Testing hypotheses about slope:

1. H_0: $\beta = \beta_0$

(N.B. The null hypothesis is that Y cannot be predicted from X)

H_A: $\beta \neq \beta_0$

2. Test statistic is:

$$t = \frac{b - \beta_0}{SE_b}$$

Standard Error of slope:

* if assumptions of linear regression are met, the sampling distribution of *b* is a normal distribution having a mean equal to β

$$SE_b = \sqrt{\frac{MS_{residual}}{\sum (X_i - \bar{X})^2}}$$

Confidence interval for b:

$$b - t_{\alpha(2),n-2} SE_b < \beta < b + t_{\alpha(2),n-2} SE_b$$

Testing hypotheses about slope:

1. $H_0: \beta = \beta_0$

   $H_A: \beta \neq \beta_0$

2. Test statistic is:

$$t = \frac{b - \beta_0}{SE_b}$$

3. State desired significance level/Determine critical value

$$dof = n-2$$

4. Conclusion (reject or FTR) **and** confidence interval

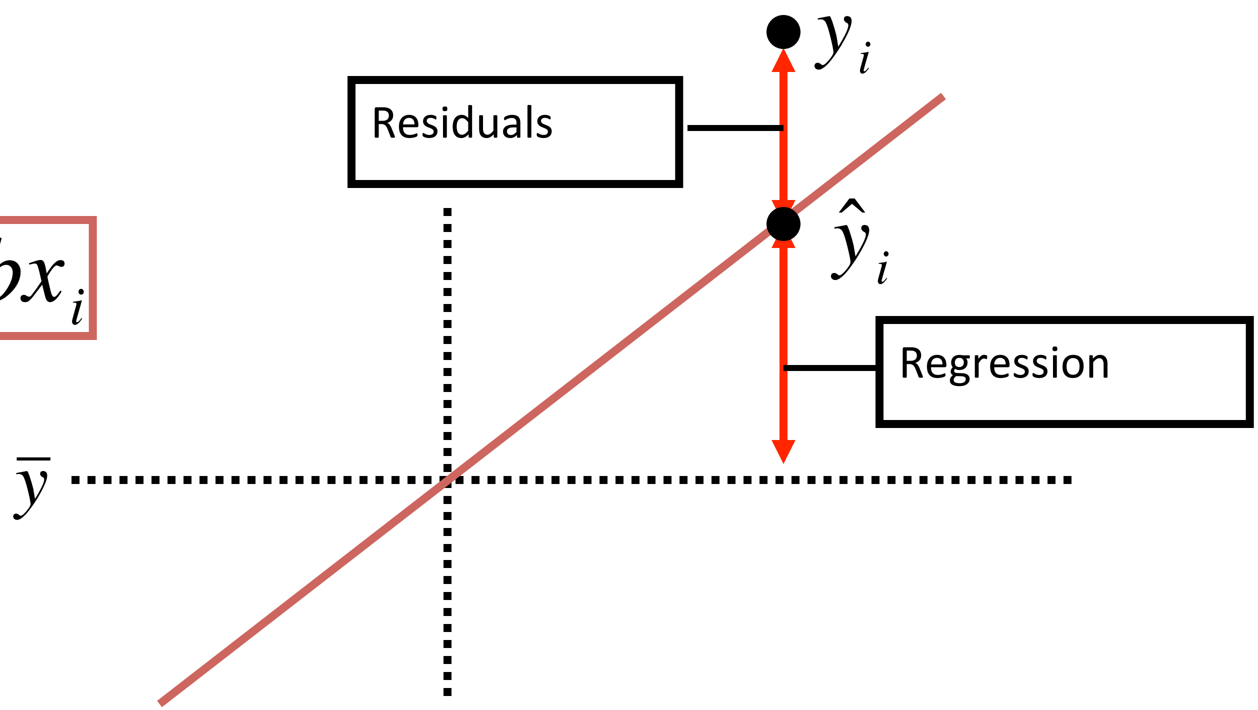
When test is two-tailed and H_0: $\beta = 0$, you can use ANOVA approach to testing regression slopes

- F-test versus t-test
- **If** H_0 is true, then the mean squares corresponding to the two components should be equal
- what are the two components?

ANOVA for testing slope:

- Residual = analogous to 'error' term
- Regression = difference in predicted value and mean value(or groups component in ANOVA)

$$\boxed{\hat{y}_i = a + bx_i}$$



Regression towards the mean:

- Francis Galton invented the term to describe the observation that tall fathers had sons of average height

- He developed "regression analysis" to study this phenomenon of "regression towards mediocrity"

- results when two variables have correlation < 1
  - Individuals who are far from the mean for one of the measurements will, on average, lie closer to the mean for the other measurement

Regression fallacy:

- Tricky concept:
  - each individual has a **true** value but the sampled value varies with time
    - the subset who scored highest on the first round included individuals who had higher values then their usual ‘true’ value
    - the second measurement captured these individuals when they happened to be closer to their own personal normal values

Regression fallacy:

- failure to consider “regression towards the mean” when interpreting the results of **observational studies**

- can be a large problem when dealing with **sick** people - they are the tail of the distribution and they might appear to improve even if the particular treatment applied has no effect

Regression fallacy:

Example: Rolling a die

Student	First	Second	Second roll lower?
1	4	5	no
2	4	3	yes
3	3	-	-
4	5	5	no
5	1	-	-
6	6	5	yes
7	5	2	yes
8	6	2	yes
9	3	-	-
10	2	-	-

Remaining students have a mean value of 5 (first roll) and 3.7 (second roll)