Lecture 3: Estimating the Mean and Variance of a Normal Population

by
Professor Scott H. Irwin

Required Reading:

Griffiths, Hill and Judge. "Statistical Inference I: Estimating the Mean and Variance of a Normal Population," Ch. 3 in Learning and Practicing Econometrics
Overview

In the previous lecture, we developed the concept of a random variable

A random variable can be thought of as the probabilistic process that creates numbers

Each random variable has two main properties (parameters) that govern the numbers it produces: mean and variance

- In developing the concept of a random variable, assumed we knew the values for mean and variance

In reality, we rarely know the true mean and variance values for a random variable (and hence, the population)

Statistical theory of estimation shows how to use sample data to estimate the properties (mean and variance) of a random variable
The concept of a **sampling distribution** is central to the statistical theory of estimation

Provides the unifying logic to all of the algebraic manipulations performed in econometrics

**Terminology Note: Relationship between Random Variable and Population**

Random variable is a probabilistic process (“machine”) that creates numbers

A listing of all of the possible values that a random variable can generate is the population

- Population is **finite** for a discrete random variable
- Population is **infinite** for a continuous random variable

So, we can refer to the **pdf** for a random variable as the **“population” distribution**, and **vice versa**

⇒ Mean and variance of the random variable and population are the same
FIGURE 18.1 A thought experiment helps explain the nature of the sampling distribution of $\bar{X}$. We start with a continuous random variable $X$ whose probability distribution is shown at the top of the figure. Next, we take $N$ different samples of size $n$, and in each sample we calculate the mean. These sample means are collected, and their frequency distribution is constructed. Theoretically, as the number of samples ($N$) increases to infinity, this frequency distribution reaches a limiting form. That limiting form is the sampling distribution of $\bar{X}$, which is the probability distribution that governs the likelihood of different values occurring for $\bar{X}$ in a sample of size $n$ drawn from the original $X$.

Key Implications of Sampling Distribution

Example

How close $\overline{x}$ is to the true mean $\beta_x$ depends on the particular sample of $x$’s drawn

- If a sample of $x$’s closely resembles the population distribution, then $\overline{x}$ will be relatively close to the true mean parameter $\beta_x$

- If a sample of $x$’s does not resemble the population distribution, then $\overline{x}$ will not be relatively close to the true mean parameter $\beta_x$

- If a sample of $x$’s tends to be drawn from the upper (lower) portion of the population distribution, then $\overline{x}$ will tend to over-estimate (under-estimate) the true mean parameter $\beta_x$
Sampling Distributions and Learning Econometrics

Assume we need to estimate the unknown mean, $\beta_x$, of random variable $X$.

The formula, or estimator, used to estimate $\beta_x$ from a sample of data is represented by $\beta^*_x$ (e.g. arithmetic mean).

1. Using $\beta^*_x$ to estimate $\beta_x$ is analogous to an econometrician obtaining an estimate by blindly reaching into the sampling distribution of $\beta^*_x$ to obtain a single estimate.

2. Because of (1), choosing between $\beta^*_x$ and an alternative estimator $\beta^{**}_x$ is based on answering the following question: Would you prefer to produce your estimate of $\beta_x$ by reaching blindly into the sampling distribution for $\beta^*_x$ or $\beta^{**}_x$?

3. Because of (2), desirable properties of an estimator $\beta^*_x$ are defined in terms of its sampling distribution.
4. The properties of the sampling distribution of an estimator depend on the process generating the data, so an estimator can be a good one in one context but a bad one in another.

5. Properties of the sampling distribution also depend on the sample size.

6. Most algebraic derivations in econometrics focus on trying to find the characteristics of the sampling distribution of an estimator, such as the mean and variance.

7. All statistics, not just parameter estimators, have sampling distributions (e.g. $t$-statistic, $F$-statistics).

**Computing Sampling Distributions**

Three techniques for determining the properties of an estimator's sampling distribution:

- Algebraic derivations for any sample size
- Algebraic derivations only for large sample sizes
- Monte Carlo simulation
Census vs. Sampling

**Census:** complete *enumeration* of population values

- In theory, allows computation of *true* value of parameter
- **Non-sampling** errors may be a big problem
- Usually prohibitively expensive

**Sample:** obtain numerical values of random variable for a *subset* of the population

- Much cheaper, so can concentrate resources on minimizing or eliminating *non-sampling errors*
- Focus on sampling error
The Problem

How much does a typical household spend on food during a week?

- Problem is of interest to food retailers and policy analysts

The Economic Model

Food expenditure depends on a number of factors such as income, family size, etc.

