Lecture 3

Chapters R.5-R.8

R.5 Sampling

Population – a group of items of events we would like to know about.

Parameter – a value that describes that population. The parameter of interest is the parameter that the researcher seeks to learn about.

Sample – a survey of a subset of the population.

Exercise 1: Come up with an example of a population, a parameter from that population, and a sample of that population.

Population – U of O students

Parameter – GPA

Sample – students taking EC 320

Exercise 2: Suppose we want to learn about the athleticism of college students. If we only survey football wide receivers, would we have a sample the represents the average colleges student? Why or why not?

No, wide receivers tend to be very athletic compared to the average person

Exercise 3: Suppose instead we decide to survey every single college student in the United States. Does this sound reasonable? Why or why not?

Not feasible. There are hundres of thousands, if not millions of college students. This would take a long time and be very costly.

Exercise 4: How could you get a sample that represents the population of college students without surveying every college student?

randomly survey a few students at multiple colleges

Simple Random Sampling – each observation is selected at random from the entire set of possible observations.

Estimator – a rule (or formula) for estimating an unknown population parameter given a sample of data.

Note: Each observation of the sample is a random variable. The estimator is a combination of random variables, so it is also a random variable.

Estimate – the specific number obtained from the estimator.

Sample mean – the estimator for population mean.

Formula – Sample mean

n = number of observations

$$\bar{X} = \frac{1}{n}(X_1 + X_2 + \ldots + X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Exercise 5: Suppose we take a survey of individuals to find out what the average number of pets owned in Eugene is. Given the sample data below, what is the estimate?

Individual	Number of Pets
1	2
2	1
3	0
4	5
5	2

$$\bar{X} = \frac{1}{5}(2+1+0+5+2) = \frac{1}{5}(10) = \frac{10}{2}$$

$$\boxed{\bar{X} = 2}$$

Exercise 6: suppose X has a distribution with variance σ_X^2. What is the variance of the sample mean?

$$\sigma_{\bar{X}}^2 = Var\left\{\frac{1}{n}(X_1 + X_2 + \ldots + X_n)\right\}$$

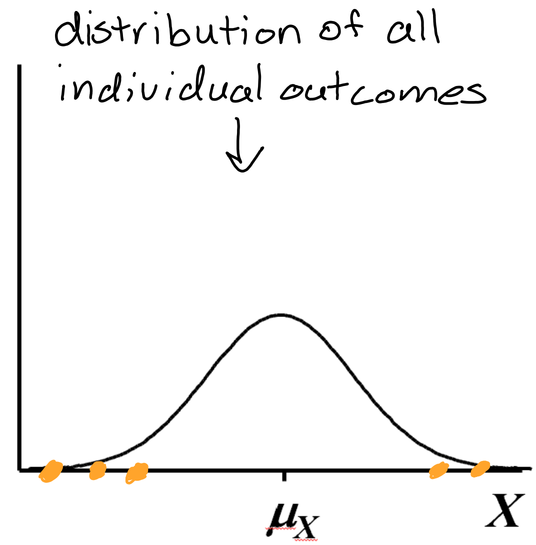
$$= \left(\frac{1}{n}\right)^2 Var(X_1 + X_2 + \ldots + X_n)$$

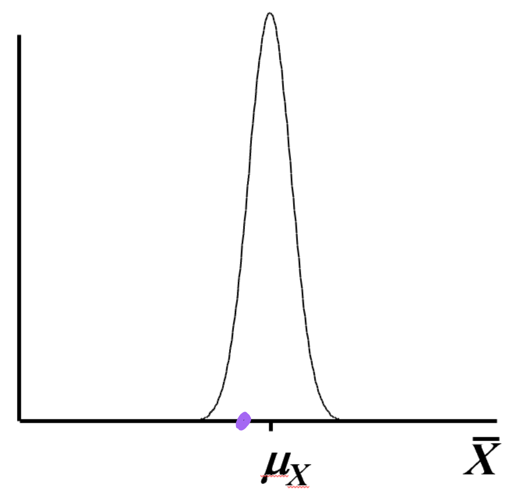
$$= \frac{1}{n^2}\left\{Var(X_1) + Var(X_2) + \ldots + Var(X_n)\right\}$$

$$= \frac{1}{n^2}(\sigma_X^2 + \sigma_X^2 + \ldots + \sigma_X^2)$$

$$\sigma_{\bar{X}}^2 = \frac{1}{n^2}(n\sigma_X^2) = \boxed{\frac{\sigma_X^2}{n}}$$

The probability density of an estimate of a parameter of X always has lower variance than distribution of X.





