Review of Estimation

- ***Estimating the population mean***
- ***Random sampling, estimators and estimates***
- ***Linear and unbiased estimators (LUEs)***
- ***Variance of LUEs***
- ***Best Linear Unbiased Estimators (BLUE)***
- ***… I: Introduction and some examples***
- ***… II: The optimization problem***
- ***… III: Wrapup***
- ***More generally: Estimators, unbiasedness, efficiency and interval estimators***
- ***Some common estimators: Sample statistics as estimators***

This review is built around what is probably the best known estimation example, estimating the mean of a distribution with randomly sampled data drawn from that unknown distribution. We will first walk through the example, and then address more general estimation issues/topics.

We start with a well-known estimation problem.

Estimating the population mean

1. Consider some population with a distribution of, say, heights, characterized by the random variable Y. The mean, μ, and variance, σ^2, of Y are unknown. You want to estimate μ, the average height in the population.

Random sampling, estimators and estimates

2. Sample independently n times from this population and use the data to estimate the unknown mean.
3. Before the data are observed (ex ante), each observation is a random variable with a value yet to be determined. After the data are observed, observations are, well, data points… sample outcomes to be used to generate estimates.
4. ***Estimators, ex ante***: Each of the n independent random draws from the distribution Y, is itself a random variable, Y_i.[1] Given the nature of the sampling process, the Y_i's are iid (independently and identically distributed) with distribution Y.
   a. A ***point estimator*** of μ will be some function of the observed values of the Y_i's, and will accordingly, be a random variable, *ex ante*. Accordingly, you can think of estimators as *rules*, which assign different estimates to different drawn samples.

---

[1] Recall that we use upper case letter to denote random variables and lower case letters to denote the actual sample values.

5. ***Estimates, ex post***: After we have the sample data $\{y_1, y_2, ..., y_n\}$, the actual ***estimate*** will be the value of the point estimator for the given set of drawn values.

6. ***Example: The Sample Mean:*** Probably the best known point estimator is the sample mean, in which the rule is that you simply average the sample observations.

   a. *ex ante*, the sample mean estimator is a random variable, which can take on different values with different probabilities, reflecting the random nature of the sampling process: $M(Y_1, Y_2, ..., Y_n) = \overline{Y} = \frac{1}{n}\sum Y_i$.

   b. *ex post*, and after the sample data are drawn, the sample mean estimator provides us with a point estimate of the unknown mean of the distribution: $M(y_1, y_2, ..., y_n) = \overline{y} = \frac{1}{n}\sum y_i$.

Linear and unbiased estimators (LUEs)

7. We often look first at linear estimators… because they are relatively simple to work with, and often a useful approximation to more complicated functional forms.

8. ***Linear***: Given the randomly sampled data, linear estimators (remember, they are random variables) will have a general linear functional form:

$$M = \beta_0 + \beta_1 Y_1 + \beta_2 Y_2 + ... + \beta_n Y_n .$$

9. ***Unbiased***: This linear estimator will be unbiased if $E(M) = \mu$, which is to say, the expected value of the estimator is the true mean, μ.

   a. Even though we don't know the true mean μ, we can often determine whether or not an estimator of μ is in fact biased or not.

10. ***LUEs***: If an estimator is linear and unbiased, we call it a Linear Unbiased Estimator (LUE). *Surprise!*

11. ***ex post***: For the actual drawn sample $\{y_1, y_2, ..., y_n\}$, the particular linear estimate m will be $m = \beta_0 + \beta_1 y_1 + \beta_2 y_2 + ... + \beta_n y_n$.

    a. m may or may not be close to μ … but on average, estimates generated in this fashion will equal μ if M is an unbiased estimator of μ.

12. Since the Y_i's are all iid with distribution Y, each Y_i has mean μ, and so the expected value of M is:

$$E(M) = \beta_0 + \beta_1\mu + \beta_2\mu + ... + \beta_n\mu = \beta_0 + \sum \beta_i\mu = \beta_0 + \mu\sum \beta_i .$$

13. Since we are focused on linear unbiased estimators of μ ... we only want to consider $\{\beta_i\}$ that satisfy $\beta_0 + \mu\sum \beta_i = \mu$, irrespective of what the particular value of μ happens to be.

