Main Components in a Gaussian Linear Response Model

- Gaussian linear response models have two main components:
  - $E(Y_i)$ is a *linear* combination of observed covariates with unknown coefficients
  - The random variable Y_i, given any covariates x_i, has a Gaussian distribution with constant standard deviation σ
- This last assumption is the **homoscedasticity assumption**
- We can use graphical means to check our assumptions

When only one covariate x is involved, we can use a scatterplot

If we have more than one covariate, we can use residual plots

Residuals

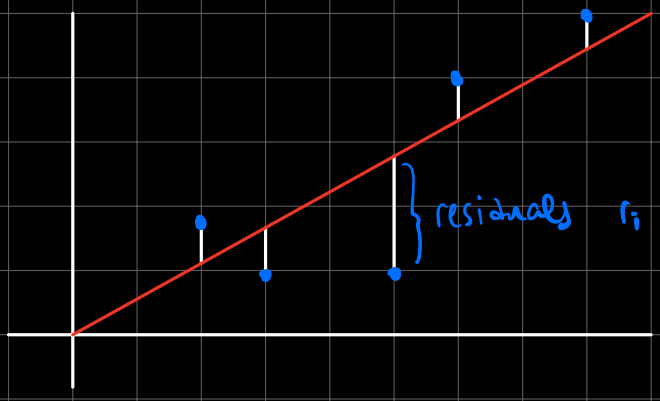
Recall:

In simple linear regression, each Yi can be expressed as

$Y_i \sim G(\mu(x_i), \sigma)$; $Y_i = \mu(x_i) + R_i$

↳ $\alpha + \beta x_i$ ↳ random variable

Actual value of R_i : $r_i = y_i - \mu(\hat{x}_i)$



Residuals should behave roughly like a random sample from a G(0,σ) distribution

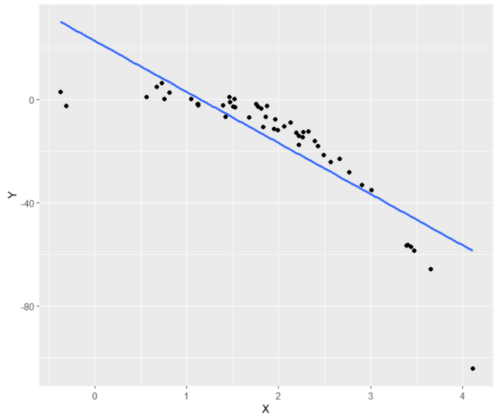
- It is common to use **standardized residuals**:

$$\hat{r}_i^* = \frac{\hat{r}_i}{s_e} = \frac{y_i - \mu(\hat{x}_i)}{s_e} = \frac{y_i - \hat{\alpha} - \hat{\beta}x_i}{s_e}$$

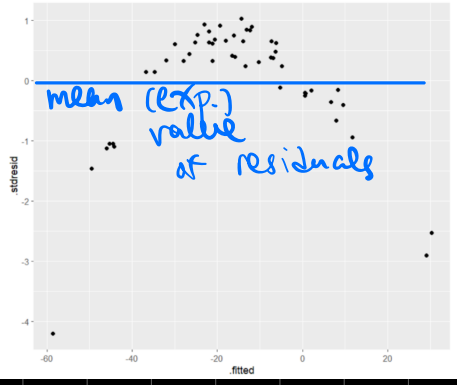
- Since the $\hat{r}_i$'s behave as a random sample from a $G(0, \sigma)$ distribution, the $\hat{r}_i^*$'s behave as a random sample from a $G(0, 1)$ distribution

- There are three types of residual plots we can consider:
  - Scatterplot of the points $(x_i, \hat{r}_i^*)$
  - Scatterplot of the points $(\mu(\hat{x}_i), \hat{r}_i^*)$
  - Gaussian Qqplot of $\hat{r}_i^*$'s

Fitted Line for Non-Linear Trend



Corresponding Residual Plot for Non-Linear Data



Y-axis: standardized residuals

Note that standardized residuals *should* follow G(0,1).

- We want an equal number of datapoints below and above the 0 line, and no discernible shapes
- Should be more or less bounded between -3 and 3 (3 standard deviations away)

X-axis — fitted values (either covariate, or linear combination of covariate)

Probably a likelihood ratio test question on the quiz