

## STAT6038 week 4 lecture 11

Rui Qiu

2017-03-16

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

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

In this example, **NO**. As it involves extrapolating (well) outside the range (of the data). Regression model is always about "local data" (i.e. within the range of data).

Conversely, when would we be interested in the value of  $\beta_0$ ? 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 (i.e. it will give a better linear approximation 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) and in general we tend to fit lower order terms as part of the model regardless of their significance.

**Note:** A correlation test focus on the problem that if the two variables are associated with each other. It's nothing about the regression model itself.

### Coefficient of Correlation

The correlation between  $X$  and  $Y$  is a standardized measure of (linear) association between  $X$  and  $Y$ .

Sample correlation coefficient is:

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

$$-1 \leq r \leq 1$$

-1 means perfectly negatively correlated. 1 means 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:  $t = \frac{r-0}{se(r)} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \sim t_{n-2}$   
where  $se(\cdot)$  is  $\sqrt{\frac{1}{n-2}(1-r^2)}$
- Step III:  $\alpha = 0.05$ , reject  $H_0$  if observed  $t < -t_{n-2}(0.025)$  or observed  $t > t_{n-2}(0.975)$
- Step IV:  $t_{17}(0.025) = -2.11, t_{17}(0.975) = 2.11$ ; obs  $t = 6.934$ , obs  $-t = -6.934$ , so within the range.
- Step V: As  $p = 0.00000242 < \alpha = 0.05$ , reject  $H_0$  in favor of  $H_A$  and conclude  $\rho \neq 0$   
**Note:** this p-value is the same as overall F-test and p-value for t-test on the slope coefficients 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)?