14. Since $\beta_0 + \mu\sum\beta_i = \mu$ for all μ, if you differentiate both sides with respect to μ**,** you find that for the linear estimator to always be unbiased, we require that

$$\beta_0 = 0 \text{ and } \sum_{i=1}^{n}\beta_i = 1$$

(the intercept term must be 0 and the slope coefficients must sum to 1; you can think of the slope coefficients as weights since they sum to one… but they do not have to be non-negative).

15. So if we restrict attention to the set (or class) of linear unbiased estimators of μ, we are considering only estimators of the form:

$$M = \beta_1 Y_1 + \beta_2 Y_2 + ... + \beta_n Y_n \text{, where } \sum_{i=1}^{n}\beta_i = 1 .$$

16. This defines the class of *LUE*s (*Linear Unbiased Estimators*) for our estimation problem.

Variance of LUEs

17. Consider the linear unbiased estimator M just defined. Since the Y_i's are pairwise independent, $Cov(Y_i, Y_j) = 0 \ \ for \ i \neq j$, and the variance of the sum is the sum of the variances. And so, $Var(M) = \beta_1^2 Var(Y_1) + \beta_2^2 Var(Y_2) + ... + \beta_n^2 Var(Y_n)$.

18. And since $Var(Y_i) = \sigma^2$ for each i,

$$Var(M) = \sigma^2 \sum \beta_i^{\,2}$$ … which will vary, depending on the β_i's.

Best Linear Unbiased Estimators (BLUE)

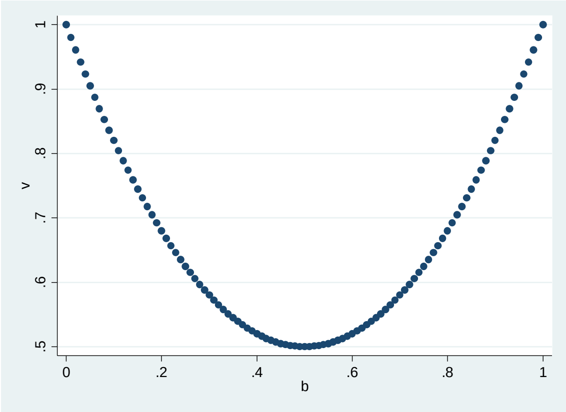
19. **BLUE I: Introduction and some examples**
    a. The Best Linear Unbiased Estimator will be the estimator in the class of LUEs that has minimum variance.
    b. So we want to consider all linear estimators of the form

    $$M = \beta_1 Y_1 + \beta_2 Y_2 + ... + \beta_n Y_n \text{, where } \sum_{i=1}^{n}\beta_i = 1 ,$$

    and find the particular set of $\{\beta_i\}$ that minimizes the variance within this group/class of estimators.
    c. Here are some unbiased estimators… unbiased, since the weights sum to one.

    Which one do you prefer?

