Linear Regression: Assumption Checking

Assumptions of the linear regression model

The assumptions for the linear regression model are:

- Linearity.
- Constant spread of residuals.
- Normality of the residuals.
- Independence of the residuals.

Assumption checking

The assumptions of the linear regression model are checked using the:

- Fitted values.
- Residuals.

Residuals

The residual for the ith point is defined as

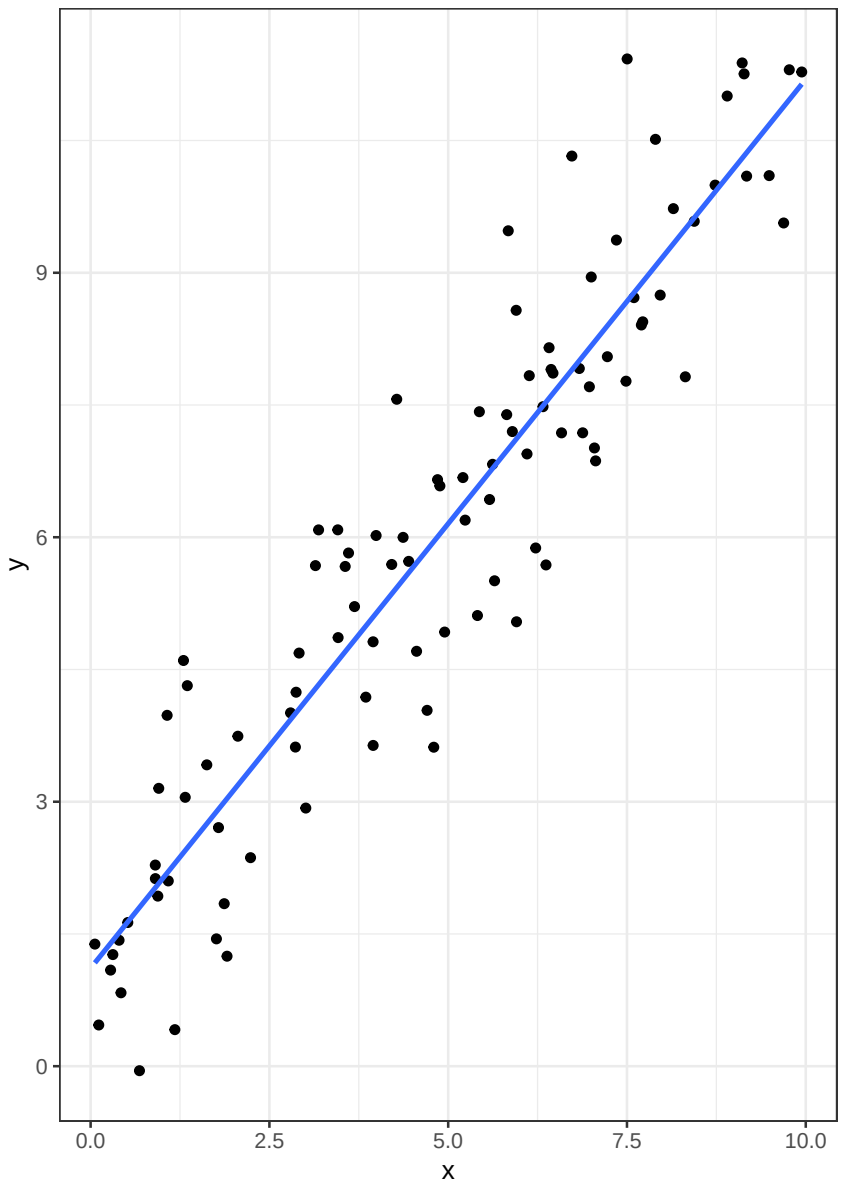
$$r_i = y_i - f_i,$$

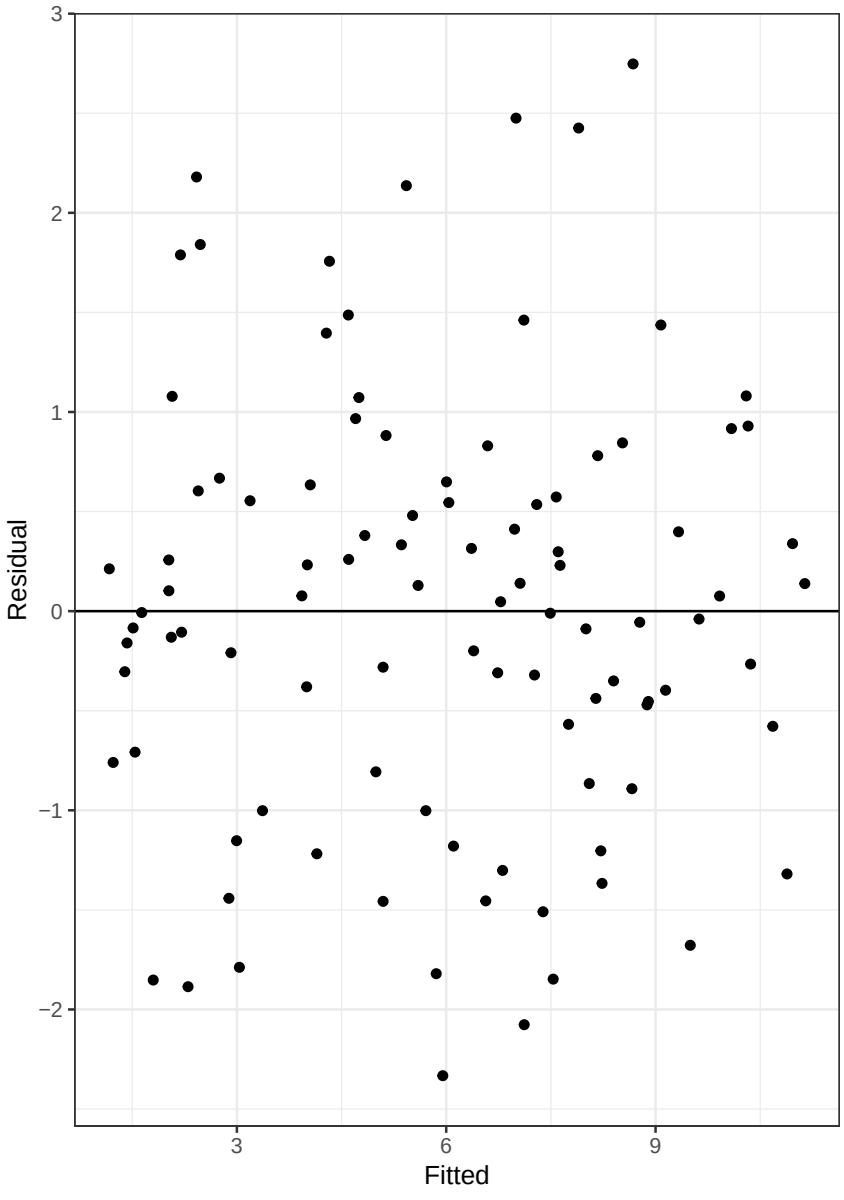
where y_i is the value of the response variable for the i observation, and f_i is the fitted value for the ith point.

