

***Residual Plots:
The Good, the Bad and the Ugly***

12.05.2017

Paul Schmidt
Biostatistics
Universität
Hohenheim



I have a model. I like my model. What do you want from me?!

- The results of any model are only okay to present if certain assumptions are met. Your results are wrong/useless if these assumptions are not met.

Oh boy! What are these assumptions?

- There are two for ANOVA and one more for regressions:

ANOVA	Regression
Variance Homogeneity	
Normally distributed	
	Linearity

How do I make sure the assumptions are met?

- By looking at residual-plots

What is a residual?

- When you run any model, it tells you for each of your observations (=data points) what it expects/predicts for this specific observation. No model is perfect and therefore the predictions are never exactly the value you actually found in your experiment.

Tell me about it - Weather forecasts are always wrong!

- Yeah... So a residual is the difference between what you measured and what the model predicted: *Residual = Observed – Predicted*

So if a prediction is perfect, the residual is 0?

- Yes!

How many residuals are there?

- Well there is one for every observation in your dataset.

Ok, so let's say I have all the residuals from my model. What now?

- You plot them* and look at them. There are multiple ways to plot them. Here are two:
 - QQ-Plot
 - Res-Pred-Plot

*Strictly speaking, you actually don't plot the raw residuals, but a „studentized“ or „pearson“ version of them. Basically people thought of smart ways to standardize the residuals and then everything is better. R and other software will give you those without any problem, so don't worry about how to do that. Look it up if you're interested, but I bet you are too lazy.

The Res-Pred-Plot



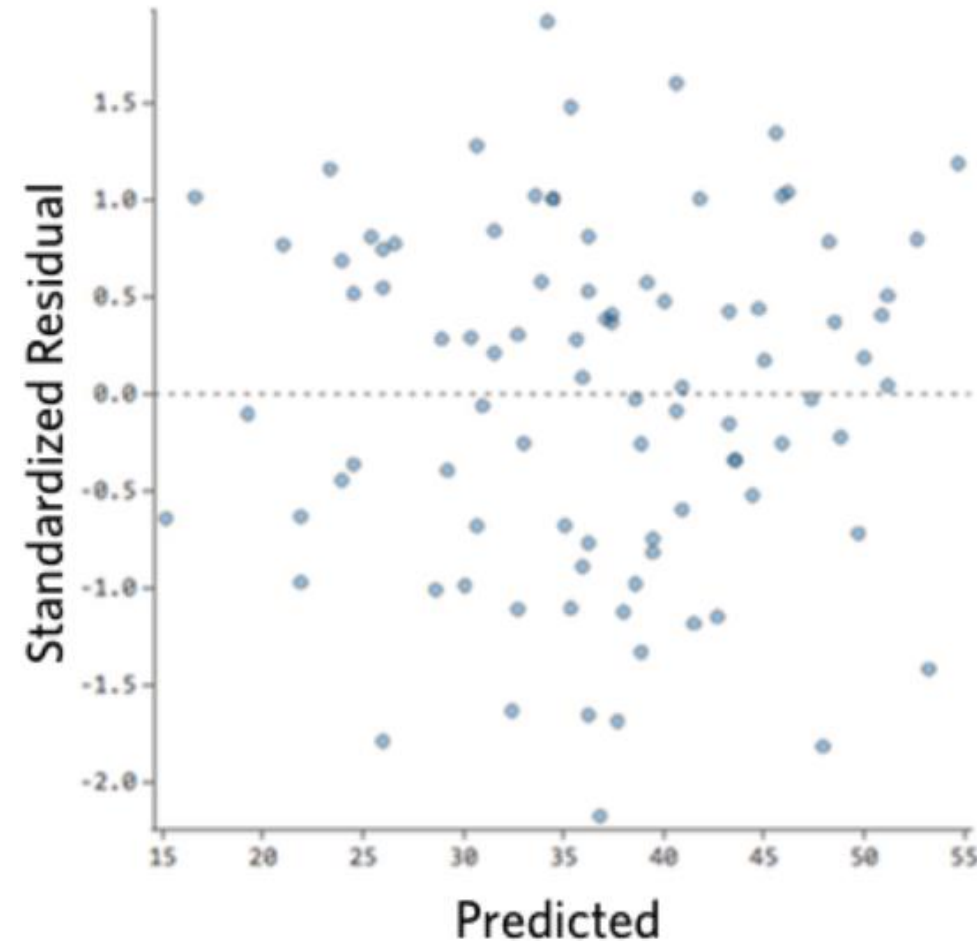
Res-Pred-Plot:

y-axis: Residuals

x-axis: Predictions

Used to evaluate:

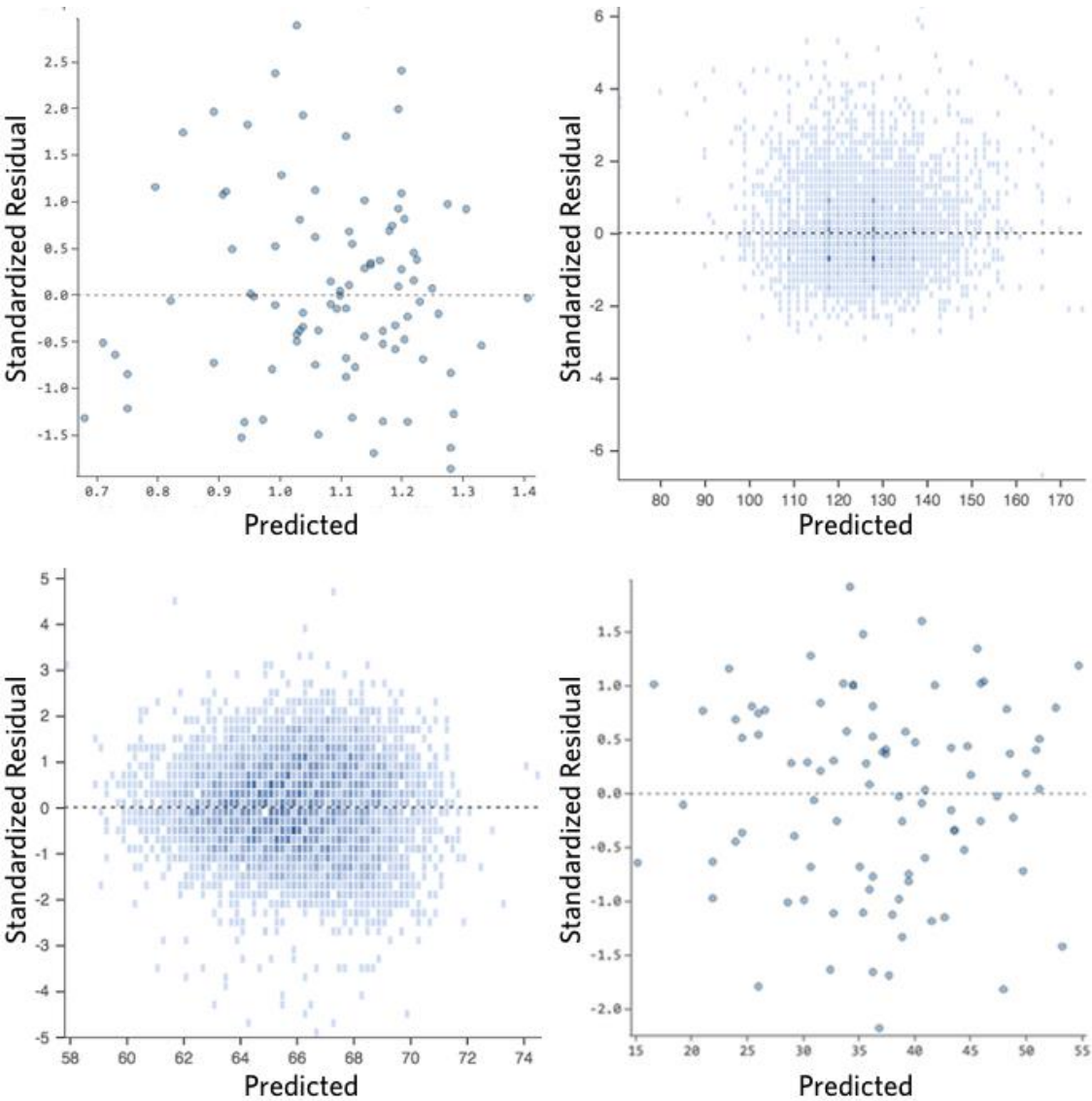
- Variance Homogeneity
- Linearity