i. $M_1 = Y_1$: $E(M_1) = \mu$ and $Var(M_1) = \sigma^2$



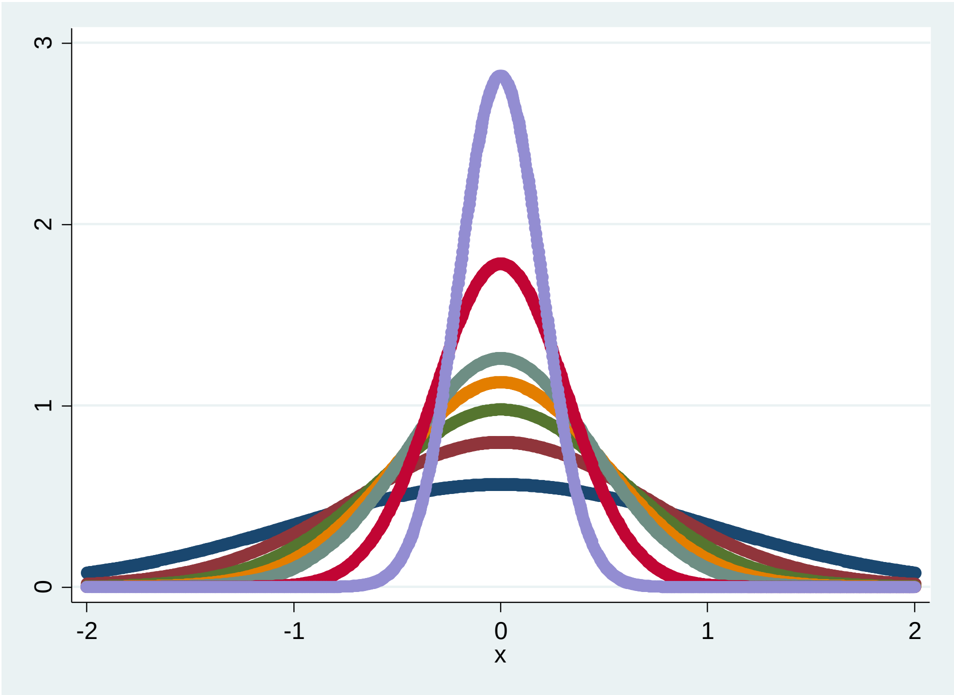
ii. $M_{2.1} = .5Y_1 + .5Y_2$: $E(M_{2.1}) = \mu$ and $Var(M_{2.1}) = .5^2\sigma^2 + .5^2\sigma^2 = \frac{2}{4}\sigma^2 = \sigma^2/2$

iii. $M_{2.2} = \beta Y_1 + (1-\beta)Y_2$ (Figure left): $E(M_{2.2}) = \mu$ and $Var(M_{2.2}) = \beta^2\sigma^2 + (1-\beta)^2\sigma^2 = \sigma^2\left\{\beta^2 + (1-\beta)^2\right\}$

1. So to minimize $Var(M_{2.2})$, set $\beta = .5$.

iv. $M_{3.1} = (1/3)Y_1 + (1/3)Y_2 + (1/3)Y_3$: $E(M_{3.1}) = \mu$ and $Var(M_{3.1}) = (1/3)^2\sigma^2 + (1/3)^2\sigma^2 + (1/3)^2\sigma^2 = \frac{3}{9}\sigma^2 = \sigma^2/3$

v. And so forth…. Here's what the distributions of the equally weighted sample means (with different sample sizes… so $M_1, M_{2.1}, M_{3.1}$...) look like…. Assuming $Y \sim N(0,1)$:

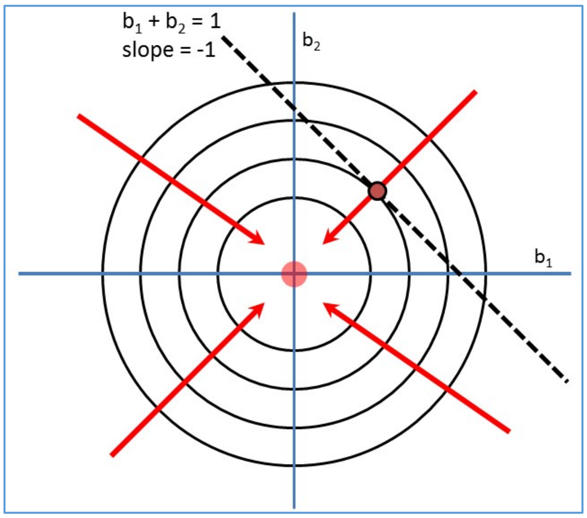


As the sample size increases, the distribution of the sample mean becomes more and more tightly concentrated around the true unknown mean, μ .

20. **BLUE II: The optimization problem**

a. To find the BLUE estimator of the unknown mean μ, we need to solve the following optimization problem:

$$\min\ Var(M) = \sigma^2 \sum \beta_i{}^2 \text{ subject to } \sum_{i=1}^{n} \beta_i = 1.$$



b. This well-studied optimization problem is called a *Quadratic Programming* (QP) problem… and features a quadratic objective function, $Var(M) = \sigma^2 \sum \beta_i{}^2$, and a linear constraint, $\sum_{i=1}^{n} \beta_i = 1$. As shown in the diagram for the two-dimensional case, the level curves of the objective function are concentric circles centered around the origin. So the goal is to be on the concentric circle closest to the origin, while also satisfying the linear constraint that $\sum_{i=1}^{n} \beta_i = 1$. As you can see, the solution will typically occur at a point of tangency between the smallest achievable concentric circle and the linear constraint.