Assumption 1: Linearity

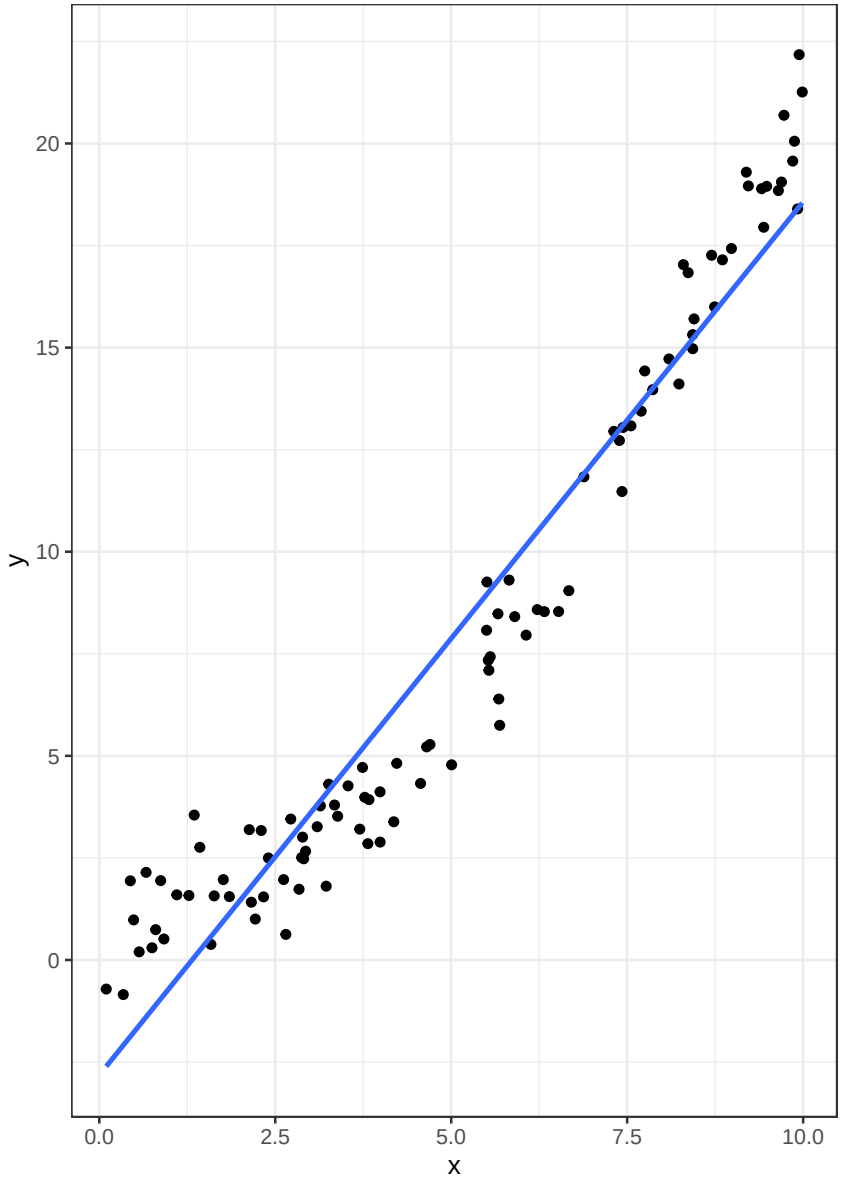
- **Where:** We check this by looking at the residual versus fitted plot.
- **Expect:** We expect to see the points **symmetrically scattered** around the zero line if this assumption is reasonable.

