

diagnostics for MLR

- ▶ linear relationships between x and y
- ▶ nearly normal residuals
- ▶ constant variability of residuals
- ▶ independence of residuals

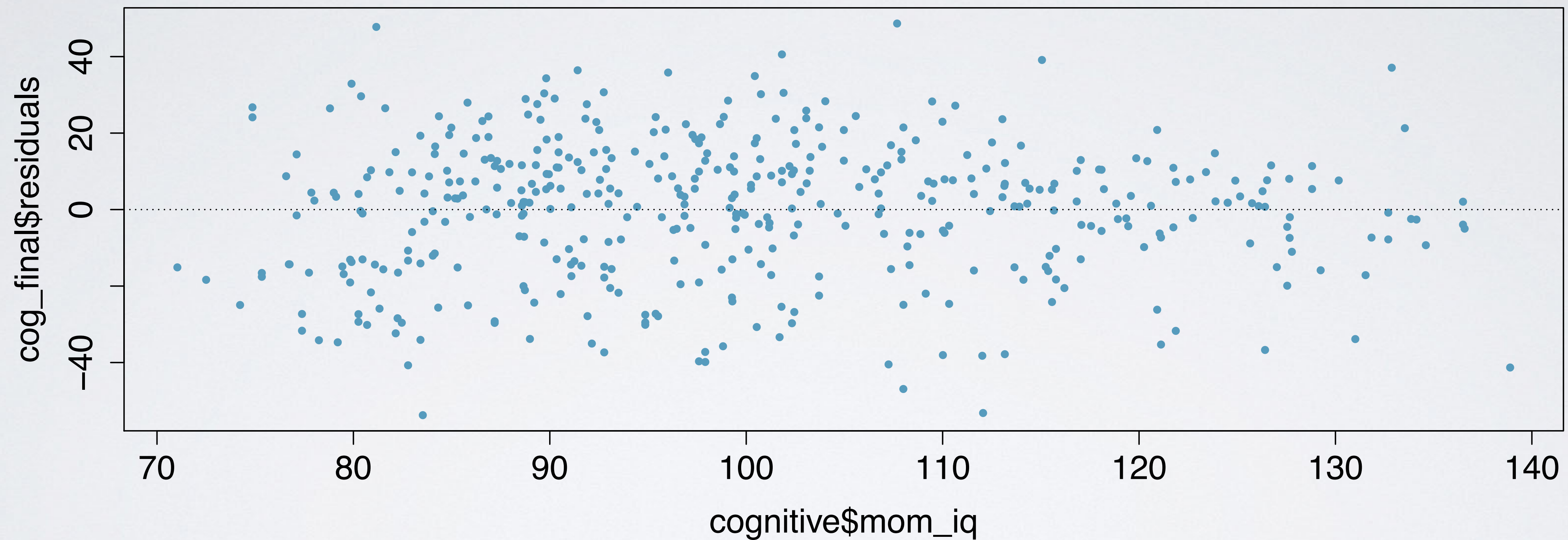
(I) linear relationships between (numerical) x and y

- ▶ each (numerical) explanatory variable linearly related to the response variable
- ▶ check using residuals plots (e vs. x)
 - ▶ looking for a random scatter around 0
 - ▶ instead of scatterplot of y vs. x : allows for considering the other variables that are also in the model, and not just the bivariate relationship between a given x and y

R

```
> cog_final = lm(kid_score ~ mom_hs + mom_iq + mom_work, data = cognitive)
> plot(cog_final$residuals ~ cognitive$mom_iq)
```