To keep the example tractable, a simple model is considered where food expenditure, $Y$, is a fixed constant,

$$ Y = \beta $$

Focus on the economic variable of weekly food expenditure by households of size 3 with income of $25,000/yr.

- In this way, at least control for household size and income
**The Statistical Model**

The simple economic model implies that food expenditure will be the same for all families of size 3 with incomes of $25,000/yr.

Actual food expenditure for this group will not likely be a constant because

- Effects of uncontrolled variables
- Measurement error
- Random component of human behavior

We add an error term to the economic model to represent these random differences

- Economic relationships are not exact in the real world
- Essential difference between an economic and statistical model
The resulting statistical model is,

\[ Y_t = \beta + e_t \quad t = 1,...,T \]

where \( Y_t \) represents the \( t^{th} \) household's food expenditure, \( e_t \) is the random error term, and \( T \) is the number of households to be sampled.

For a given observation \( t \), \( Y_t \) can be thought of as having two components:

- A **systematic** component \( \beta \), that is determined by an **economic** process
- A **random** component \( e_t \), that is determined by a **probabilistic** process

Another way of saying the same thing is that the random variable \( Y_t \) is simply a linear transformation of the random variable \( e_t \)

\[ Y_t = a + be_t \quad \text{where} \ a = \beta \ and \ b = 1 \]
Since the error term (disturbance) is the random variable that "drives" the model, we need to be more explicit about its properties.

Our first impulse is to think of the error term, $e_t$, as a single random variable for which we will take one draw of $T$ observations.

It is actually more useful to envision the error term as consisting of a set of single draws from $T$ identical random variables.

For each of the $T$ random variables $e_t$, we assume

- $E(e_t) = 0$
- $\text{var}(e_t) = E[(e_t - E(e_t))^2] = E[e_t]^2 = E[e_t^2] = \sigma^2$
- $e_t$ follows a normal distribution
- $e_t$ are independent so that $\text{Cov}(e_t, e_s) = 0$ for all $t \neq s$
The assumptions for the error term can be summarized using the following notation,

\[ e_t \sim N(0, \sigma^2) \quad t = 1, \ldots, T \]

Often this is referred to as the *iid* normality assumption, which is shorthand for *identical, independently distributed normal random variables*

- Think of each \( e_t \) as being a "lottery", where the winning number ("draw") is produced by a random number generator
- The random number generators are all identical and follow a normal distribution with mean 0 and variance \( \sigma^2 \)
- The random number generators are *independent*, which means the outcome of one "lottery" does not affect the outcome of another "lottery"
As a result of these assumptions, the expected value and variance for $Y_t$ are,

$$E(Y_t) = E(\beta + e_t) = E(\beta) + E(e_t) = \beta + 0 = \beta$$

$$\text{var}(Y_t) = \text{var}(\beta + e_t) = \text{var}(e_t) = \sigma^2$$

which can be represented as,

$$Y_t \sim N(\beta, \sigma^2) \quad t = 1,\ldots,T$$
Assumptions of the Constant Mean Model

CM1. $y_t = \beta + e_t, \quad t = 1,...,T$

CM2. $E(e_t) = 0 \iff E(y_t) = \beta$

CM3. $\text{var}(e_t) = \text{var}(y_t) = \sigma^2$

CM4. $\text{cov}[e_t,e_s] = \text{cov}[y_t,y_s] = 0 \quad t \neq s$

CM5. $e_t \sim N(0,\sigma^2) \iff y_t \sim N(\beta,\sigma^2)$
Figure 3.2 The probability density functions of $e_i$ and $Y$.

Notes:

- Sometimes the statistical model is referred to as the "data generating process" for $Y$

- This simple model is a useful “training ground” for the more complex linear regression model

- It is important to emphasize that we now consider any sample observation to be one of many that might have occurred due to random chance
Estimating the Mean

After specifying the economic and statistical models, we can collect data and begin the process of estimating the unknown parameter $\beta$

Data on 40 randomly selected households in Table 3.1

- Random selection should result in sample data that reflects general characteristics of population

- Denote the observed sample values of $Y_t$ as $y_t$

First question: How do we use the sample data to estimate the population mean $\beta$?