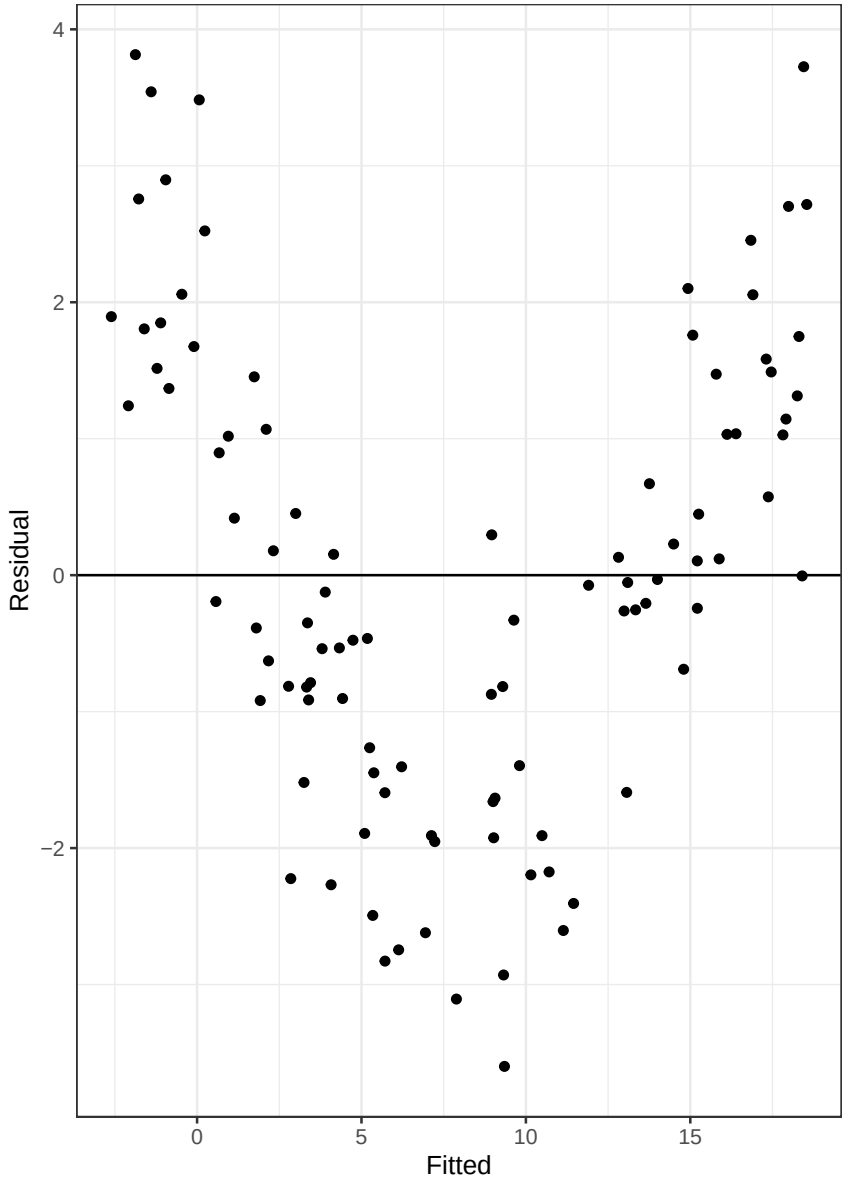
Example of reasonable linearity





An example of unreasonable linearity

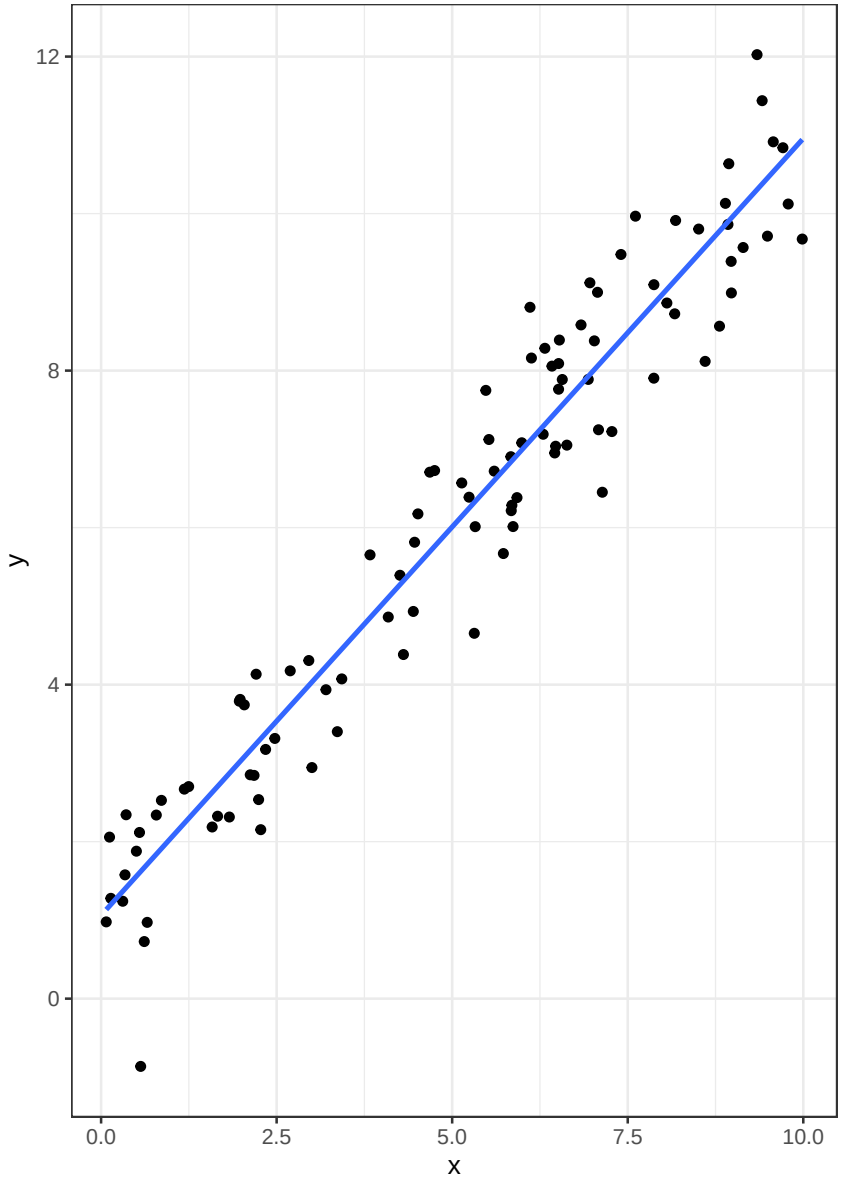


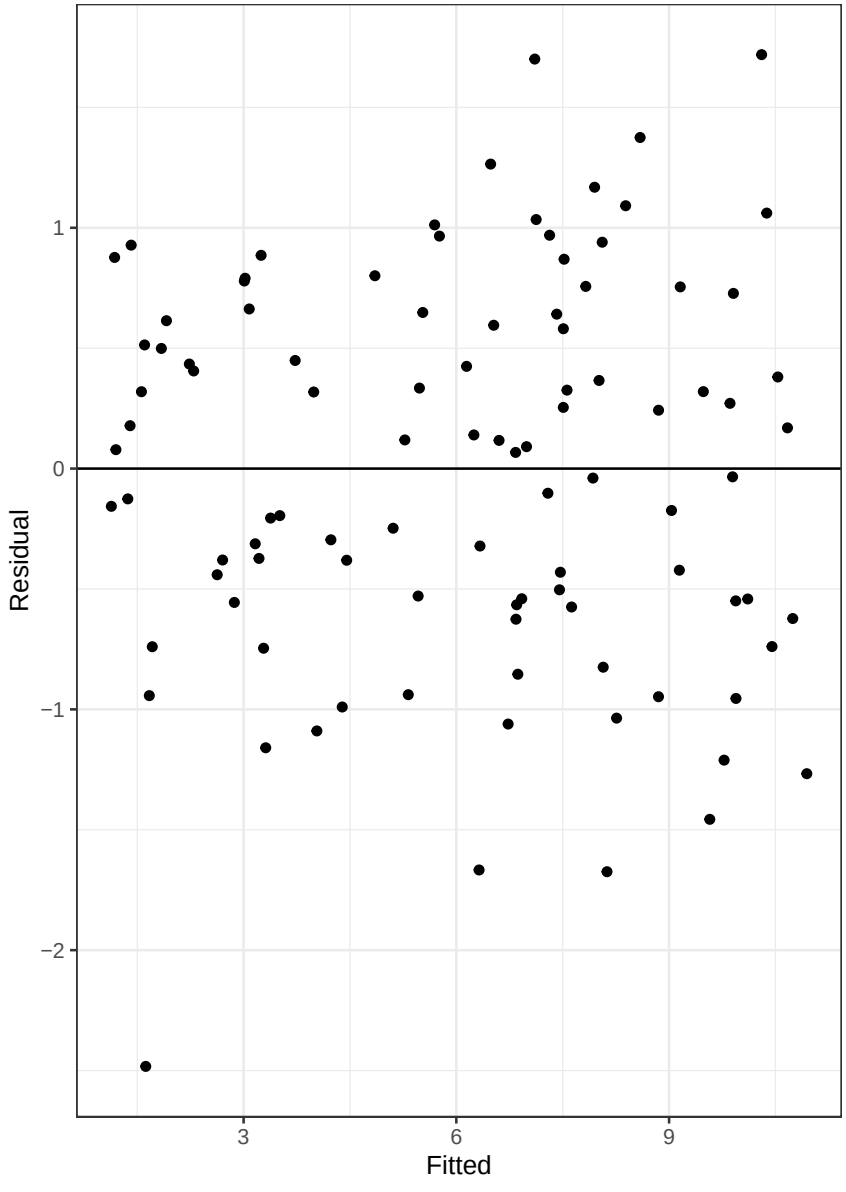


Assumption 2: Constant spread

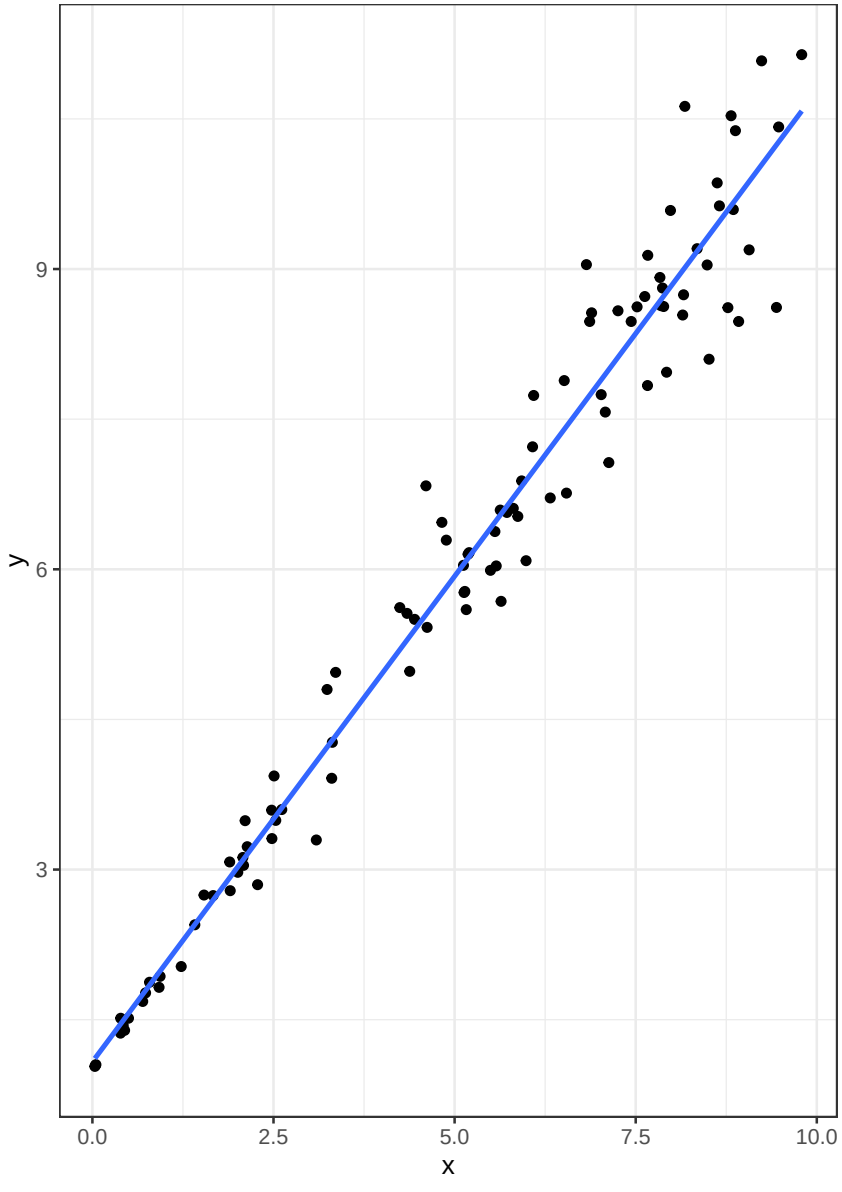
- **Where:** We check this by looking at the residual versus fitted plot.
- **Expect:** We expect to see **roughly equal spread** of the points as we go from left to right.

