Statistics vs. parameters: descriptive vs. inferential statistics

	Sample Statistic	Population Parameter
Mean	$\overline{Y}$	μ (mu)
Standard deviation	s	σ ($sigma$)

Sampling distribution

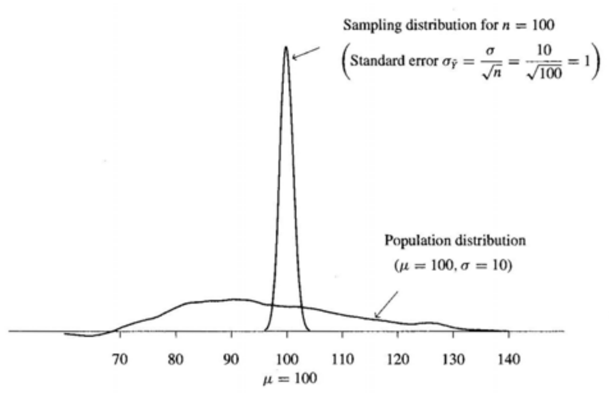
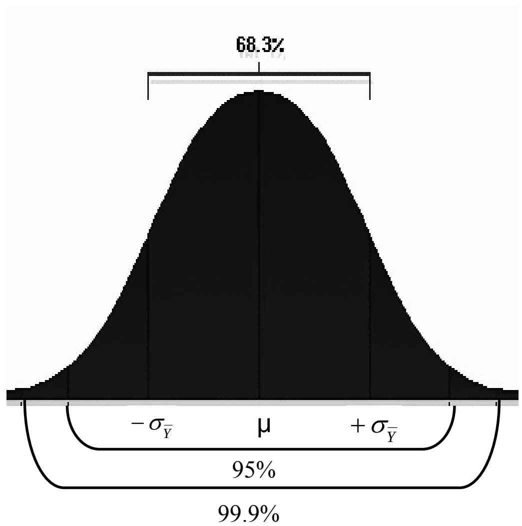


Figure 4.12 A Population Distribution and the Sampling Distributions of $\bar{Y}$ for $n = 100$

Source: Agresti and Finlay, 1997, Fig. 4.12

Sampling distribution



Standard error

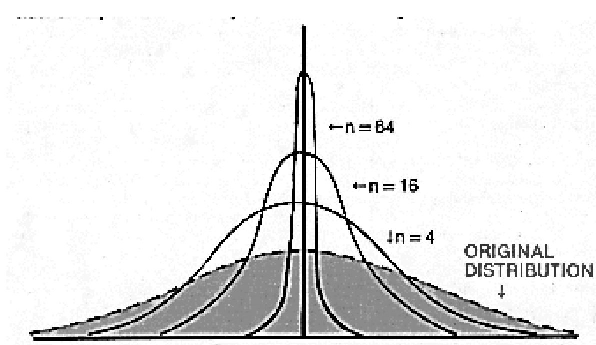
$$\sigma_{\overline{Y}} = \sigma / \sqrt{n}$$

Standard error estimate

(if don't know population σ)

$$\hat{\sigma} = s = \sqrt{\frac{\sum (Y_i - \overline{Y})^2}{n-1}}$$

Sampling distribution: normally distributed



Simple regression
(Murdock data)

Country	Political integration (X)	Stratification (Y)	Country	Political integration (X)	Stratification (Y)
001	2	1	011	1	1
003	3	2	013	0	0
005	3	3	015	2	1
007	4	2	017	2	1
009	0	0	019	4	2

Simple regression

$$\hat{Y} = a + bX$$

$$where\,, a = \text{intercept}\,, b = \text{slope}$$

$$b = \left[\frac{\Delta \text{Y}}{\Delta \text{X}}\right]$$

Step 1: calculate b
(unstandardized regression coefficient)

$$b_{yx} = \frac{\text{cov}(x, y)}{\text{var}(x)}$$

$$b_{yx} = \frac{\sum_{i=1}^{n} (X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n} (X_i - \overline{X})^2}$$

Step 1: calculate b
(unstandardized regression coefficient): computational formula

$$= \frac{n\sum X_i Y_i - (\sum X_i)(\sum Y_i)}{n\sum X_i^2 - (\sum X_i)^2}$$

Step 1: calculate b

(how does stratification change relative to integration: computation formula)

$$= \frac{10(38) - (21)(13)}{10(63) - (21)^2}$$

$$= \frac{107}{189}$$

$$b_{yx} = .566$$

Step 2: find a

$$\overline{Y} = a + b\overline{X}$$

$$\overline{Y} - b\overline{X} = a$$

$$or,\ a = \overline{Y} - b\overline{X}$$

$$a = 1.3 - (.566)(2.1)$$

$$a = .111$$

Step 2: prediction equation

thus,

$$\hat{Y} = .111 + .566X$$

Step 2: standardized coefficient

$$B_{yx} = b_{yx}(s_x / s_y)$$
$$= .566(s_x / s_y)$$

Interpretation

- b_{yx} (unstandardized coeff.) = 1 unit change in political integration produces a .566 change in stratification
- B_{yx} (standardized coeff.) = relative effect, or strength of association (= r_{yx} in 1 variable case)

Predicting

$$\hat{Y} = .111 + .566X$$

$$\hat{Y} = .111 + .566\,(1)$$

$$\hat{Y} = .677\ (i.e., 1, .677)$$

r^2: how good is prediction?

$$SSE = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2$$

$$TSS = \sum_{i=1}^{n} (Y_i - \overline{Y})^2$$

$$r^2 = \frac{TSS - SSE}{TSS}$$

$$\frac{\text{total SS} - \text{residual SS}}{\text{total SS}}$$

r^2: how good is prediction?

r^2 interpretation:

"Proportion of variance in dependent variable "explained" by knowing X"

How good is prediction?

r = measure of goodness of fit

$$r = \frac{\operatorname{cov}(x, y)}{\sqrt{\operatorname{var} x}\sqrt{\operatorname{var} y}}$$

How good is prediction?
Computation formula

$$r = \frac{n\sum X_i Y_i - (\sum X_i)(\sum Y_i)}{\sqrt{[n\sum X_i^2 - (\sum X_i)^2][n\sum Y_i^2 - (\sum Y_i)^2]}}$$

How good is prediction?

$$r = \frac{10(38) - (21)(13)}{\sqrt{[10(63) - (21)^2][10(25) - (13)^2]}}$$

How good is prediction?

r = .865

= Byx in simple regression case

Interpretation: suggests strong + relationship between political integration and stratification

Characteristics of r's

1) Symmetrical ($r_{xy} = r_{yx}$)
2) Standardized
3) Range: -1 to +1
4) r has same sign as b and B
5) Larger absolute value, stronger the association
6) r not appropriate if nonlinear

How good is prediction?

$r^2 = .748$

Interpretation: 75% of the variance in social stratification is explained by political integration