- Note that $E(Y_t) = \beta$, or the "center" of the pdf for $Y_t$

- Suggests "center" of sample data may yield a good estimate of the population mean $\beta$

- Need a rule to find the "center" of sample data
### Table 3.1 Weekly Expenditures on Food by \( T = 40 \) Randomly Selected Households of Size 3 and $25,000 Income

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>49.98</td>
<td>55.33</td>
<td>50.39</td>
<td>46.83</td>
</tr>
<tr>
<td>47.43</td>
<td>48.47</td>
<td>52.73</td>
<td>52.32</td>
</tr>
<tr>
<td>50.35</td>
<td>47.57</td>
<td>49.52</td>
<td>51.56</td>
</tr>
<tr>
<td>55.17</td>
<td>52.65</td>
<td>51.43</td>
<td>45.96</td>
</tr>
<tr>
<td>50.86</td>
<td>52.57</td>
<td>50.10</td>
<td>49.74</td>
</tr>
<tr>
<td>49.07</td>
<td>48.91</td>
<td>58.06</td>
<td>51.26</td>
</tr>
<tr>
<td>51.52</td>
<td>51.10</td>
<td>52.92</td>
<td>53.07</td>
</tr>
<tr>
<td>49.35</td>
<td>54.85</td>
<td>54.14</td>
<td>47.75</td>
</tr>
<tr>
<td>50.88</td>
<td>49.44</td>
<td>53.02</td>
<td>53.50</td>
</tr>
<tr>
<td>49.65</td>
<td>52.39</td>
<td>49.81</td>
<td>51.85</td>
</tr>
</tbody>
</table>

Consider the $T$ sample points $y_1, y_2, \ldots, y_T$ as stores along a street front

- If you want to visit the stores one at a time, where should you park your car to minimize the distance walked?

- The principle of least squared distance can be used to find the central parking spot that minimizes walking

This analogy demonstrates the motivation for the use of the least squares principle in finding the center of a data sample

Graphically, we want to minimize the sum of squares of the vertical distances between a flat line and the sample observations
Minimizing the sum of squared errors
More formally, the least squares "center" of a data sample is the value for $\beta$ that minimizes

$$S = \sum_{t=1}^{T} (y_t - \beta)^2 = \sum_{t=1}^{T} e_t^2$$

Since the values for $y_t$ are known, $S$ is solely a function of the unknown parameter $\beta$

Multiplying the expression out, $S$ can be rewritten as,

$$S = \sum_{t=1}^{T} y_t^2 - 2\beta \sum_{t=1}^{T} y_t + T\beta^2$$

which is a quadratic in terms of $\beta$
Figure 3.3 The sum of the squares parabola.

Mathematically, minimum (point of zero slope) is found by differentiating $S$ with respect to $\beta$,

$$\frac{dS}{d\beta} = -2 \sum_{t=1}^{T} y_t + 2T \beta$$

The value of $\beta$ that makes the derivative zero is the least squares estimate of $\beta$, which is denoted as $b$,

$$-2 \sum_{t=1}^{T} y_t + 2Tb = 0$$

Solving for $b$,

$$b = \frac{\sum_{t=1}^{T} y_t}{T}$$

**Important comments**

- $b$, the least squares estimate of $\beta$, is nothing more than the *arithmetic average* of the sample data

- In intro stats, $b$ is usually called $\bar{y}$
• For the sample of 40 households,

\[
b = \frac{\sum_{t=1}^{T} y_t}{T} = \frac{2043.30}{40} = 51.08
\]

• **Estimate** of **population** mean weekly food expenditure is $51.08

• Recall the statistical model,

\[Y_t = \beta + e_t \quad t = 1, \ldots, T\]

Implies individual household weekly food expenditure is equal to the **population mean expenditure**, estimated to be $51.08, and a **random error** that reflects household tastes, preferences, and other factors not controlled.
<table>
<thead>
<tr>
<th>Weekly Expenditures on</th>
<th>49.98</th>
<th>55.33</th>
<th>50.39</th>
<th>46.83</th>
</tr>
</thead>
<tbody>
<tr>
<td>Food by T = 40</td>
<td>47.43</td>
<td>48.47</td>
<td>52.73</td>
<td>52.32</td>
</tr>
<tr>
<td>Randomly Selected</td>
<td>50.35</td>
<td>47.57</td>
<td>49.52</td>
<td>51.56</td>
</tr>
<tr>
<td>Households of Size 3</td>
<td>55.17</td>
<td>52.65</td>
<td>51.43</td>
<td>45.96</td>
</tr>
<tr>
<td>and $25,000 Income</td>
<td>50.86</td>
<td>52.37</td>
<td>50.10</td>
<td>49.74</td>
</tr>
<tr>
<td></td>
<td>49.07</td>
<td>48.91</td>
<td>58.06</td>
<td>51.26</td>
</tr>
<tr>
<td></td>
<td>51.52</td>
<td>51.10</td>
<td>52.92</td>
<td>53.07</td>
</tr>
<tr>
<td></td>
<td>49.35</td>
<td>54.85</td>
<td>54.14</td>
<td>47.75</td>
</tr>
<tr>
<td></td>
<td>50.88</td>
<td>49.44</td>
<td>53.02</td>
<td>53.50</td>
</tr>
<tr>
<td></td>
<td>49.65</td>
<td>52.39</td>
<td>49.81</td>
<td>51.85</td>
</tr>
</tbody>
</table>

Sampling Distribution of the Mean Estimator

Using the least squares estimation rule, we computed the sample value of $b$ to equal $51.08$

- Key question: Is $51.08$ a good estimate of the true population mean?

