

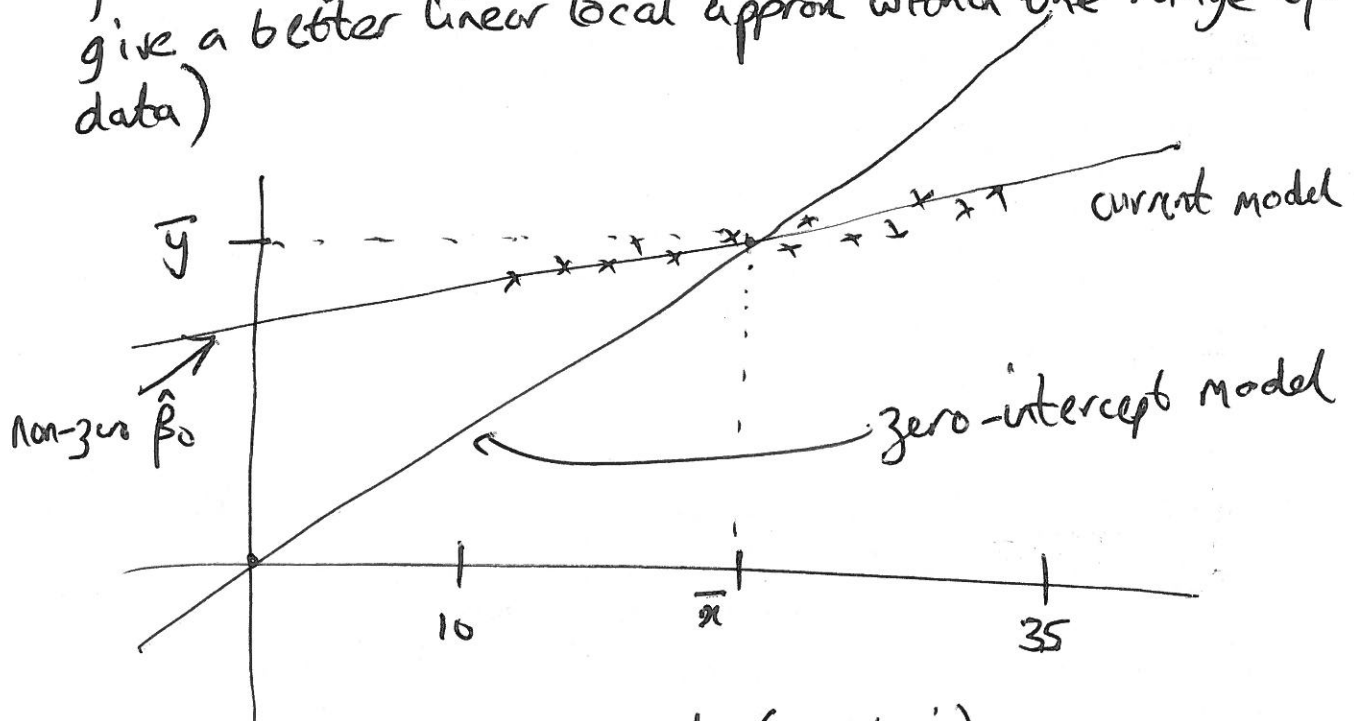
How relevant (to the research question) is the test on the intercept?

Are we really interested in whether $\beta_0 = 0$?

(well) In this example, NO, as it involves extrapolating outside the range of the data.

Conversely, when would ^{we} be interested in the value of β_0 ? When $x=0$ is within the range of the x values included in the data.

Note: Allowing a non-zero intercept gives maximum flexibility in how the model fits the data (as it will give a better linear local approx within the range of the data)

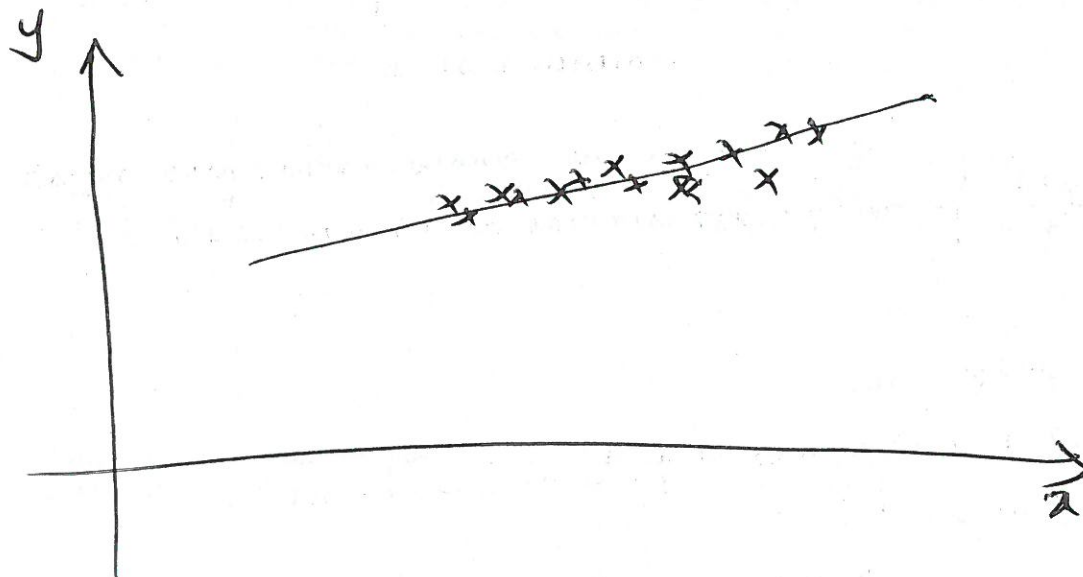


Unless there is a good (a priori) reason, we should always fit a model with an intercept (to observational data) & in general we tend to fit lower order terms as part of the model regardless of their significance

