interpretation...)?
    - iii. Are there references?
6. Does it contain a summary of the regression diagnostics?
7. Does the output contain mathematical equations?
8. Is the output well-structured and learner friendly?
9. Does it generate reproducible Python code for each generated output?
10. Is the prompt reproducible over several datasets (with ChatGPT4o)?

Criteria	Result
1	No
2	No
3	No
4	No, outliers and influential points are missing.
5a	Only a short introduction
5b	Only normality of residuals is tested with a plot and a statistical test
5ci	Yes. No negative examples for the dataset cacao.csv
5cii	Plots: axes and lines are not explained. Statistical tests: Test statistic or p-value is given to make assumptions  Used methods are not explained, neither plots nor statistical tests. No information about the used method (e.g. why is the method used, are there different methods, what can you do with this method, ...)
5ciii	No
6	Yes
7	No
8	1. No, the output is structured by the steps performed by the AI. These steps do not lead to the best structure for a learner

	<ol style="list-style-type: none"> <li>2. The notes delivered by <code>model.summary()</code> and included in the output (Notes: [1] Standard Errors assume that the covariance matrix of the errors is correctly specified. [2] The condition number is large, <math>4.1e+05</math>. This might indicate that there is strong multicollinearity or other numerical problems) These notes are not explained. Either one should explain them or hide them in the output.</li> <li>3. The statistics delivered by <code>model.summary()</code>: Durbin-Watson and Jarque-Bera are not explained. Either one should explain them or hide them in the output.</li> <li>4. No need for model summary before diagnostics</li> <li>5. Each line of the output should have a meaning for the learner</li> </ol>
9	Yes
10	To be checked: Current issues with uploading files to ChatGPT