c. Not surprisingly, there are many ways to solve this QP problem… probably the easiest is to just incorporate the *constraint* into the *objective function*.

d. The constraint requires that $\sum_{i=1}^{n} \beta_i = 1$, or put differently, that $\beta_n = 1 - \sum_{i=1}^{n-1} \beta_i$ (note that the summation runs from 1 to n-1. If we incorporate the constraint in the objective function, then we have a new (unconstrained) optimization problem:

$$\min\ Var(M) = \sigma^2 \sum_{i=1}^{n-1} \beta_i^2 + \sigma^2 \beta_n^2 = \sigma^2 \left\{ \sum_{i=1}^{n-1} \beta_i^2 + \left[1 - \sum_{i=1}^{n-1} \beta_i \right]^2 \right\}.$$

e. We can solve this with n-1 FOCs (First Order Conditions):

$$\text{For } j = 1, \dots, n-1\text{: } \min \frac{\partial}{\partial \beta_j} Var(M) = \sigma^2 \left\{ 2\beta_j + 2\left[1 - \sum_{i=1}^{n-1} \beta_i \right](-1) \right\} = 0, \text{ or}$$

$$\beta_j^* = \left[1 - \sum_{i=1}^{n-1} \beta_i^* \right].$$

f. Since the RHS of the last expression doesn't depend on j, the β_j^*'s are all the same. Call their common value β^*.

g. The from the last expression we have: $\beta^* = \left[1 - \sum_{i=1}^{n-1} \beta^*\right] = 1 - (n-1)\beta^*$, or $n\beta^* = 1$ or

$$\beta_i^* = \frac{1}{n} \;\; for\ i = 1, ..., n-1.$$ And since $\beta_n^* = 1 - \sum_{i=1}^{n-1} \beta_i^*$, $\beta_n^* = \frac{1}{n}$ as well.

h. But then M is just the Sample Mean: $M = \bar{Y} = \frac{1}{n}\sum Y_i$.

i. Or put differently, ***The Sample Mean is BLUE!***

21. **BLUE III: Wrapup**

a. **BLUE:** Since the Sample Mean is unbiased and has minimum variance in the class of LUE's, it is a BLUE… the best (minimum variance) estimator in the class of linear unbiased estimators (LUE's).

   i. Notice that this result holds, even if we don't know the actual value of σ^2.

   ii. Also notice that this holds for any distribution (we haven't yet said anything about the particular distribution of Y).

b. **Your sample:** For a particular sample $\{y_1, y_2, ..., y_n\}$, $m = \bar{y} = \frac{1}{n}\sum y_i$ may or may not be close to μ … but on average, estimates generated in this fashion will equal μ**,** since $M = \bar{Y}$ is an unbiased estimator of μ**.**

   i. The Sample Mean estimator, $M = \bar{Y} = \frac{1}{n}\sum Y_i$, is a random variable, taking on different values with different probabilities depending on the actual drawn sample. The distribution of M is called a sampling distribution.

c. **Review**:

   i. (ex ante) Estimators are random variables, taking on different values depending on the drawn sample.

   ii. (ex post) Estimates are numbers, the value of the estimator for our particular drawn sample (set of observations).

   iii. We bless estimates not because we know them to be specifically praiseworthy, but rather because we praise the process/rule/estimator that generated the estimate. Or put differently: We can say something about the quality of estimator… but we don't have much to say about the quality of specific estimates, unless of course, we know something about the representativeness of our sample.

Review of Estimation

As promised, we now turn to more general estimation topics.

More generally: Estimators, unbiasedness, efficiency and interval estimators

22. **Estimators**:

   a. Suppose that you have an iid random sample $\{Y_1, Y_2, \ldots Y_n\}$ from the distribution, Y, and you want to use this data to estimate some unknown parameter of the distribution, θ. A point estimator will be a function of the Y_i's, say $W = h(Y_1, Y_2, \ldots Y_n)$, and will itself be a random variable (taking on different values with different probabilities).