Even if the points we sample are far from the mean

The value obtained for $\bar{X}$ will be closer to the mean

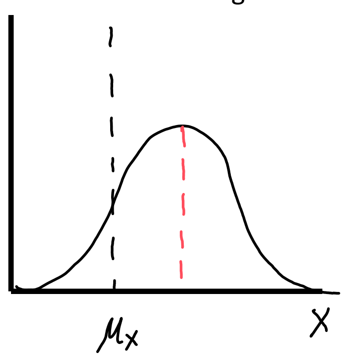
R.6 Unbiasedness and Efficiency

We need two things for a reliable estimator.

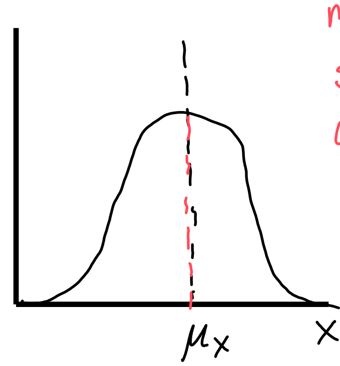
1. Unbiasedness – the expected value of the estimator of a parameter equals the true value of the parameter.
2. Efficiency – an unbiased estimator that has low variance.

Exercise 7: Label the following distributions as biased or unbiased.

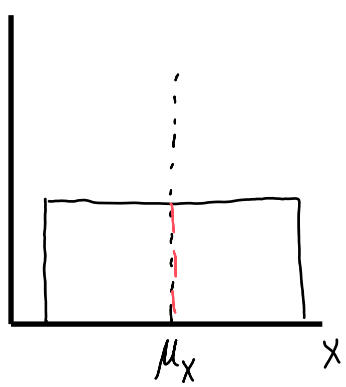
mean of sample distribution



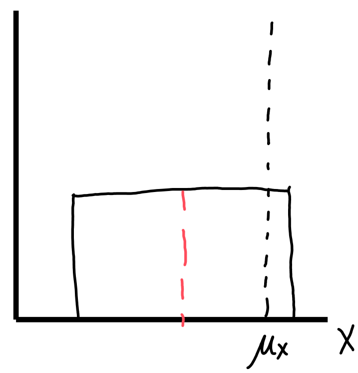
Biased



Unbiased



Unbiased



Biased

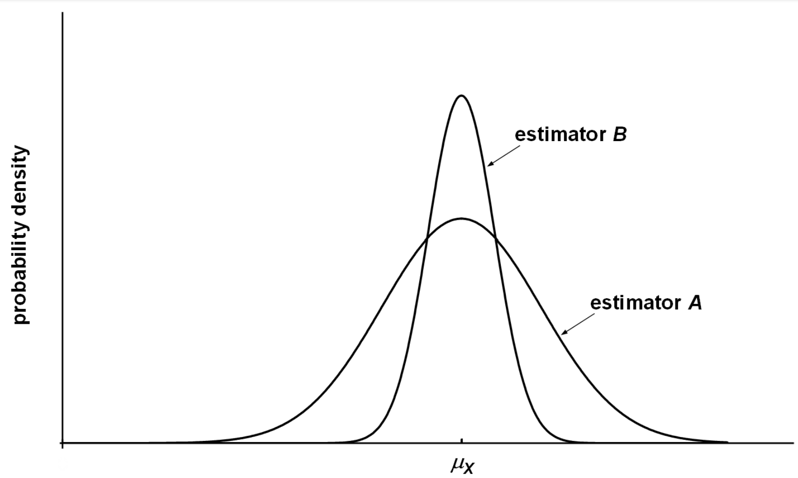
Exercise 8: Show that sample mean is an unbiased estimator of population mean.