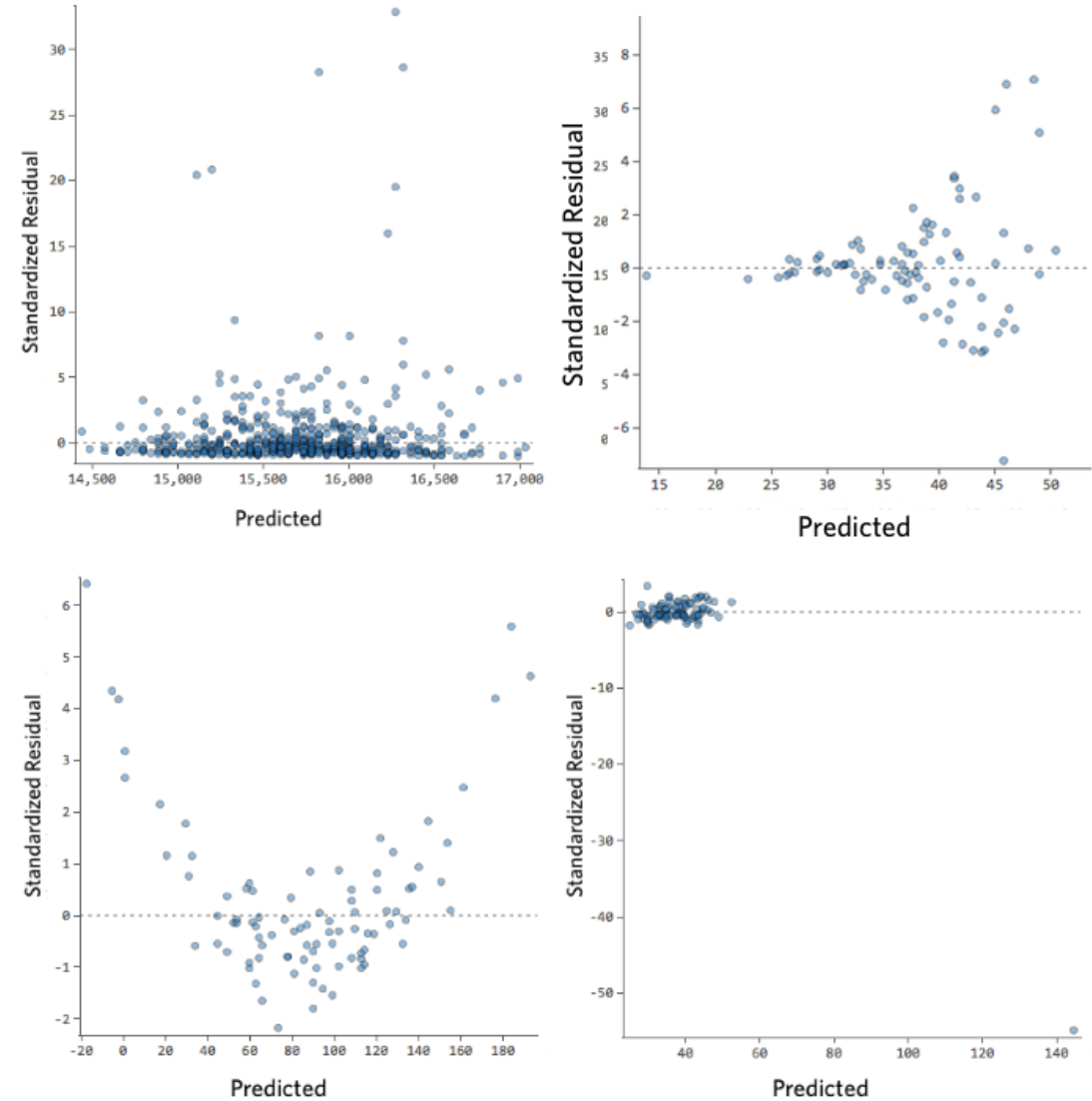
The Res-Pred-Plot



Good:



Bad & Ugly:



The QQ-Plot



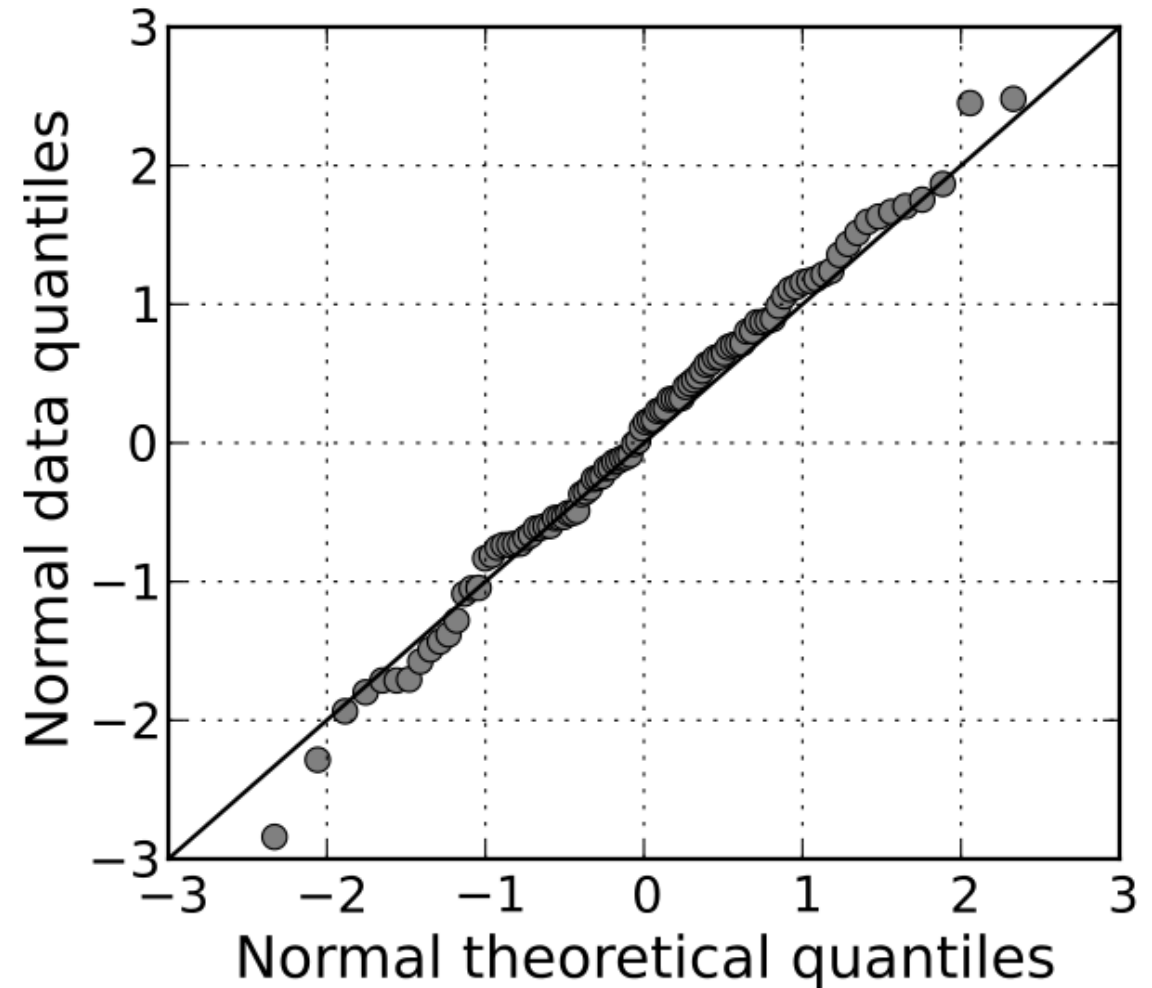
QQ-Plot:

one axis: Sample Quantile*

other axis: Theoretical Quantile*

Used to evaluate:

- (Normal) Distribution

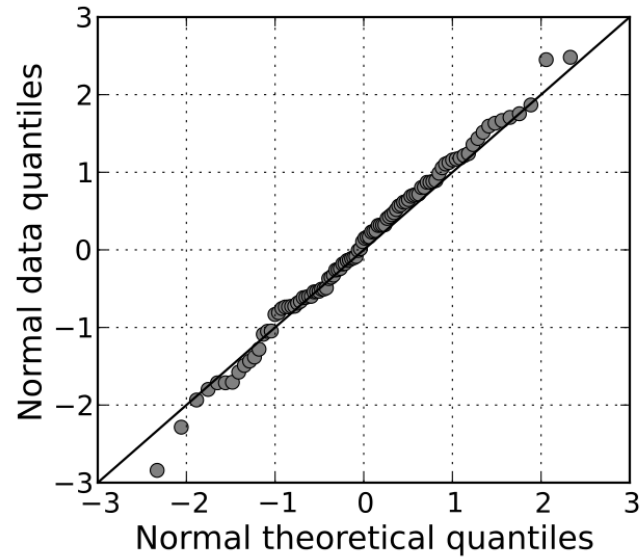
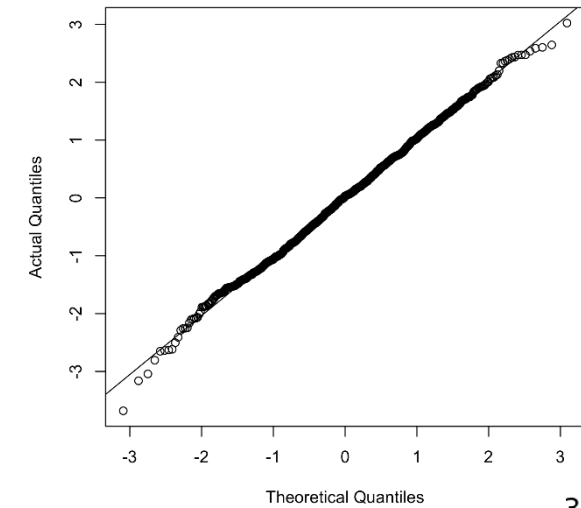


*You don't know what this means? That's okay, you can google it. I just wrote this here so you know how to label your axes.

The QQ-Plot

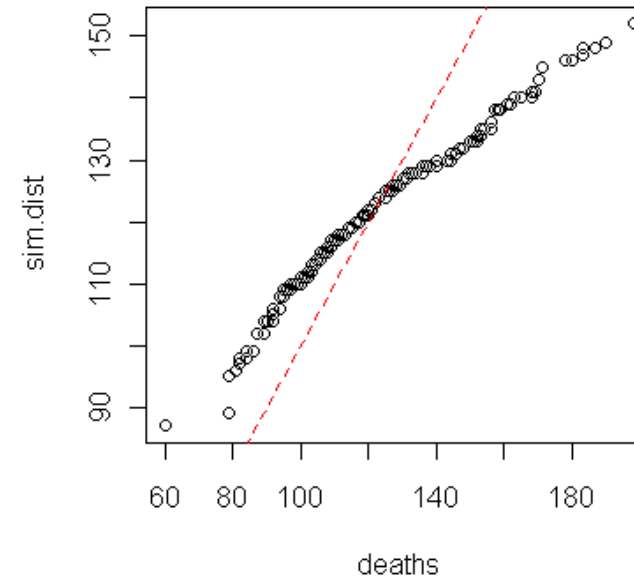


Good:

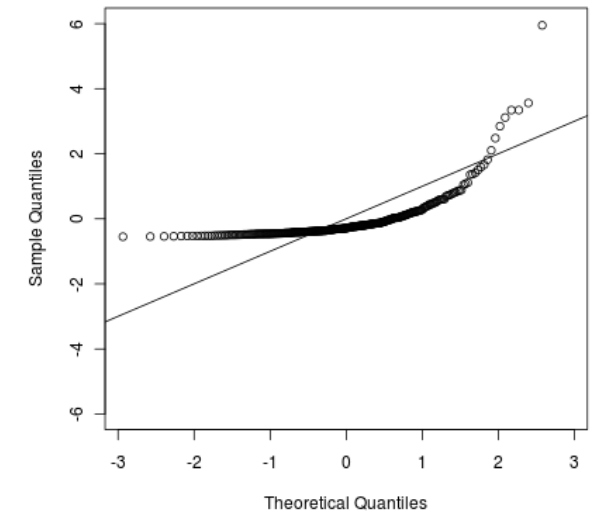


Bad & Ugly:

Q-Q Plot



Normal Q-Q log.normal.data



Got it! So what if my plot does not look good?

- You try transforming your data and look at the new residuals.

And if that does not help?

- You consult a statistician and hope for the best.