- The problem: To answer this question, the true value of $\beta$ must be known!

- The result: Accuracy of sample estimate cannot ever be directly assessed

⇒ Strictly speaking, we cannot make any statement about the accuracy of an estimate from a sample (unsettling!!)

We can only examine the properties of the least squares rule, or estimator, rather than a particular estimate

- We want to “hit” the estimator with many hypothetical samples and analyze the performance of the estimator across these hypothetical samples
• Analagous to using a flight simulator to assess performance of airline pilots

**Basic Idea**

Before actually generating the sample, the values of \( Y_t \) are unknown

Hence, before sampling the \( Y_t \) are random variables

The least squares rule states that no matter what the \( Y_t \) turn out to be, calculate,

\[
b = \frac{\sum_{t=1}^{T} Y_t}{T}
\]

**Key observation:** Since \( b \) is a function of the \( T \) random variables \( Y_t \), \( b \) is a random variable as well!

• In other words, different samples of data will yield different values for \( b \)

• For this reason, the variability of the estimation rule \( b \) is called sampling variability
• The random variable $b$ has a distribution (pdf) called a **sampling distribution**, which has characteristics called **sampling properties**

• Analysis of estimator focuses on sampling properties

**Summary:**

We cannot ask whether $b = 51.08$ is a good estimate, but we can ask whether the estimator

$$b = \frac{\sum_{i=1}^{T} Y_i}{T}$$

is a “good” rule to use.

*In other words, how “good” is the formula or “recipe” by which the data are transformed into an actual estimate?*
Properties of "Good" Estimators

Since there are many possible rules to use in estimating population parameters, such as $\beta$, we need criteria to separate the "good" from the "bad"

*Any fool can produce an estimator of $\beta$, since literally an infinite number of them exist... What distinguishes an econometrician is the ability to produce “good” estimators, which in turn produce “good” estimates.*

---Peter Kennedy

A number of criteria have been proposed by econometricians to measure the “goodness” of an estimator

The four main criteria are:

- **Computational cost**: Simplicity and ease of calculation

  $\Rightarrow$ Estimator is a linear function of sample data
• **Unbiasedness**: In repeated sampling, the estimator generates estimates that on average equal the population parameter

⇒ Hits the target on average

• **Efficiency**: Of all possible unbiased estimators, there is no other estimator that has a smaller variance in repeated sampling

⇒ For a given sample size, no other unbiased rule will produce more precise shots at the target

• **Consistency**: As the sample size increases the probability mass of the estimator "collapses" on the population parameter

⇒ Increasing the sample size causes the shots to get "closer" to the target

We will examine the least squares estimator \( b \) to see if it meets these four criteria
Sampling Properties of the Least Squares Estimator $b$

Computational Cost

We want to determine whether $b$ is a linear estimator, which generally implies it will be simple and relatively costless to compute.

It is useful to rewrite the formula for $b$ as,

$$b = \frac{1}{T} \sum_{t=1}^{T} Y_t = \frac{1}{T} Y_1 + \frac{1}{T} Y_2 + \ldots + \frac{1}{T} Y_T$$

which shows that $b$ is a weighted-average of the $Y_t$, with the weights equaling $1/T$.

So we can further re-state the rule as,

$$b = \sum_{t=1}^{T} a_t Y_t = a_1 Y_1 + a_2 Y_2 + \ldots + a_T Y_T$$

where $a_t = 1/T$

- Any estimator that can be written in the above form is a linear estimator.
Unbiasedness

We want to know whether the expected value of \( b \) is in fact equal to \( \beta \).

Begin by re-stating the formula,

\[
b = \frac{1}{T} \sum_{t=1}^{T} Y_t = \frac{1}{T} Y_1 + \frac{1}{T} Y_2 + \ldots + \frac{1}{T} Y_T
\]

Taking expectations,

\[
E[b] = E \left[ \frac{1}{T} \sum_{t=1}^{T} Y_t \right] = E \left[ \frac{1}{T} Y_1 \right] + E \left[ \frac{1}{T} Y_2 \right] + \ldots + E \left[ \frac{1}{T} Y_T \right]
\]

\[
E[b] = \frac{1}{T} E[Y_1] + \frac{1}{T} E[Y_2] + \ldots + \frac{1}{T} E[Y_T]
\]

\[
E[b] = \frac{1}{T} \beta + \frac{1}{T} \beta + \ldots + \frac{1}{T} \beta
\]

\[
E[b] = \beta
\]
Efficiency

For a given sample size, we want to know whether the sampling variance of $b$ is smaller than any other linear estimator.

