# AREC422 Notes

Youpei Yan

February 12, 2019

Exercise: One of the studies examining the housing price is to see if the distance to a garbage incinerator has any effect. The corresponding estimation is (with n=135, and $R^2 = 0.162$:

$$\widehat{log(price)} = 9.40 + 0.312 log(dist)$$

1. Does the sign make sense? Yes.

2. What's the elasticity of *price* with respect to *dist*? 0.312.

3. What does it mean to have a low $R^2$?
16.2% of the log(price) variation is explained by log(dist).
Also: Something in the error terms that (1) are not included in the model and (2) have impacts to price. It may includes the size of the house, number of bathrooms, age of the house, etc.

4. What situation may lead to the violation of the "zero conditional mean" assumption?
All the above 3 missing (omitted) variables, although they have impacts to the housing price, they are unlikely to be correlated with $log(dist)$.
However, say, neighborhood quality, such as number of schools nearby, area of forest covers in the region, etc. can also influence price. They are in the error term and they are likely to be correlated with $log(dist)$.
Because $corr(neighborhood\_quality, log(dist)) > 0$ violates $E(u|x) = 0$, we say that $\hat{\beta}_1$'s estimation is wrong, and $\hat{\beta}_1$ is biased ($E(\hat{\beta}_1) \neq \beta_1$).

How can we guarantee that the estimate $\hat{\beta}_1$ is not biased? We need Gauss Markov Assumptions (GMA).

## 1 Gauss Markov Assumptions

(1) "Linear" Model.
(2) Random Sampling.
(3) $x_i$ are not identical.
(4) Zero conditional mean.
*(5) Constant variance: $var(u|x) = \sigma^2$
We need the GMA1-4 to guarantee the unbiasedness. The 5th one is called homoscedasticity. With it, we can guarantee that the estimated variance of residuals $E(\sigma^2)$ is also unbiased: $E(\sigma^2) = \sigma^2$.

Proof of $E(\hat{\beta}_0) = \beta_0$ and $E(\hat{\beta}_1) = \beta_1$.:

$$E(\hat{\beta}_1) = E\left[\frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right] = E\left[\frac{\sum_{i=1}^{n}(x_i - \bar{x})y_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right]$$

This is because $\sum_{i=1}^n (x_i - \bar{x})\bar{y} = \bar{y}\sum_{i=1}^n (x_i - \bar{x}) = \bar{y} \cdot 0 = 0$. Then:

$$E(\hat{\beta}_1) = E\left[\frac{\sum_{i=1}^n (x_i - \bar{x})^2)(\beta_0 + \beta_1 x_i + u_i)}{SST_x}\right] = E\left(\beta_1 + \frac{\sum_{i=1}^n (x_i - \bar{x})u_i}{SST_x}\right)$$

Because $\beta_0 \sum_{i=1}^n (x_i - \bar{x}) = \beta_0 \cdot 0 = 0$ and $\beta_1 \sum_{i=1}^n (x_i - \bar{x})x_i = \beta_1 \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x}) = \beta_1 SST_x$
Recall that $E(u_i) = 0$, then $E(\sum_{i=1}^n (x_i - \bar{x}x_i)) = 0$, and we have:

$$E(\hat{\beta}_1) = \beta_1$$

Similarly, $E(\hat{\beta}_0) = E(\bar{y} - \hat{\beta}_1\bar{x}) = E(\beta_0 + \beta_1\bar{x} + \bar{u} - \hat{\beta}_1\bar{x}) = \beta_0$ We have proved that both the estimates are unbiased, i.e., their expected values are the same as the true values.