Residuals vs. mom_iq



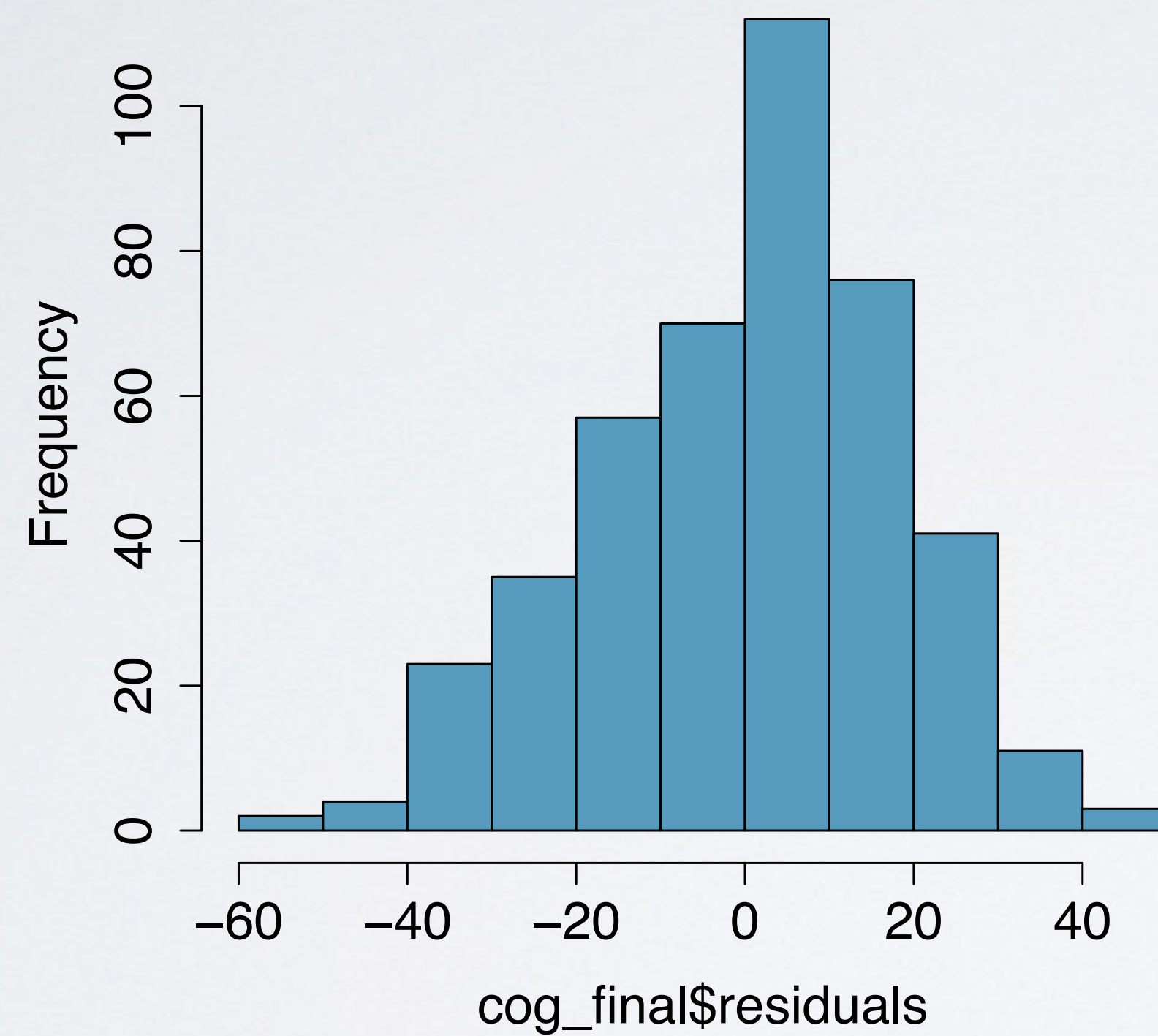
(2) nearly normal residuals with mean 0

- ▶ some residuals will be positive and some negative
- ▶ on a residuals plot we look for random scatter of residuals around 0
- ▶ this translates to a nearly normal distribution of residuals centered at 0
- ▶ check using histogram or normal probability plot