23. **Unbiasedness**:

   a. An estimator is unbiased if, on average, it's right… or more formally, if it's expected value is the true value of the parameter. $E(W) = \theta$.

      i. We've shown above that the Sample Mean is an unbiased estimator of the true mean, μ.

      ii. The bias of the estimator is the difference between its expectation and the true value of the unknown parameter: $Bias(W) = E(W) - \theta$.

24. Efficiency of (unbiased) estimators

   a. Consider two unbiased estimators of θ, W_1 and W_2. Then W_1 is more *efficient* than W_2 if for every parameter value θ, $Var(W_1) \le Var(W_2)$, and for at least one value of θ, $Var(W_1) < Var(W_2)$ (so W_1 never has higher variance than W_2).

   b. Why restrict to unbiased estimators? Otherwise, it's easy to find (really bad) estimators with zero variance. For example, spoze that my estimator of θ is $W_1 = 7$. My crummy stupid estimator of course has nothing to do with θ, but it does have zero variance!

25. Confidence intervals as interval estimators

   a. So far we have focused on point estimators, which provide a single (point) estimate of the unknown parameter. Alternatively, we could work with interval estimators, for which estimates are intervals (a range of values) rather than points (specific values). The most common interval estimator is the *Confidence Interval*:

   b. Consider the point estimator $L = l(Y_1, Y_2, \ldots Y_n)$ for the lower bound of the confidence interval and $U = u(Y_1, Y_2, \ldots Y_n)$ for its upper bound. L and U are both random variables, taking on different values depending on the drawn sample.

      i. The randomly generated confidence interval [L,U] will be an interval estimator, where L and U are the values of the interval endpoints.

      ii. Note that if it's the case that 95% of the time, intervals generated in this fashion contain the true mean μ, then we say that we have a 95% *Confidence Interval* (estimator) for μ.

Review of Estimation

Some common estimators: Sample statistics as estimators

26. ***Mean***: The ***sample mean***, $\bar{Y} = \frac{1}{n}\sum Y_i$, is an unbiased estimator (in fact, it's BLUE) of the mean of Y, μ . $E(\bar{Y}) = E(Y) = \mu$.

    a. The particular estimated sample mean is $\bar{y} = \frac{1}{n}\sum y_i$.

27. ***Variance***: The ***sample variance***, $S_{YY} = S_Y^2 = \frac{1}{n-1}\sum (Y_i - \bar{Y})^2$, is an unbiased estimator of the variance of Y, σ^2, when the mean of Y is unknown: $E(S_{YY}) = Var(Y) = \sigma^2$.[2]

    a. The particular estimated sample variance is $S_{yy} = S_y^2 = \frac{1}{n-1}\sum (y_i - \bar{y})^2$.

28. ***Standard deviation***: The **sample standard deviation** is the square root of the sample variance, $S_Y = \sqrt{S_{YY}} = \sqrt{\frac{1}{n-1}\sum (Y_i - \bar{Y})^2}$. It is generally a biased estimator of the Standard Deviation of Y, σ_Y, since in general, the expected value of the square root of something is not equal to the square root of the expected value of that something. … But that fact doesn't stop us from using it!

    a. The particular estimated standard deviation is $S_y = \sqrt{\frac{1}{n-1}\sum (y_i - \bar{y})^2}$.

29. ***Covariance***: The ***sample covariance***, $S_{XY} = \frac{1}{n-1}\sum (X_i - \bar{X})(Y_i - \bar{Y})$, is an unbiased estimator of the Covariance of X and Y, $E(S_{XY}) = Cov(X,Y) = \sigma_{XY}$, when the means of X and Y (μ_X *and* μ_Y) are unknown.

    a. The particular estimated sample covariance is $S_{xy} = \frac{1}{n-1}\sum (x_i - \bar{x})(y_i - \bar{y})$.

30. ***Correlation***: The ***sample correlation*** estimator, $\rho_{XY} = \frac{S_{XY}}{S_X S_Y}$, is generally a biased estimator of the correlation of X and Y, $corr(X,Y) = \rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$.

    a. The particular estimated sample correlation is $\rho_{xy} = \frac{S_{xy}}{S_x S_y}$.

---

[2] We divide by n-1 to generate an unbiased estimator. There are circumstances under which you might want to divide by n, or even n+1 … but with large samples, the consequential differences are typically quite small.