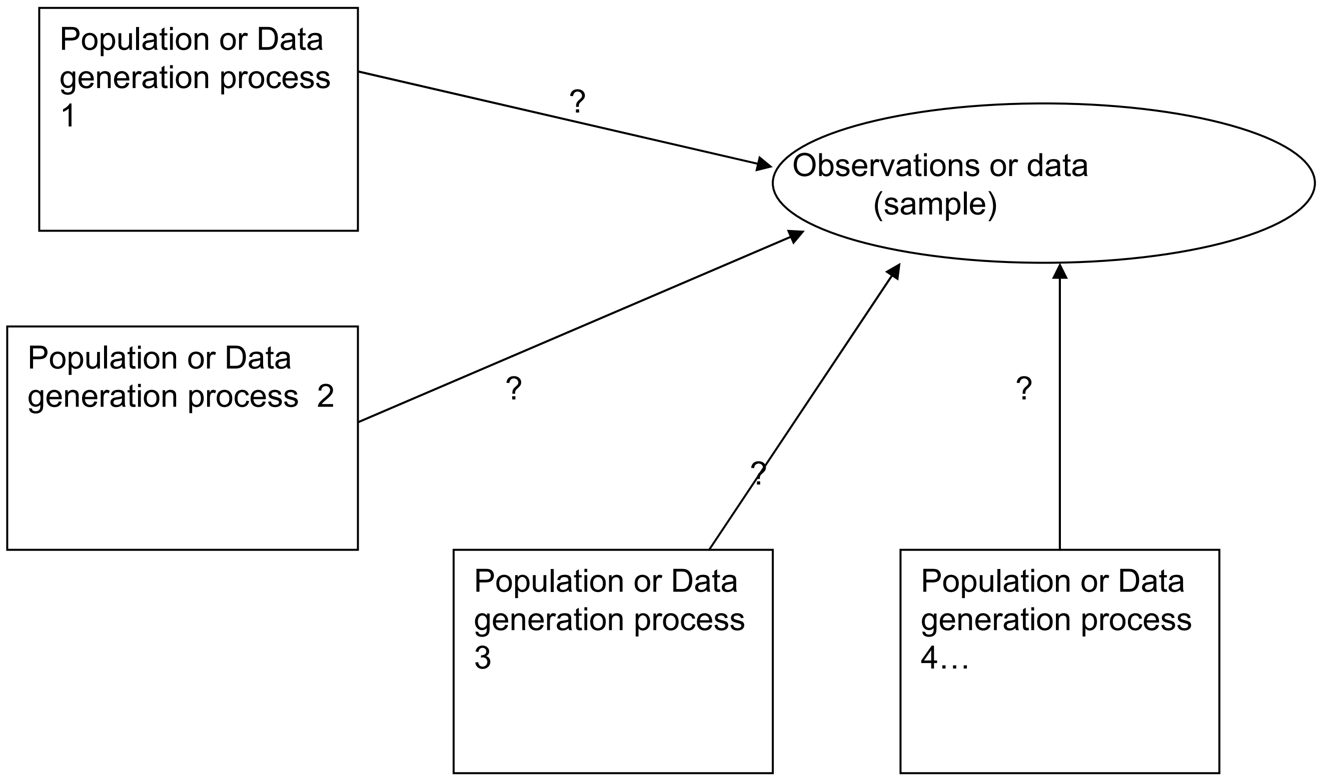
Lecture 5:
Point and Interval Estimation

Quantitative Methods for
Regulation and Competition

Our obsession…
Where does our data come from?



Some Examples

Group	**Attribute**	**Characteristic of interest**	**Sample**
People entitled to vote in the next election	Voting intention	Proportion of voters for Party A	Voting intentions of 1,000 people called
All past and future students of City University	Ownership of laptops	Proportion of owners	Ownership in the class
Grains of sand in a bucket	Mass	Average mass	Mass of 10 grains
100 students of a course	IQ level	Average IQ	IQ levels of 16 students in one class

Today's lecture

- Point estimators
- Examples: mean and variance
- Which one is better?
- Interval estimators
- Examples: mean, proportion and variance

- Reading: Chapters 7 and 8 in Ashenfelter et al.

Point Estimation

- Estimator:

  "Random variable used to estimate a characteristic (parameter) or relationship in the population"

  Formula specified before gathering the sample!

  The actual numerical value obtained is called an estimate

- Example: if we want to estimate the (unknown) mean of the population, we could use the…
  - Sample mean: $$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$
  - Or an average of the largest and the smallest values observed: $$\frac{\max\{X_i\} + \min\{X_i\}}{2}$$
  - Or…anything you can think of!

Point Estimation (2)

- Example: if we want to estimate the (unknown) variance of the population, we could use the…
  - "Sample variance":

$$S_x^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \overline{X} \right)^2$$

  - Or an alternative…

$$S_n^2 = \frac{1}{n} \sum_{i=1}^{n} \left(X_i - \overline{X} \right)^2$$

Which one is better?

- An estimator is **unbiased**…

  "if the expected value is equal to the true value"

  Examples:

$$E\left(\overline{X}\right) =$$

$$E\left(\frac{X_1 + X_6}{2}\right) =$$

$$E\left(S_n^2\right) =$$

$$E\left(S_X^2\right) =$$

First, we want an unbiased estimator!

Which one is better?

- An estimator is more **efficient** than another…

  "if it has a smaller variance"

  In other words, the distribution of the estimator is less dispersed

$$Var\left(\overline{X}\right) =$$

$$Var\left(\frac{X_1 + X_6}{2}\right) =$$

Second, among the unbiased estimators, we want the most efficient one!