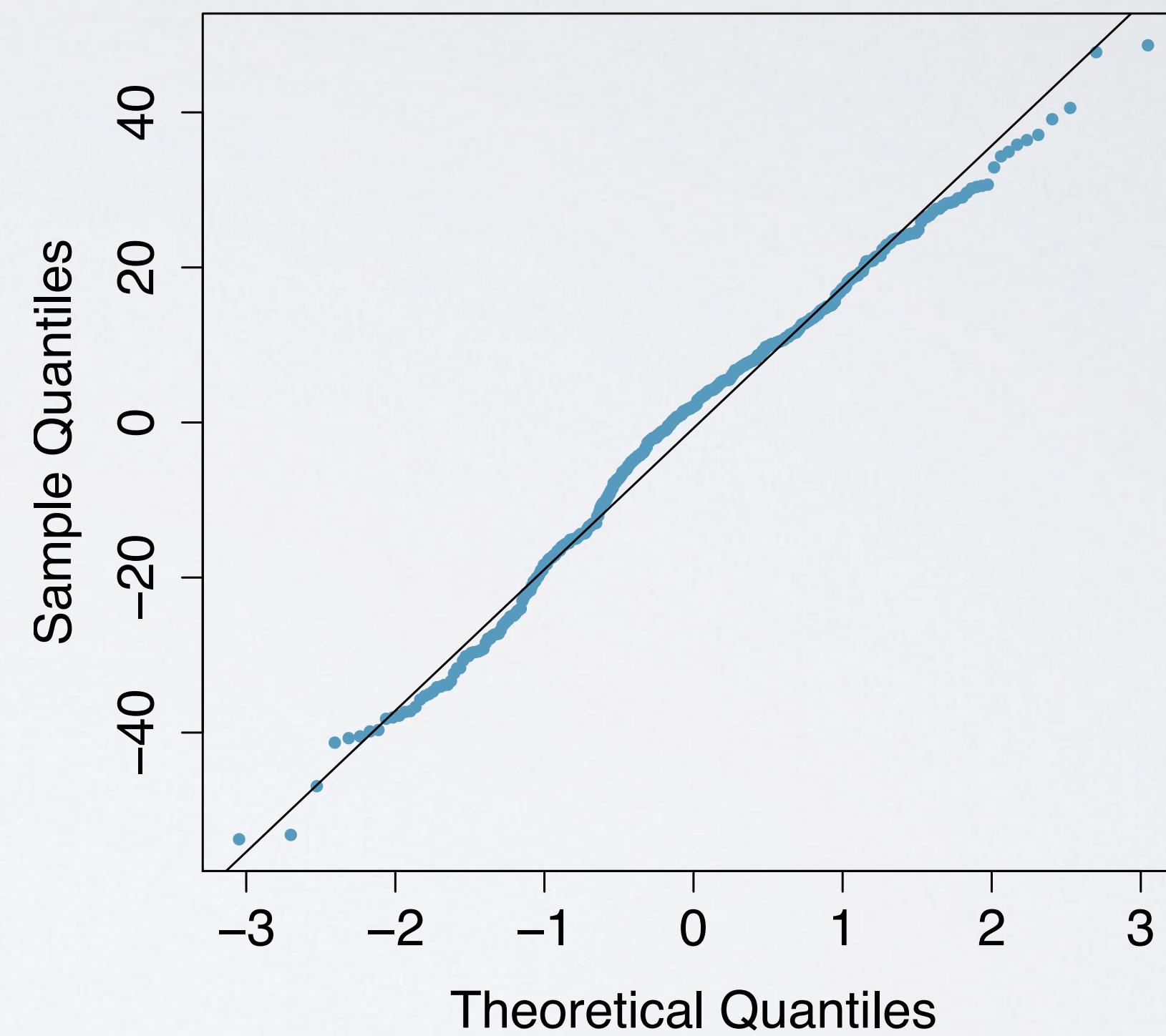
R

```
> hist(cog_final$residuals)
> qqnorm(cog_final$residuals)
> qqline(cog_final$residuals)
```