To begin then, we need to derive the formula for the sampling variance of $b$.

This derivation is based on the earlier result that the variance of a sum of random variables equals the sum of the variances if the random variables are independent.

Since the formula for $b$ is,

$$ b = \frac{\sum_{t=1}^{T} Y_t}{T} = \frac{1}{T} Y_1 + \frac{1}{T} Y_2 + \ldots + \frac{1}{T} Y_T $$

the sampling variance of $b$ is derived as follows,

$$ \text{var}[b] = \text{var} \left[ \frac{\sum_{t=1}^{T} Y_t}{T} \right] = \text{var} \left[ \frac{1}{T} Y_1 \right] + \text{var} \left[ \frac{1}{T} Y_2 \right] + \ldots + \text{var} \left[ \frac{1}{T} Y_T \right] $$
\[
\text{var}[b] = \frac{1}{T^2} \text{var}[Y_1] + \frac{1}{T^2} \text{var}[Y_2] + \ldots + \frac{1}{T^2} \text{var}[Y_T]
\]

\[
\text{var}[b] = \frac{1}{T^2} \sigma^2 + \frac{1}{T^2} \sigma^2 + \ldots + \frac{1}{T^2} \sigma^2
\]

\[
\text{var}[b] = \frac{T}{T^2} \sigma^2
\]

\[
\text{var}[b] = \frac{\sigma^2}{T}
\]

This shows the important result that the sampling variance of \( b \) decreases as the sample size \( T \) increases.

The proof of the efficiency of the sampling variance of \( b \) is rather lengthy and tedious.

- Let it suffice to say that this can be proved.
- For all unbiased estimators, \( \text{var}[b] \) is the smallest sampling variance possible.

Proof is found on pp. 85-86 of Griffith, Hill, and Judge.
Figure 3.5 Probability distributions of $b$ and $b^*$. 

**Consistency**

We want to show that as the sample size increases the probability mass of the estimator "collapses" on the population parameter.

This can be demonstrated informally by noting the formula for the sampling variance of $b$,

$$\text{var}[b] = \frac{\sigma^2}{T}$$

and that

$$\lim_{T \to \infty} \text{var}[b] = \frac{\sigma^2}{T} = 0$$

This can also be seen visually by inspecting Figure 3.4.
Figure 3.4 Probability density functions for the least squares estimator based on different sample sizes $T_1 > T_2 > T_3$.

Summary of Sampling Properties Discussion

Through algebraic derivations, we have shown that the least squares estimator $b$ of the population mean $\beta$ is,

- Linear
- Unbiased
- Efficient
- Consistent

The first three properties are sufficient to prove that $b$ is the best linear unbiased estimator (BLUE) of $\beta$

- Note that proof is based on algebra, not application of the estimator to various data samples

- Implies that no other estimator is “better” for any data sample (under the assumptions of the derivations)

- In this context "best" implies minimum variance

- BLUE estimators have an important place in econometrics
Normality of Sampling Distribution for $b$

It is straightforward to show that the sampling distribution of $b$ is normal.

The formula for the least squares estimator $b$ can be written as,

$$ b = \sum_{t=1}^{T} a_t Y_t = a_1 Y_1 + a_2 Y_2 + \ldots + a_T Y_T $$

where $a_t = 1/T$

This version makes it clear that $b$ is a linear combination of the $Y_t$ random variables.

- The $Y_t$ random variables are normally distributed because $e_t$ is normally distributed.
- A linear combination of normally distributed random variables is itself normally distributed.

$\Rightarrow b$ is normally distributed.
We can state that the sampling distribution of $b$ is normal with mean $\beta$ and variance $\frac{\sigma^2}{T}$, or,

$$b \sim N\left(\beta, \frac{\sigma^2}{T}\right)$$

This result is the foundation of confidence interval construction and hypothesis tests.
Estimating the Variance

To this point, we have (exhaustively!) discussed the estimation of $\beta$, the population mean.

Recall that the statistical model implies that,

$$Y_t \sim N(\beta, \sigma^2) \quad t = 1, ..., T$$

Unless $\sigma^2$ is known, which is highly unlikely, it will have to be estimated as well.

Estimation of $\sigma^2$ proceeds differently than in the case of $\beta$.