μ_X $\bar{X}$

Unbiased $\longrightarrow E(\bar{X}) = \mu$

$$E(\bar{X}) = E\left\{\frac{1}{n}(X_1 + X_2 + \dots + X_n)\right\}$$
$$= \frac{1}{n} E(X_1 + X_2 + \dots + X_n)$$
$$= \frac{1}{n}\{E(X_1) + E(X_2) + \dots + E(X_n)\}$$
$$= \frac{1}{n}(\mu_X + \mu_X + \dots + \mu_X) = \frac{1}{n}(n\mu_X) = \boxed{\mu_X}$$

Exercise 9: Which estimator would you rather use A or B? Why?

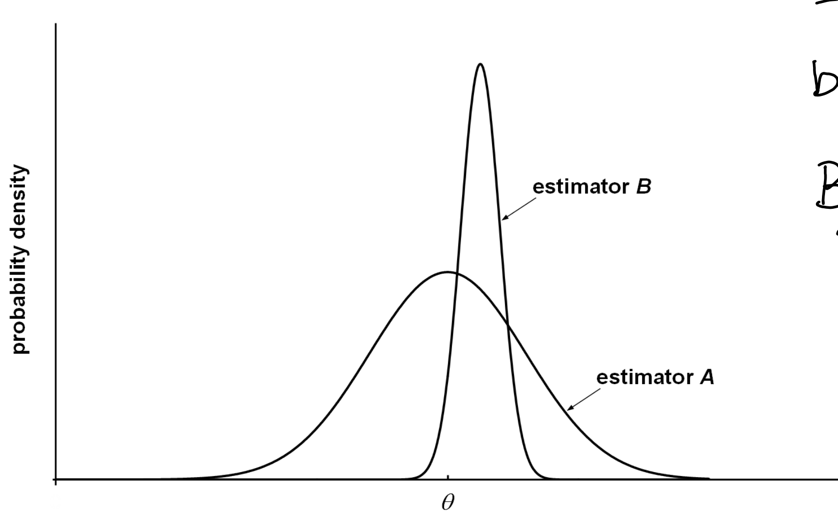


B. it has a lower variance than A and is unbiased

Exercise 10: Which estimator would you rather use, A or B? Why?

It depends. A is unbiased but has high variance.

B has low variance but is biased.



R.7 Estimators of variance, covariance, and correlation

Formula – sample variance

$$S_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

Formula – sample covariance

$$S_{XY} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})$$

Formula – sample correlation

$$r_{X,Y} = \frac{S_{XY}}{\sqrt{S_X^2 S_Y^2}}$$

Exercise 11: Find the sample means, variances, covariance, and correlation of the following sample data.

X	Y
1	4
2	5
3	6

$$\bar{X} = \tfrac{1}{3}(1+2+3) = \tfrac{1}{3}(6) = \boxed{2}$$

$$\bar{Y} = \tfrac{1}{3}(4+5+6) = \tfrac{1}{3}(15) = \boxed{5}$$

$$S_X^2 = \frac{1}{3-1}\left[(1-2)^2 + (2-2)^2 + (3-2)^2\right] = \tfrac{1}{2}\left[(-1)^2 + 0^2 + 1^2\right]$$

$$= \tfrac{1}{2}[1+1] = \tfrac{1}{2}(2) = \boxed{1}$$

$$S_Y^2 = \frac{1}{3-1}\left[(4-5)^2 + (5-5)^2 + (6-5)^2\right] = \tfrac{1}{2}\left[(-1)^2 + 0^2 + 1^2\right]$$

$$= \tfrac{1}{2}[1+0+1] = \tfrac{1}{2}(2) = \boxed{1}$$

$$S_{XY} = \frac{1}{3-1}\left[(1-2)(4-5) + (2-2)(5-5) + (3-2)(6-5)\right]$$

$$= \tfrac{1}{2}\left[(-1)(-1) + 0(0) + 1(1)\right] = \tfrac{1}{2}[1+0+1] = \tfrac{1}{2}[2] = \boxed{1}$$

$$r_{XY} = \frac{1}{\sqrt{1\cdot 1}} = \frac{1}{\sqrt{1}} = \frac{1}{1} = \boxed{1}$$