[β_1 is part of a term involving x^1 , whilst β_0 could be seen as part of a term involving x^0 , i.e. lower order]

Coefficient of Correlation

Correlation coef. of X & Y is a standardised measure of (linear) association between X & Y



Sample correlation coefficient is :

$$r = \frac{\text{cov}(x, y)}{\sqrt{\text{var}(x) \cdot \text{var}(y)}} = \frac{\frac{1}{n-1} \sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\frac{1}{n-1} \sum (x_i - \bar{x})^2 \cdot \frac{1}{n-1} \sum (y_i - \bar{y})^2}}$$

$$= \frac{S_{xy}}{\sqrt{S_{xx} \cdot S_{yy}}}$$

$-1 \leq r \leq +1$
 perfectly negatively correlated perfectly positively correlated

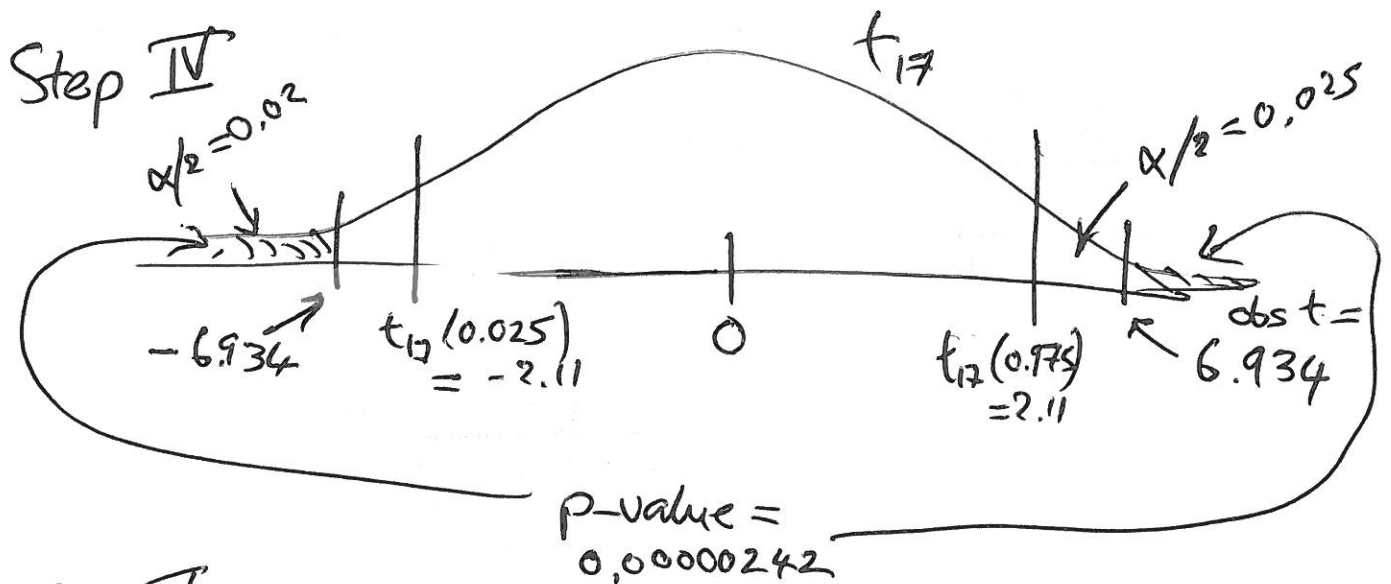
Hypothesis test on the correlation coefficient

Step I $H_0: \rho_{x,y} = 0$ vs $H_A: \rho_{x,y} \neq 0$

Step II $t = \frac{r - 0}{\text{se}(r)} = \frac{r \sqrt{n-2}}{\sqrt{1-r^2}} \sim t_{n-2}$

\searrow
 $\sqrt{\frac{1}{n-2} (1-r^2)}$

Step III $\alpha = 0.05$, reject H_0 if obs $t < -t_{n-2}^{(0.025)}$
or obs $t > t_{n-2}^{(0.975)}$



Step V As $p = 0.00000242 < \alpha = 0.05$
reject H_0 in favour of H_A & conclude $\rho \neq 0$

Note: this p-value is same as overall F test & p-value for t-test on the slope coefficient as all three tests are equivalent (proof - see page 29

of ch. 1 of the brick)

→ for SLR, the three tests all address the question "are X and Y associated (linearly related)"