We cannot use the least squares principle, as $\sigma^2$ does not appear in the sum of squares function,

$$S = \sum_{t=1}^{T} (y_t - \beta)^2$$

Instead, we apply a "heuristic" procedure based on the definition of $\sigma^2$. 
The original definition of $\sigma^2$ in the statistical model is,

$$\text{var}(Y_t) = \text{var}(e_t) = \sigma^2 = E[e_t^2]$$

In other words, the variance is the expected value of the squared errors (Why?)

Given this definition, it would be natural to estimate $\sigma^2$ as the average of the squared errors

In order to do this, we must first obtain estimates of the population errors using our sample data as

$$\hat{e}_t = y_t - b$$

We can then develop our sample estimator of $\sigma^2$ as,

$$\hat{\sigma}^2 = \frac{\hat{e}_1^2 + \hat{e}_2^2 + \ldots + \hat{e}_T^2}{T - 1} = \frac{\sum_{i=1}^{T} \hat{e}_i^2}{T - 1}$$

Notice the squared sample errors are averaged by dividing by $T-1$ not $T$

- This accounts for the fact that the $T$ sample errors are not independent (Why?)
Sampling Properties of Variance Estimator

The rule derived for estimating the population variance ($\sigma^2$) is,

$$\hat{\sigma}^2 = \frac{\sum_{t=1}^{T} e_t^2}{T - 1}$$

Just as was the case with $b$, the estimator of the population mean, we are interested in the sampling properties of $\hat{\sigma}^2$

The same four criteria are applied when asking whether $\hat{\sigma}^2$ is a "good" estimation rule

- Linearity, unbiasedness, efficiency, consistency

It is obvious that $\hat{\sigma}^2$ is not a linear estimator

It can be shown that $\hat{\sigma}^2$ is unbiased, efficient, and consistent

- Best unbiased estimator (BUE)
- Proof can be found in advanced econometrics books
The next issue is the form of the sampling distribution of the variance estimator $\hat{\sigma}^2$

Before deriving the sampling distribution, we need to introduce a new distribution

Let $Z_1, Z_2, \ldots, Z_m$ denote $m$ independent standard normal random variables, where each is distributed $N(0,1)$

Next, form a new random variable as

$$V = Z_1^2 + Z_2^2 + \ldots + Z_m^2$$

We say $V$ is distributed as a chi-square with $m$ degrees of freedom

- Mean: $E[V] = m$

- Variance: $\text{var}[V] = E[V - m]^2 = 2m$
Figure 3.6 The chi-square distribution.

To derive the sampling distribution of the variance estimator $\hat{\sigma}^2$, we first note that the random errors are distributed as

$$e_t \sim N(0, \sigma^2) \quad t = 1, \ldots, T$$

Consequently,

$$\frac{e_t}{\sigma} \sim N(0, 1) \quad t = 1, \ldots, T$$

and squaring,

$$\left(\frac{e_t}{\sigma}\right)^2 \sim \chi_1 \quad t = 1, \ldots, T$$

If we sum over the $T$ transformed random error terms

$$\sum_{t=1}^{T} \left(\frac{e_t}{\sigma}\right)^2 \sim \chi_T$$
Based on the previous result, we can generate the sampling distribution of $\hat{\sigma}^2$,

$$\hat{\sigma}^2 \sim \frac{\sigma^2}{T-1} \chi^2_{T-1}$$

Note that the sampling distribution of $\hat{\sigma}^2$ is proportional to a chi-square with $T-1$ degrees of freedom.

Finally, using $\hat{\sigma}^2$ we can obtain a best unbiased estimator of the variance of the least squares estimator $b$,

$$\text{vâr}(b) = \frac{\hat{\sigma}^2}{T}$$

The square root of this estimator is called the "standard error" of $b$,

$$\text{se}(b) = \sqrt{\text{vâr}(b)} = \sqrt{\frac{\hat{\sigma}^2}{T}} = \frac{\hat{\sigma}}{\sqrt{T}}$$

