

## 1 Is there a difference between $\hat{y}$ and $\bar{y}$ ?

I refer to  $\hat{y}_i$  as the predicted value from the fitted regression model for the  $i$ th observation.  $\bar{y}$  is the sample mean.

- 2 A list of all notations and what they represent?
- 3 What is the parameter we are trying to find in the ratio linear regression?
- 4 Why are there t-statistics and F-statistics given as outputs of a regression model? Why are we analyzing t if the F is what is usually reported for regression models?
- 5 Kind of confused about the overall goal of a regression... Is it just a faster way to get all the individual outputs (like the estimates, p-value, standard error, etc.), or does it have some extra use in and of itself?
- 6 Why is south estimate  $\log(\theta \text{ hat})$ , why isnt it  $\log(u) + \log(\theta)$
- 7 Why is the df n-p and not  $n_1+n_2-1$ ?
- 8 When calculating residual  $(y_i - \hat{y}_i)$ , is  $\hat{y}_i$  the sample mean for the entire sample, or the sample mean for North and for South? And if it's the latter, are you subtracting  $\hat{y}_i$  for North from all the North  $y$ 's and  $\hat{y}_i$  for South for all the South  $y$ 's, or subtracting both  $\hat{y}_i$  for each  $y$ ?