Histogram of residuals



Normal probability plot of residuals



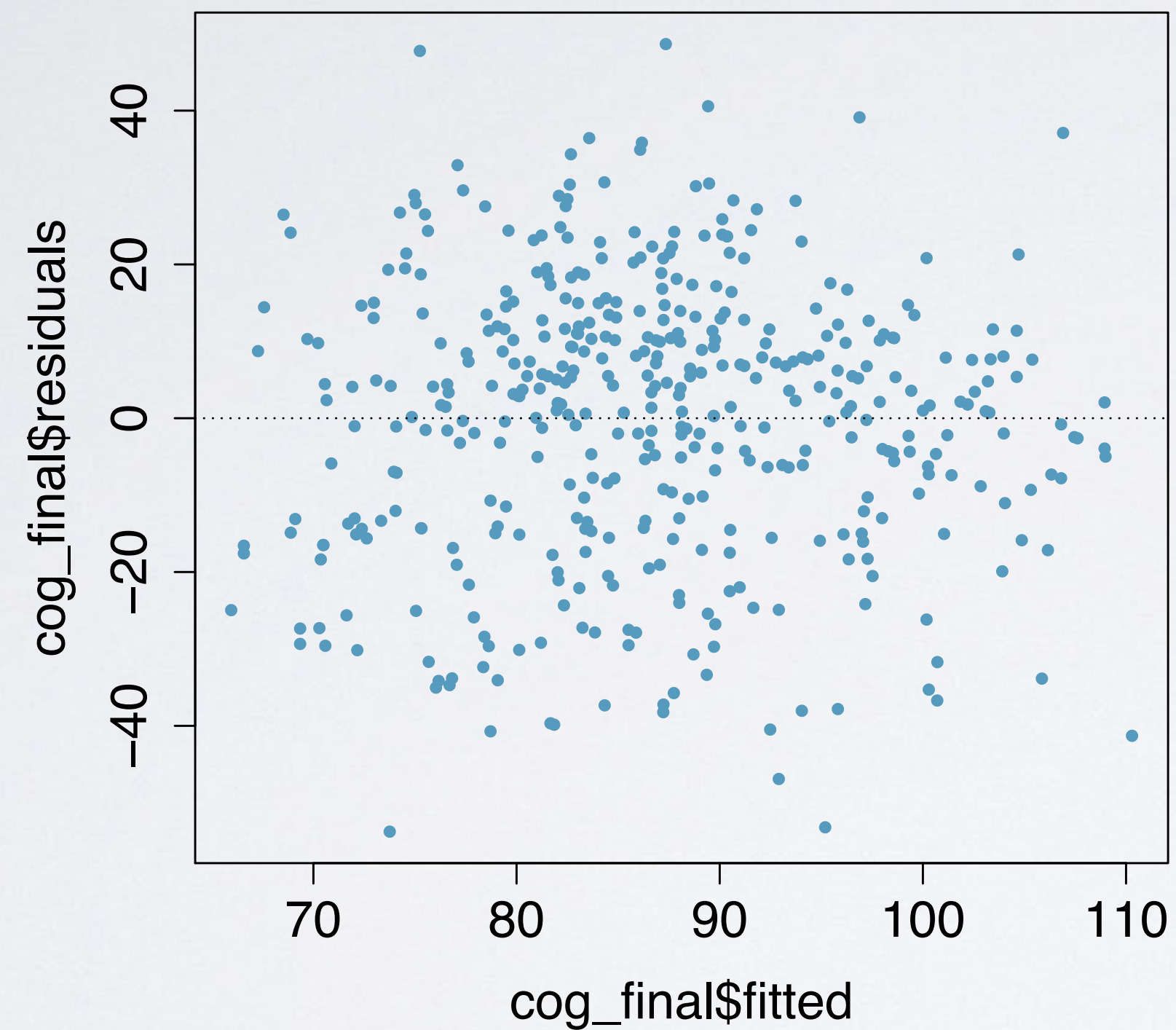
(3) constant variability of residuals

- ▶ residuals should be equally variable for low and high values of the predicted response variable
- ▶ check using residuals plots of residuals vs. predicted (e vs. \hat{y})
 - ▶ residuals vs. predicted instead of residuals vs. x because it allows for considering the entire model (with all explanatory variables) at once
 - ▶ residuals randomly scattered in a band with a constant width around 0 (no fan shape)
 - ▶ also worthwhile to view absolute value of residuals vs. predicted to identify unusual observations easily