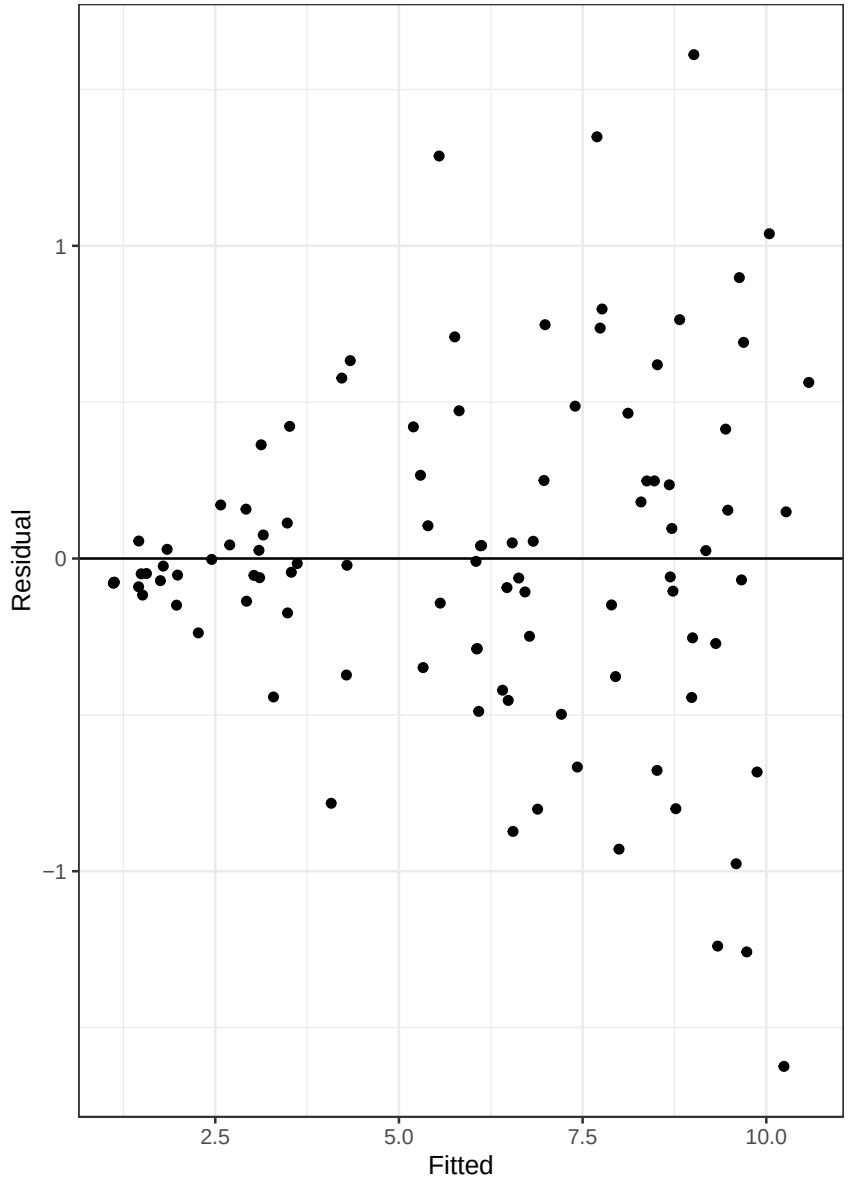
Example of reasonable constant spread



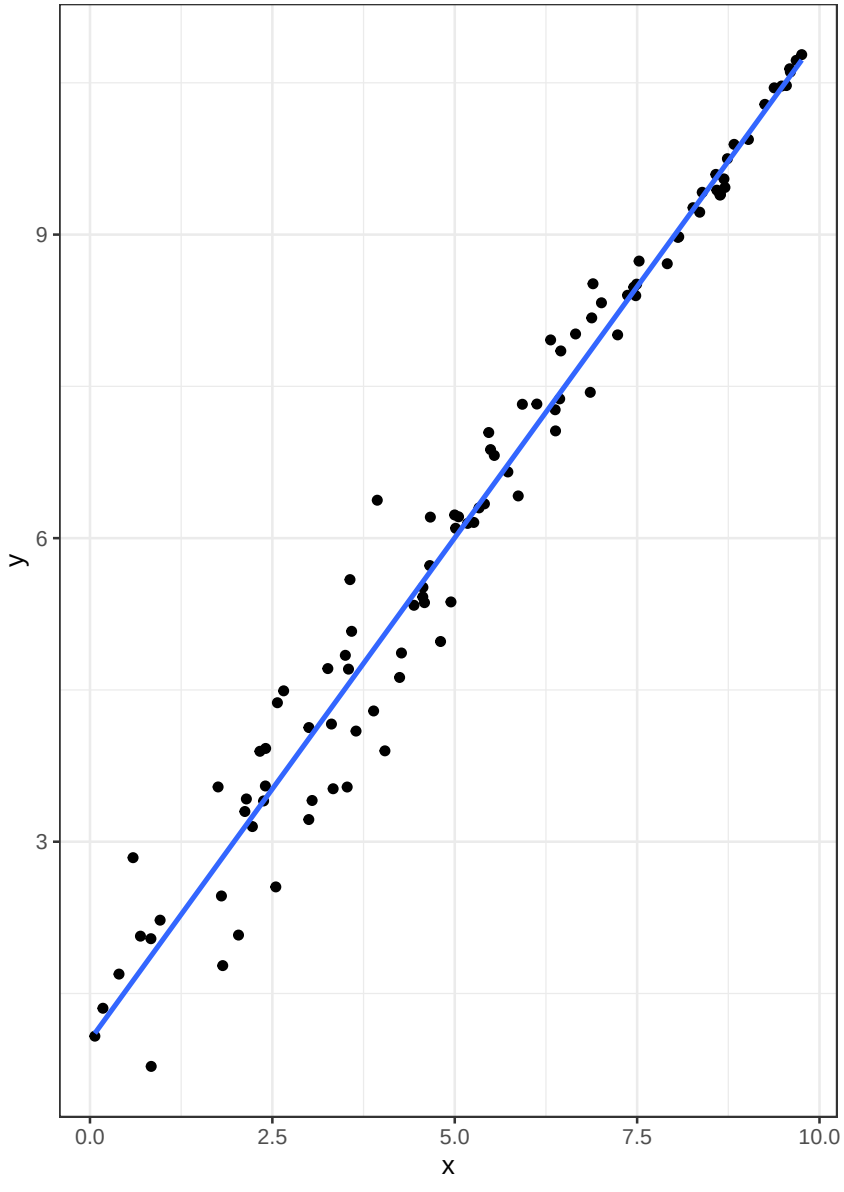


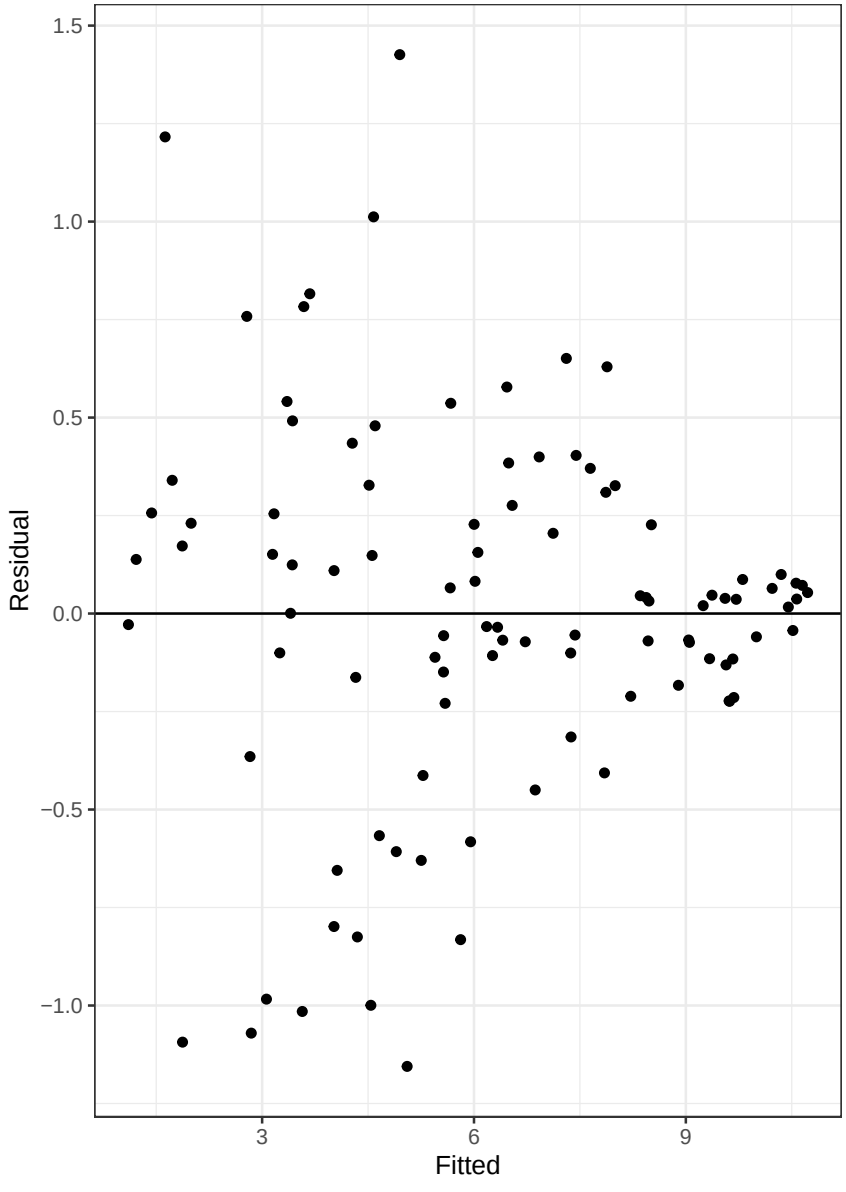