• Whenever you see the term "standard error" remember it applies to an estimate.
<table>
<thead>
<tr>
<th>Household</th>
<th>$Y_i$</th>
<th>$\hat{\varepsilon}_i$</th>
<th>Household</th>
<th>$Y_i$</th>
<th>$\hat{\varepsilon}_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>49.98</td>
<td>-1.10</td>
<td>21</td>
<td>49.07</td>
<td>-2.01</td>
</tr>
<tr>
<td>2</td>
<td>55.33</td>
<td>4.25</td>
<td>22</td>
<td>48.91</td>
<td>-2.17</td>
</tr>
<tr>
<td>3</td>
<td>50.39</td>
<td>-0.69</td>
<td>23</td>
<td>58.06</td>
<td>6.98</td>
</tr>
<tr>
<td>4</td>
<td>46.83</td>
<td>-4.25</td>
<td>24</td>
<td>51.26</td>
<td>0.18</td>
</tr>
<tr>
<td>5</td>
<td>47.43</td>
<td>-3.65</td>
<td>25</td>
<td>51.52</td>
<td>0.44</td>
</tr>
<tr>
<td>6</td>
<td>48.47</td>
<td>-2.61</td>
<td>26</td>
<td>31.10</td>
<td>0.02</td>
</tr>
<tr>
<td>7</td>
<td>52.73</td>
<td>1.65</td>
<td>27</td>
<td>52.92</td>
<td>1.84</td>
</tr>
<tr>
<td>8</td>
<td>52.32</td>
<td>1.24</td>
<td>28</td>
<td>53.07</td>
<td>1.99</td>
</tr>
<tr>
<td>9</td>
<td>50.35</td>
<td>-0.73</td>
<td>29</td>
<td>49.35</td>
<td>-1.73</td>
</tr>
<tr>
<td>10</td>
<td>47.57</td>
<td>-3.51</td>
<td>30</td>
<td>54.85</td>
<td>3.77</td>
</tr>
<tr>
<td>11</td>
<td>49.52</td>
<td>-1.56</td>
<td>31</td>
<td>54.14</td>
<td>3.06</td>
</tr>
<tr>
<td>12</td>
<td>51.56</td>
<td>0.48</td>
<td>32</td>
<td>47.75</td>
<td>-3.33</td>
</tr>
<tr>
<td>13</td>
<td>55.17</td>
<td>4.09</td>
<td>33</td>
<td>50.88</td>
<td>-0.20</td>
</tr>
<tr>
<td>14</td>
<td>52.65</td>
<td>1.57</td>
<td>34</td>
<td>49.44</td>
<td>-1.64</td>
</tr>
<tr>
<td>15</td>
<td>51.43</td>
<td>0.35</td>
<td>35</td>
<td>53.02</td>
<td>1.94</td>
</tr>
<tr>
<td>16</td>
<td>45.96</td>
<td>-5.12</td>
<td>36</td>
<td>53.50</td>
<td>2.42</td>
</tr>
<tr>
<td>17</td>
<td>50.86</td>
<td>-0.22</td>
<td>37</td>
<td>49.65</td>
<td>-1.43</td>
</tr>
<tr>
<td>18</td>
<td>52.37</td>
<td>1.29</td>
<td>38</td>
<td>52.39</td>
<td>1.31</td>
</tr>
<tr>
<td>19</td>
<td>50.10</td>
<td>-0.98</td>
<td>39</td>
<td>49.81</td>
<td>-1.27</td>
</tr>
<tr>
<td>20</td>
<td>49.74</td>
<td>-1.34</td>
<td>40</td>
<td>51.85</td>
<td>0.77</td>
</tr>
</tbody>
</table>

Summary of Empirical Sample Estimates

\[ b = \$51.08 \]
\[ \hat{\sigma}^2 = 6.38 \]
\[ \text{vár}(b) = 0.159 \]
\[ s\hat{e}(b) = \$0.399 \]

So, we estimate the distribution of food expenditure to be

\[ Y_i \sim N(51.08, 6.38) \]

And we estimate the sampling distribution of \( b \) as

\[ b \sim N(51.08, 0.159) \]

Notes:

- We cannot directly state the accuracy of the above estimates

- All we can do is fall back on performance of estimators in repeated sampling
A Monte Carlo Sampling Experiment

As mentioned earlier in the lecture, sometimes it is difficult to gain an intuitive understanding of the sampling properties of an estimator.

Sampling experiments, or Monte Carlo simulations, can be very helpful in understanding sampling properties.

Use a computer to actually carry out a "thought experiment" for repeated samples.
FIGURE 18.1 A thought experiment helps explain the nature of the sampling distribution of $\bar{X}$. We start with a continuous random variable $X$ whose probability distribution is shown at the top of the figure. Next, we take $N$ different samples of size $n$, and in each sample we calculate the mean. These sample means are collected, and their frequency distribution is constructed. Theoretically, as the number of samples ($N$) increases to infinity, this frequency distribution reaches a limiting form. That limiting form is the sampling distribution of $\bar{X}$, which is the probability distribution that governs the likelihood of different values occurring for $\bar{X}$ in a sample of size $n$ drawn from the original $X$.