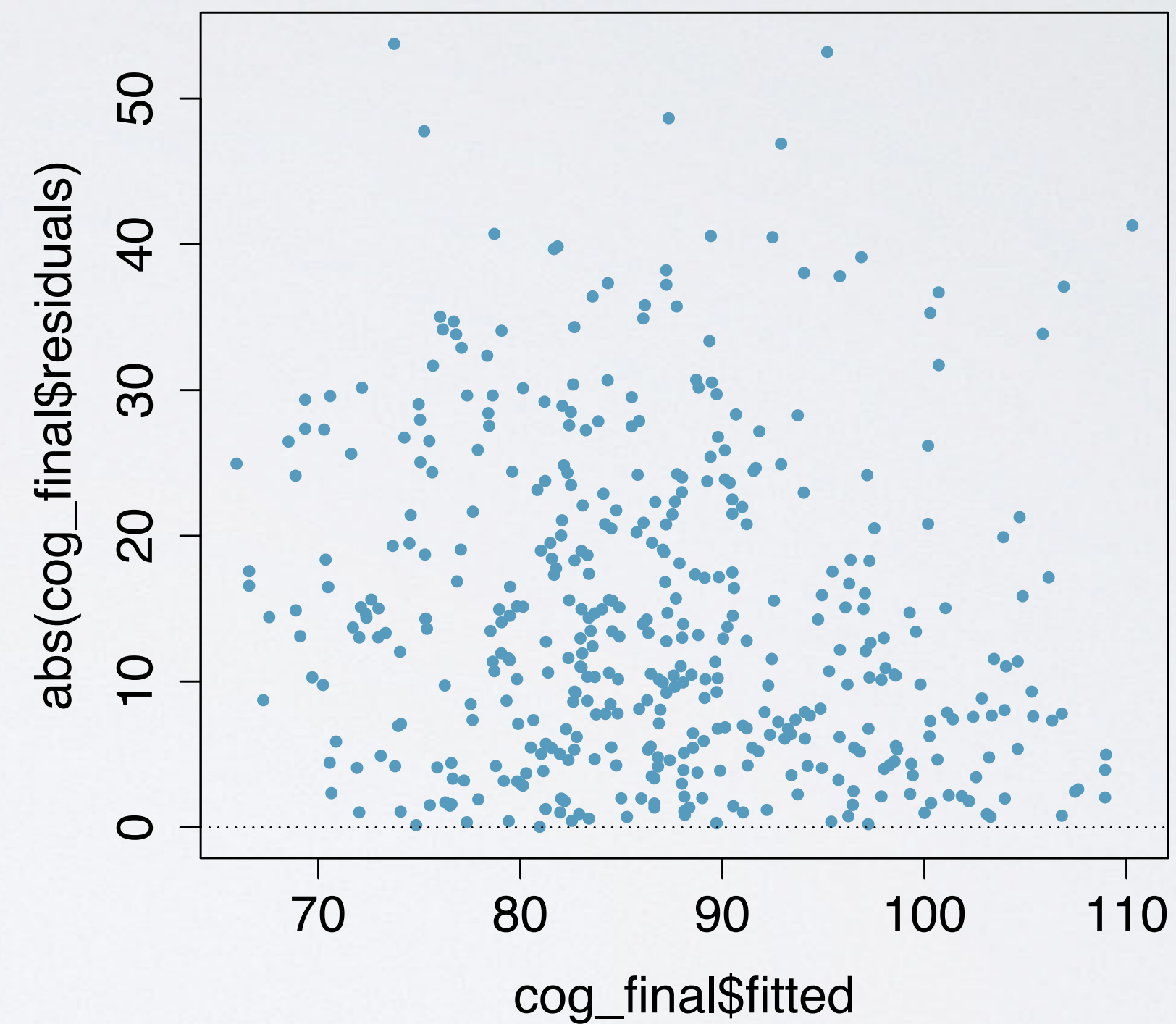
R

```
> plot(cog_final$residuals ~ cog_final$fitted)
> plot(abs(cog_final$residuals) ~ cog_final$fitted)
```

Residuals vs. fitted



Absolute value of residuals vs. fitted



(4) independent residuals

- ▶ independent residuals → independent observations
- ▶ if time series structure is suspected check using residuals vs. order of data collection
- ▶ if not, think about how the data are sampled

R

```
> plot(cog_final$residuals)
```

