

COEFFICIENT OF DETERMINATION

As with simple regression, in the multiple linear regression model, we can interpret

$$\frac{SYY - RSS}{SYY}$$

as the fraction of the variability in Y explained by including the terms u_1, u_2, \dots, u_{k-1} in the mean function (as compared to the constant mean function).

In the multiple regression context, $\frac{SYY - RSS}{SYY}$ is denoted as R^2 (with *capital R*).

R^2 is called the *coefficient of (multiple) determination* or (misleadingly) the *squared multiple correlation*.

Note:

- R alone (unsquared) has no meaning in *multiple* regression.
- By convention, we use small r for the *sample* correlation in *simple* regression.
- In *multiple* regression, we can talk about correlation between two variables (i.e., just two at once).
- In particular, in multiple regression, r_{ij} is often used to denote the sample correlation coefficient between terms u_i and u_j .
- R^2 is sometimes used for comparing models. *But caution is needed:*
 - It only makes sense to use for comparing models that are in the same units (e.g., submodels of the same full model).
 - A submodel of a model will always have a smaller R^2 than the larger model.
 - As discussed above and below, many other considerations should be taken to account in selecting a model.