Example of unreasonable constant spread



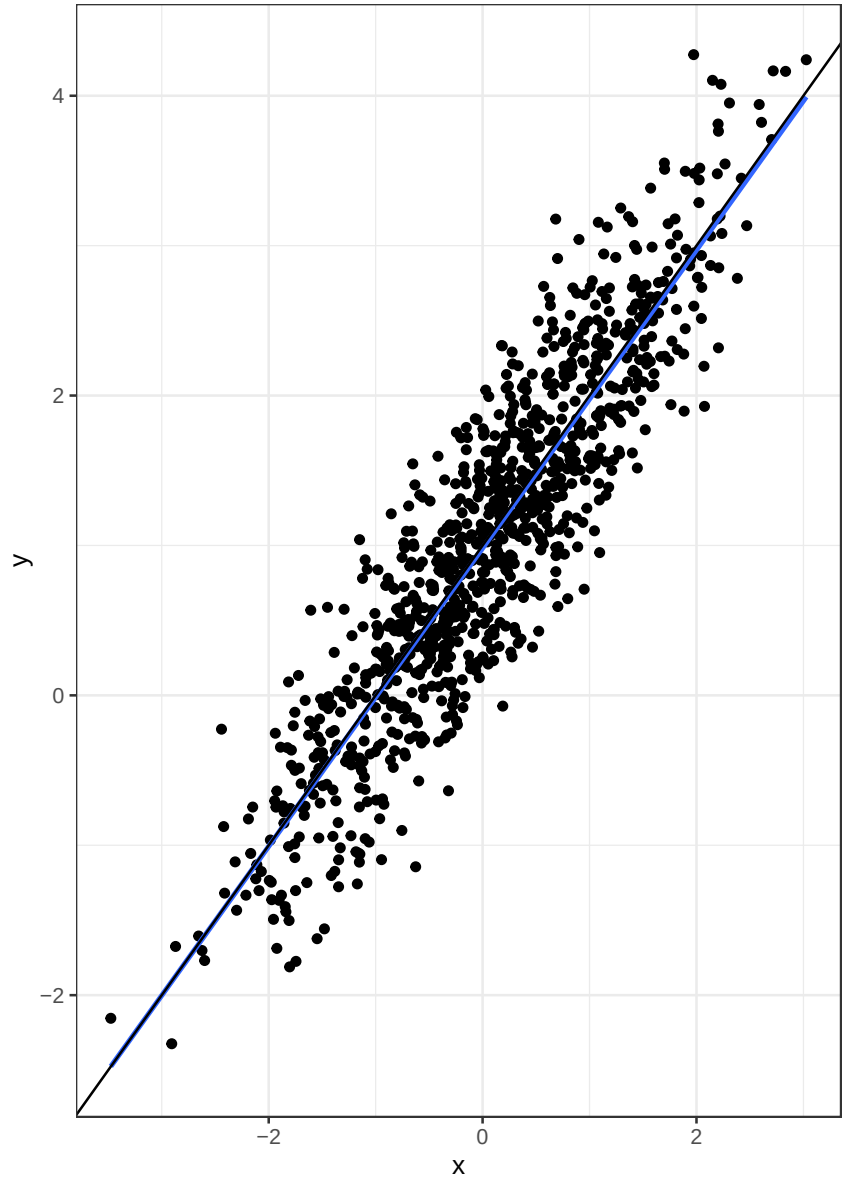


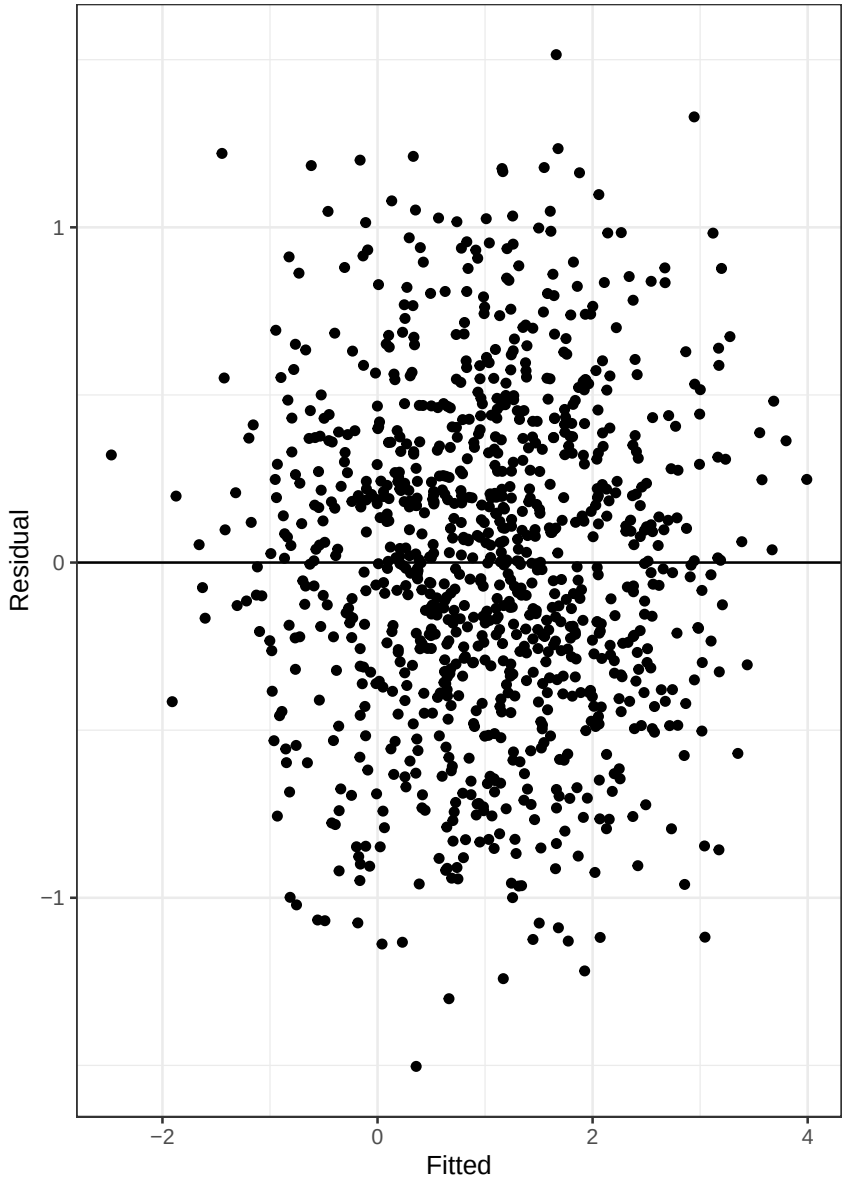
Another example of unreasonable constant spread