Interval Estimation

- An interval estimator:

  "Range within which we have some degree of certainty the true population parameter lies"

  The actual range is called *interval estimate*

  The probability that the true parameter lies in the interval estimate is called *confidence coefficient*

  confidence coefficient = $P(\text{ LB} \leq \text{true parameter} \leq \text{UP}) = 1 - \alpha$

  The *significance level* is α

Example: Interval Estimation for the Population Mean

- Suppose that n is large. Then…

$$\overline{X} \sim N\left(\mu_X, \frac{\sigma_X^2}{n}\right)$$

- And…

$$Z = \frac{\overline{X} - \mu_X}{\sqrt{\frac{\sigma_X^2}{n}}} = \frac{\overline{X} - \mu_X}{\frac{\sigma_X}{\sqrt{n}}} \sim N(0,1)$$

- Define z_x as $F(z_x) = x$ where F is the standard normal distribution
- Note: since F is symmetric, $z_x = -z_{1-x}$

Example: Interval Estimation for the Population Mean

We have that…

$$P(-\, z_{1-\alpha/2} \le Z \le z_{1-\alpha/2}) = 1 - \alpha$$

$$\Rightarrow P\left(-z_{1-\alpha/2} \le \frac{\overline{X} - \mu_X}{\frac{\sigma_X}{\sqrt{n}}} \le z_{1-\alpha/2}\right) = 1 - \alpha$$

And therefore…

$$P\left(\overline{X} - z_{1-\alpha/2}\frac{\sigma_X}{\sqrt{n}} \le \mu_X \le \overline{X} + z_{1-\alpha/2}\frac{\sigma_X}{\sqrt{n}}\right) = 1 - \alpha$$

What if we don't know the variance of the population?

- Since…

$$\frac{\overline{X} - \mu_X}{\frac{S_X}{\sqrt{n}}} \sim t_{n-1}$$

- Following the same procedure…

$$P\left(\overline{X} - t_{n-1,1-\alpha/2}\frac{s_X}{\sqrt{n}} \le \mu_X \le \overline{X} + t_{n-1,1-\alpha/2}\frac{s_X}{\sqrt{n}}\right) = 1-\alpha$$

- However if n is large (n>120), then

$t_{n-1,1-\alpha/2} = Z_{1-\alpha/2}$ and we can use either of them

What if our sample is small?

- Then if the population is normal, we still have that

$$\overline{X} \sim N\left(\mu_X, \frac{\sigma_X^2}{n}\right)$$

- And therefore…

$$P\left(\overline{X} - z_{1-\alpha/2}\frac{\sigma_X}{\sqrt{n}} \leq \mu_X \leq \overline{X} + z_{1-\alpha/2}\frac{\sigma_X}{\sqrt{n}}\right) = 1 - \alpha$$

- And if we do not know the population variance…

$$P\left(\overline{X} - t_{n-1,1-\alpha/2}\frac{s_X}{\sqrt{n}} \leq \mu_X \leq \overline{X} + t_{n-1,1-\alpha/2}\frac{s_X}{\sqrt{n}}\right) = 1 - \alpha$$

Summarizing…

Table 8.1

Assumption	$100(1-\alpha)\%$ Confidence interval (two-sided)
n large and σ_X^2 known, or population normal and σ_X^2 known.	$P\left(\overline{X} - z_{1-\alpha/2}\frac{\sigma_X}{\sqrt{n}} \leq \mu_X \leq \overline{X} + z_{1-\alpha/2}\frac{\sigma_X}{\sqrt{n}}\right) = 1 - \alpha$
n large and σ_X^2 unknown	$P\left(\overline{X} - z_{1-a/2}\frac{s_X}{\sqrt{n}} \leq \mu_X \leq \overline{X} + z_{1-\alpha/2}\frac{s_X}{\sqrt{n}}\right) = 1 - \alpha$
n small, population normal, and σ_X^2 unknown	$P\left(\overline{X} - t_{n-1,1-a/2}\frac{s_X}{\sqrt{n}} \leq \mu_X \leq \overline{X} + t_{n-1,1-\alpha/2}\frac{s_X}{\sqrt{n}}\right) = 1 - \alpha$

Example: Interval Estimation for the Population Proportion

- We have a sample $X_1,..X_n$ can take values 0 and 1
- The unknown proportion of 1's in the population is π
- Take the following estimator… $$\hat{\pi} = \frac{1}{n}\sum_{i=1}^{n} X_i$$
- Is it unbiased? What is the variance?
- If n is large, we have that.. $$\hat{\pi} \sim N\left(\pi, \frac{\pi(1-\pi)}{n}\right)$$
- And therefore…

$$P\left(\hat{\pi} - z_{1-\alpha/2}\frac{\sqrt{\hat{\pi}(1-\hat{\pi})}}{\sqrt{n}} \le \pi \le \hat{\pi} + z_{1-\alpha/2}\frac{\sqrt{\hat{\pi}(1-\hat{\pi})}}{\sqrt{n}}\right) = 1-\alpha$$

Example: Interval Estimation for the Population Variance

- Since…

$$\frac{(n-1)s_X^2}{\sigma_X^2} \sim \chi^2_{n-1}$$

- Then…

$$P\left(\chi^2_{\alpha/2,n-1} \leq \frac{(n-1)s_X^2}{\sigma_X^2} \leq \chi^2_{1-\alpha/2,n-1}\right) = 1-\alpha$$

- Following the same procedure…

$$P\left(\frac{(n-1)s_X^2}{\chi^2_{1-\alpha/2,n-1}} \leq \sigma_X^2 \leq \frac{(n-1)s_X^2}{\chi^2_{\alpha/2,n-1}}\right) = 1-\alpha$$