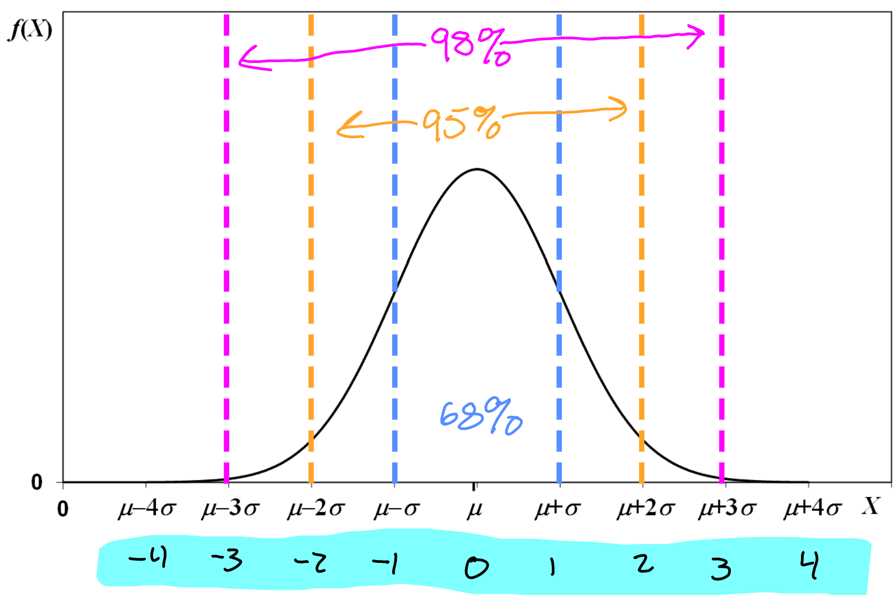
Note: S_X^2, S_Y^2, S_{XY} & r_{XY} are not always the same. It just happenned to be so for this particular example.

R.8 The normal distribution

There are 4 continuous distributions that we will use in econometrics.

1. Normal distribution
2. t distribution
3. F distribution
4. χ^2 distribution (chi-squared)

The normal distribution – a distribution that is “bell shaped” around the population mean. Most of it’s probability lies close to the middle, with relatively little far away.



Standard normal distribution – a normal distribution with mean = 0 and standard deviation = 1.

To find probabilities for a standard normal distribution, we use a z-table.

Exercise 12: What is the probability that X < 1 for a standard normal distribution?

Find this area

$$Pr(X \leq 1) = \boxed{0.8413}$$

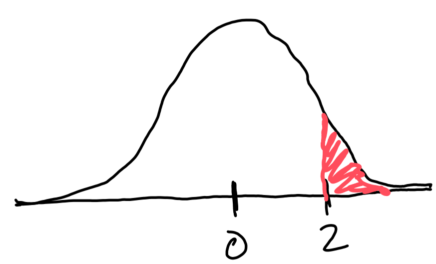
Find 1.00 on the edges, the probability is on the inside.

Exercise 13: What is the probability that X > 2 for a standard normal distribution?

$$Pr(X \geq 2) = 1 - Pr(X \leq 2)$$

$$= 1 - 0.9772$$

$$\boxed{= 0.0228}$$



Exercise 14: What is the probability that -1 < X < 0.5 for a standard normal distribution?

$Pr(-1 \leq X \leq 0.5) = Pr(X \leq 0.5) - Pr(X \leq -1)$

$= 0.6915 - 0.1587$

$\boxed{= 0.5328}$

Exercise 15: What is Z if $\Pr(X < Z) = 0.5120$?

Find 0.5120 inside the table then get the z from the edges

$\boxed{Z = 0.03}$

Exercise 16: What is Z if $\Pr(X < Z) = 0.877$?

$\boxed{Z = 1.16}$