Reasonable

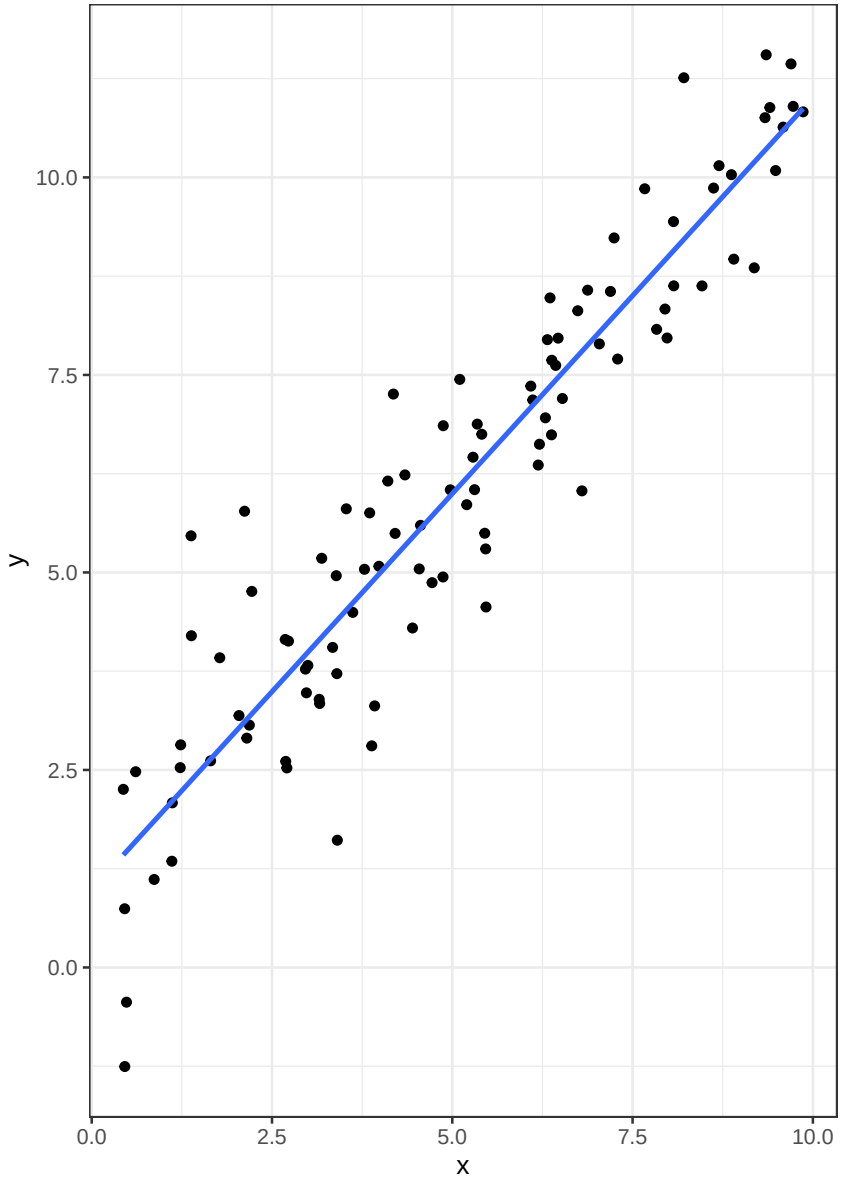


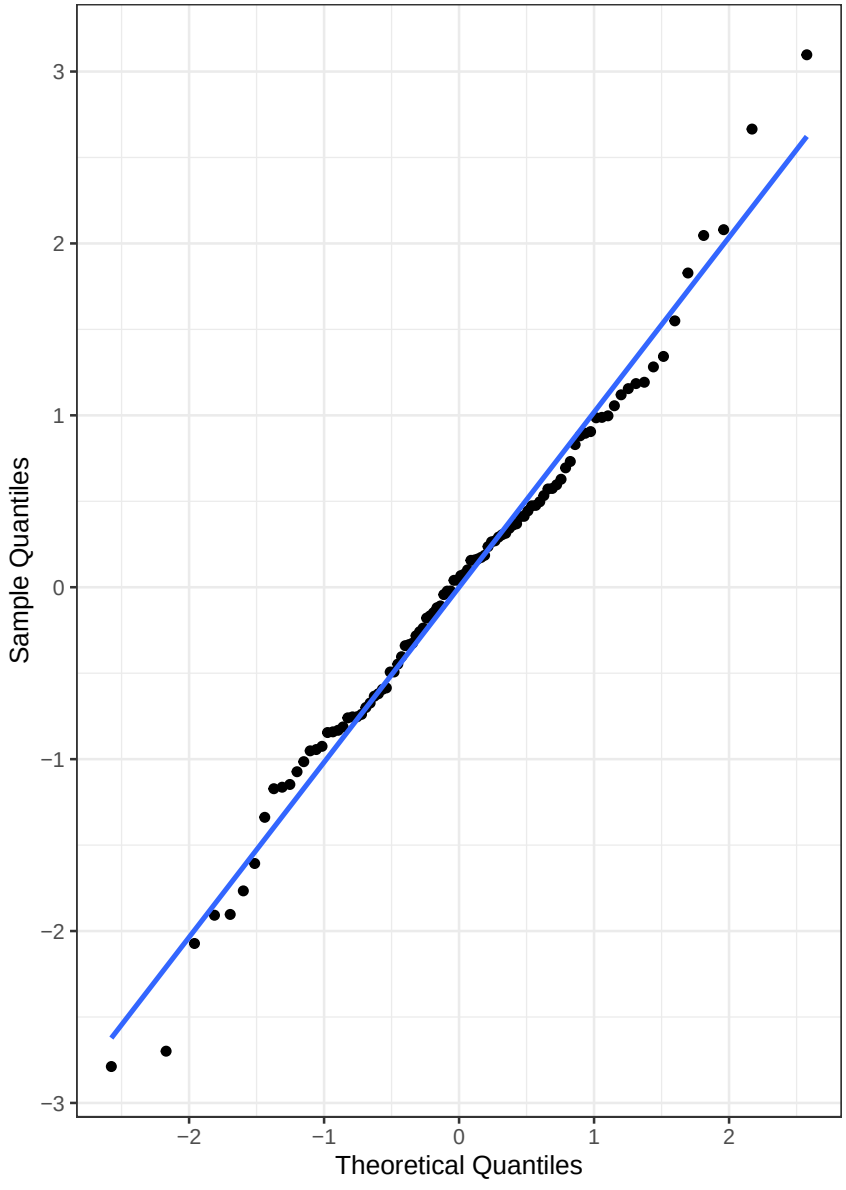


Assumption 3: Normality

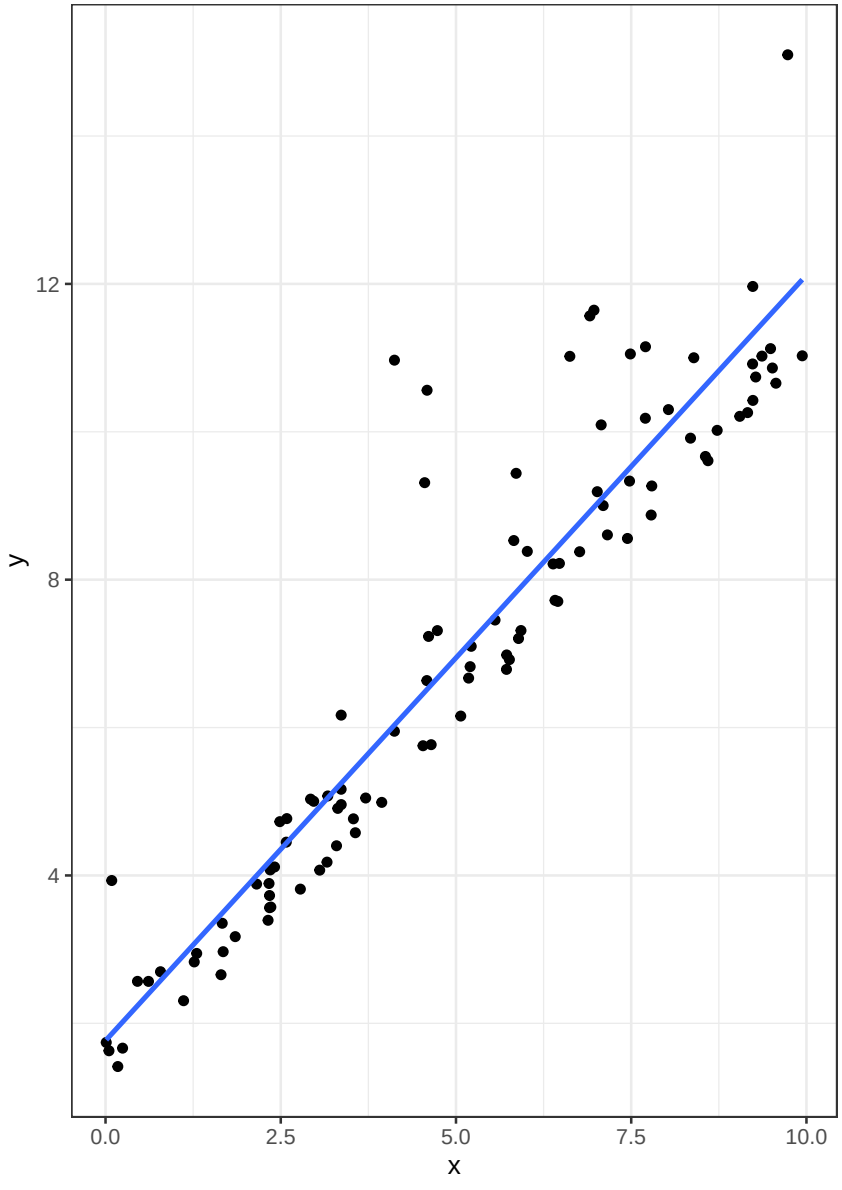
- **Where:** We check this by looking at the normal quantile-quantile plot of the residuals.
- **Expect:** We expect to see **roughly linear** points, i.e., the points lie close to the line (fat pen test).

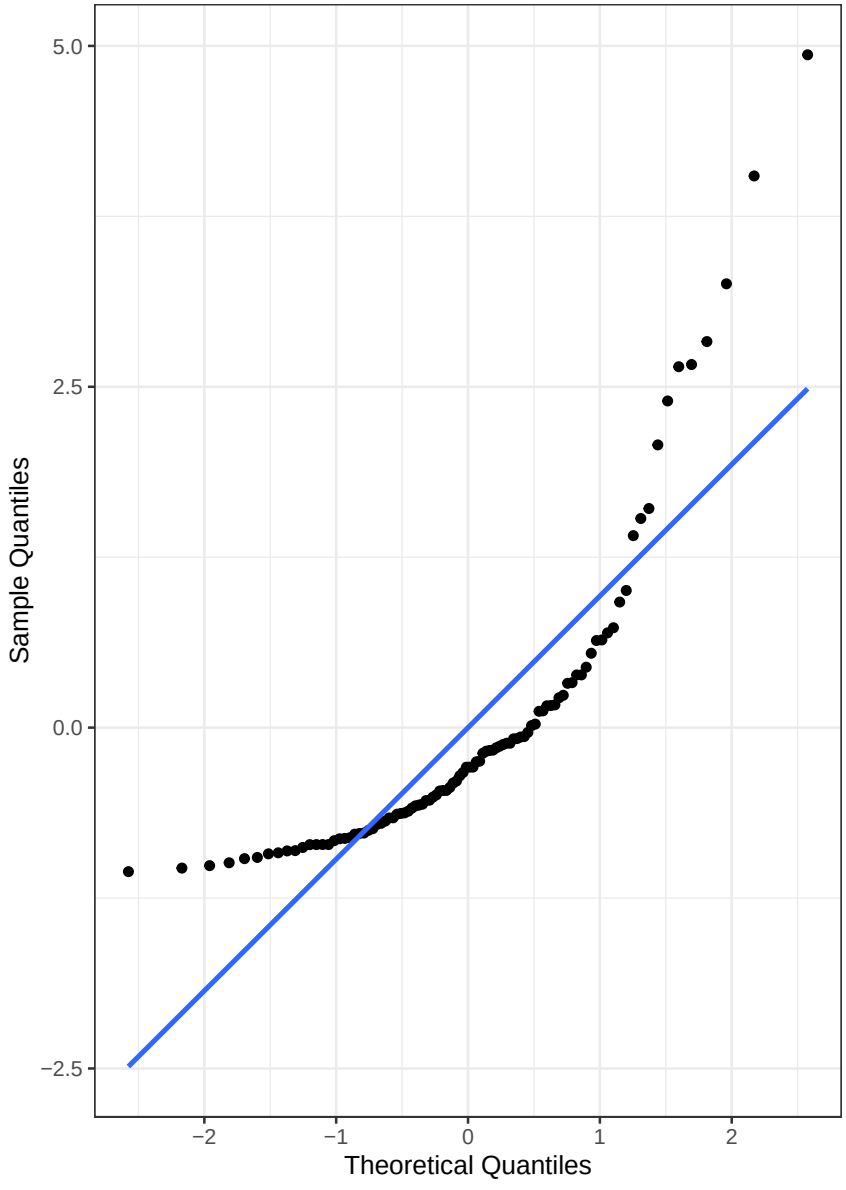
Example of reasonable normality



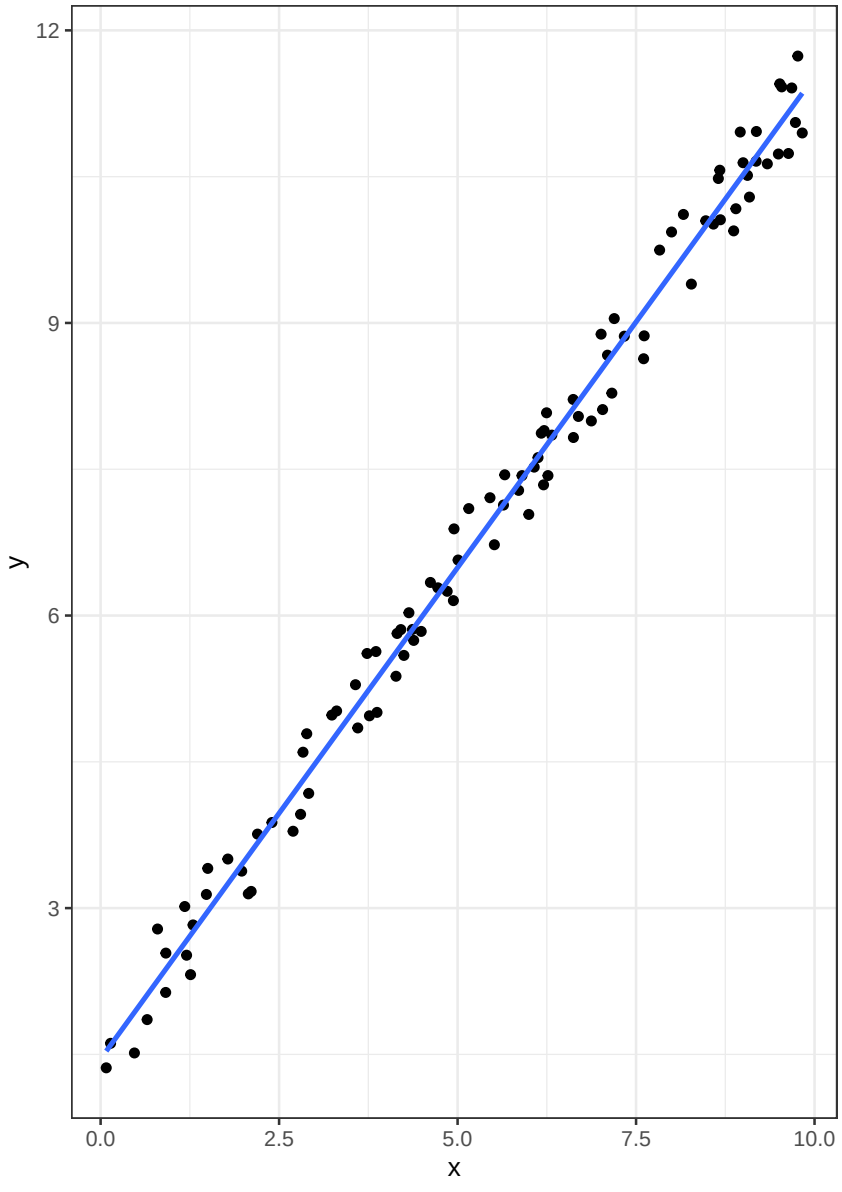


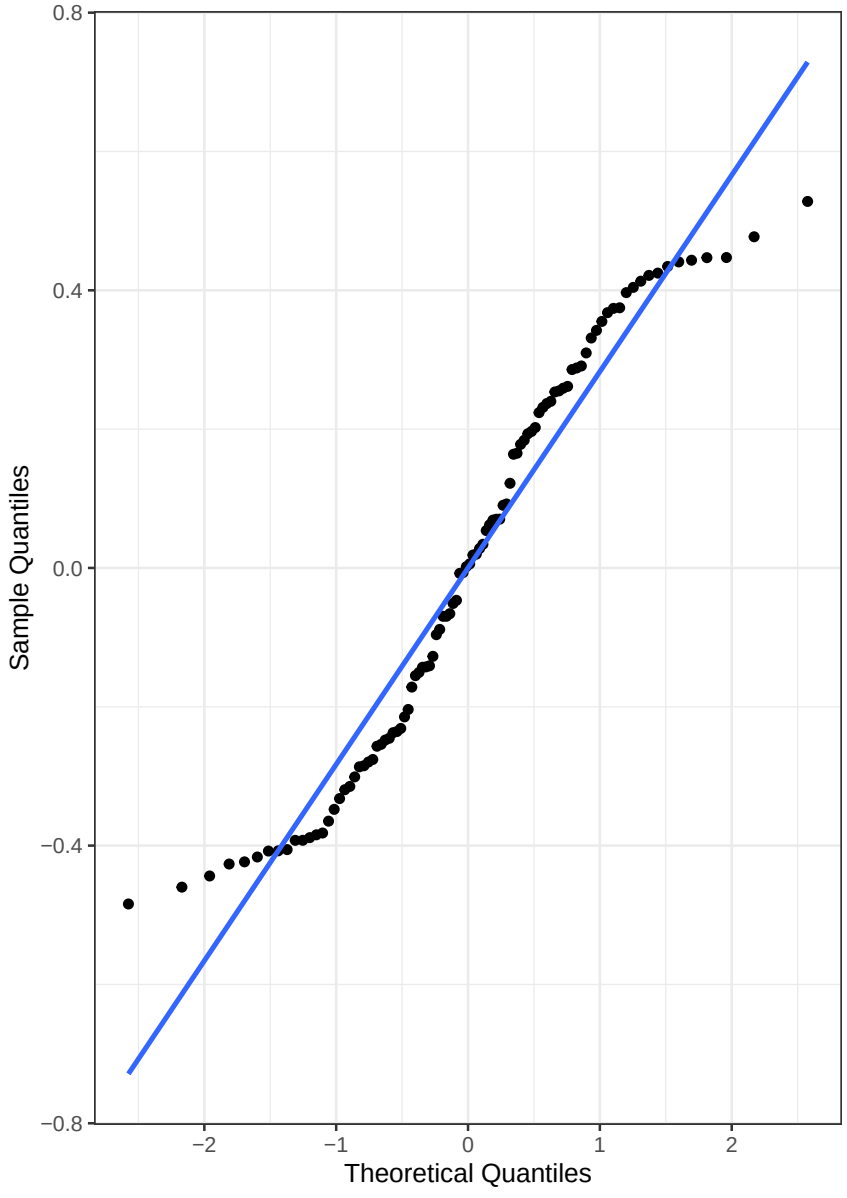
Example of unreasonable normality



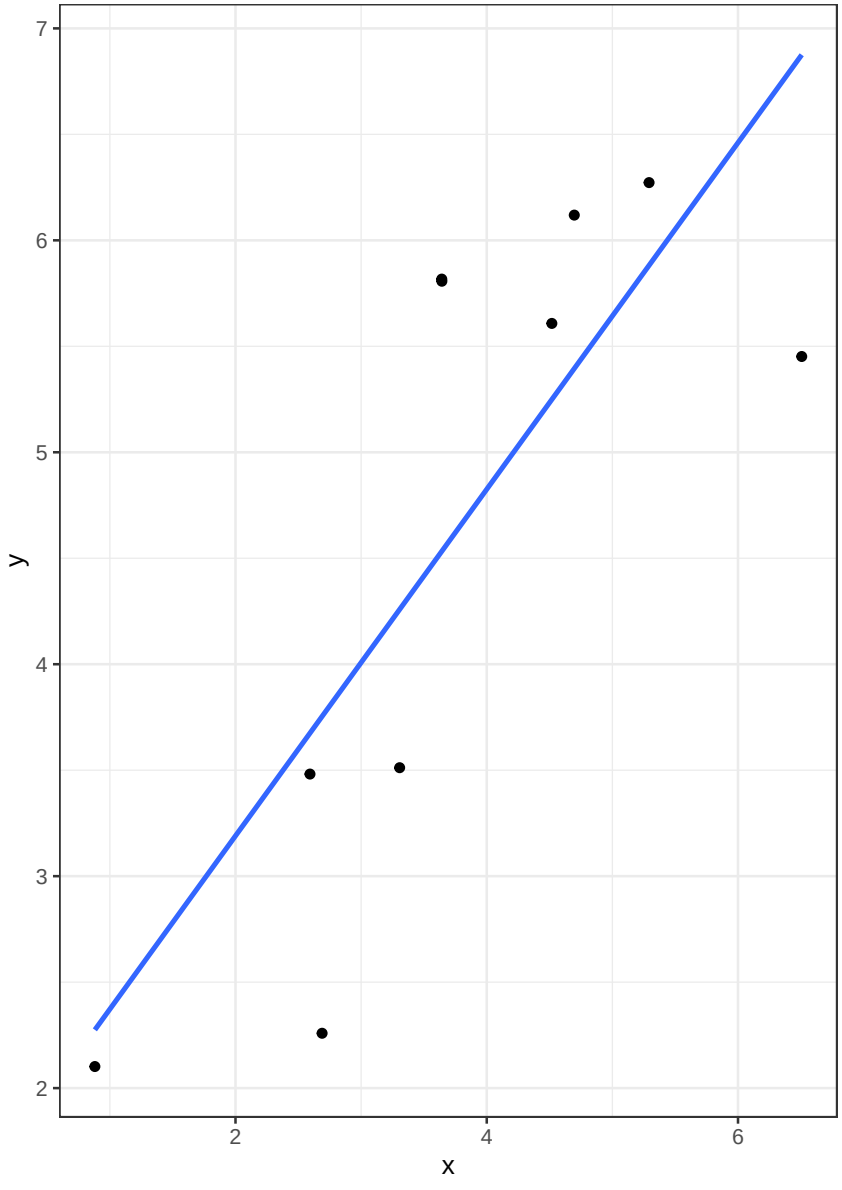


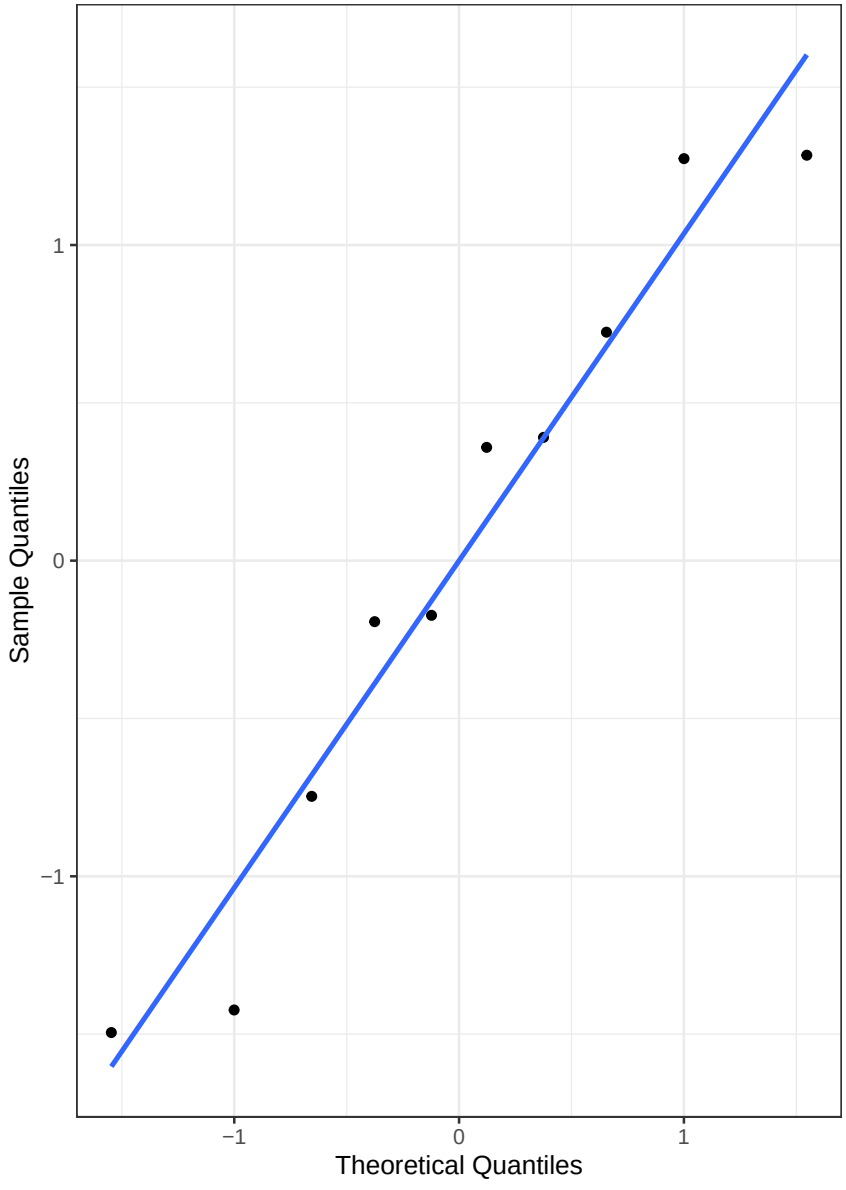
Another example of unreasonable normality





Example of small sample dataset





Assumption 4: Independence of residuals

- **Where:** We check this by looking at the experimental design.
- **Expect:** We expect to see that one observation should not give us information about the other observations.

Examples of unreasonable independence

- **Clustering:** For example, if we had closely related subjects (twins, litter of puppies).
- **Spatial relationship:** For example, two neighbouring fields.
- **Temporal relationship:** For example, manufacturing process deteriorating over time.

Summary

Linearity:

- Where: Residual versus fitted plot.
- Good: Symmetrically scattered around zero line.
- Bad: Obvious pattern.

Constant spread:

- Where: Residual versus fitted plot.
- Good: Roughly equal spread left to right (use boxes).
- Bad: Very different spread at different places.

Summary

Normality:

- Where: Normal quantile plot of residuals.
- Good: Roughly linear.
- Bad: Obvious pattern.

Independence:

- Where: Story.
- Good: Observations do not give information about others, randomness.
- Bad: Relationship between some observations.