Basic Steps in a Monte Carlo Simulation

1. For the constant mean statistical model

\[ Y_t = \beta + e_t \quad t = 1, \ldots, T \]

select values for \( T, \beta \) and \( \sigma^2 \) (variance of \( e_t \))

2. Use a computer to draw a random sample of \( T \) error observations from a random variable with a mean of zero and variance of \( \sigma^2 \)

3. Add the selected value for \( \beta \) to each of the \( T \) error observations drawn in step (2) above. This generates \( T \) sample observations on \( y_t \).

4. Calculate mean and variance estimates for the \( T \) sample observations on \( y_t \)

5. Repeat the process in steps two, three and four a "large" number of times, say 1,000

6. Examine the 1,000 sample mean and variance estimates to see if they conform to the theoretically derived properties
Monte Carlo Simulation Example

The statistical model we will use for the simulation experiment is the same as before,

\[ Y_t = \beta + e_t \quad t = 1, \ldots, T \]

To create an artificial sample based on this model, we have to specify the sample size, mean and population variance.

Let's pick \( T=10 \), \( \beta=20 \), and \( \sigma^2=10 \), which implies the following,

\[ e_i \sim N(0,10) \]
\[ Y_t \sim N(20,10) \]
\[ b \sim N(20,1) \]

Generation of the artificial samples

- Computer packages contain random number generators that can be programmed to produce the artificial samples.

- Used to be a fairly complex programming problem, today such random number generators can be found in many packages (e.g. Excel, SAS).
### Table 3.3 One Artificial Sample from a $N(20, 10)$ Population

<table>
<thead>
<tr>
<th>$i$</th>
<th>$z_i$</th>
<th>$e_i$</th>
<th>$y_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.193</td>
<td>0.610</td>
<td>20.610</td>
</tr>
<tr>
<td>2</td>
<td>-0.504</td>
<td>-1.594</td>
<td>18.406</td>
</tr>
<tr>
<td>3</td>
<td>-2.318</td>
<td>-7.331</td>
<td>12.669</td>
</tr>
<tr>
<td>4</td>
<td>0.257</td>
<td>0.813</td>
<td>20.813</td>
</tr>
<tr>
<td>5</td>
<td>1.784</td>
<td>5.641</td>
<td>25.641</td>
</tr>
<tr>
<td>6</td>
<td>1.627</td>
<td>5.145</td>
<td>25.145</td>
</tr>
<tr>
<td>7</td>
<td>0.708</td>
<td>2.238</td>
<td>22.238</td>
</tr>
<tr>
<td>8</td>
<td>2.252</td>
<td>7.120</td>
<td>27.120</td>
</tr>
<tr>
<td>9</td>
<td>-0.462</td>
<td>-1.460</td>
<td>18.540</td>
</tr>
<tr>
<td>10</td>
<td>-1.348</td>
<td>-4.261</td>
<td>15.739</td>
</tr>
</tbody>
</table>

### Table 3.4 Least Squares Estimates of the Population Mean for 10 Samples of Size $T = 10$

<table>
<thead>
<tr>
<th>Sample</th>
<th>$b_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>20.692</td>
</tr>
<tr>
<td>2</td>
<td>21.326</td>
</tr>
<tr>
<td>3</td>
<td>20.318</td>
</tr>
<tr>
<td>4</td>
<td>20.717</td>
</tr>
<tr>
<td>5</td>
<td>22.043</td>
</tr>
<tr>
<td>6</td>
<td>19.586</td>
</tr>
<tr>
<td>7</td>
<td>19.651</td>
</tr>
<tr>
<td>8</td>
<td>21.501</td>
</tr>
<tr>
<td>9</td>
<td>19.575</td>
</tr>
<tr>
<td>10</td>
<td>19.304</td>
</tr>
</tbody>
</table>

Figure 3.7 Relative frequency histogram for standardized least squares estimates of the population mean.

Table 3.5 Estimates of $\beta$ and $\sigma^2$ for 10 Samples of Size $T = 10$

<table>
<thead>
<tr>
<th>Sample</th>
<th>$b_n$</th>
<th>$\hat{\sigma}^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>20.692</td>
<td>20.757</td>
</tr>
<tr>
<td>2</td>
<td>21.326</td>
<td>6.376</td>
</tr>
<tr>
<td>3</td>
<td>20.318</td>
<td>20.819</td>
</tr>
<tr>
<td>4</td>
<td>20.717</td>
<td>17.638</td>
</tr>
<tr>
<td>5</td>
<td>22.043</td>
<td>10.427</td>
</tr>
<tr>
<td>6</td>
<td>19.586</td>
<td>11.336</td>
</tr>
<tr>
<td>7</td>
<td>19.651</td>
<td>16.044</td>
</tr>
<tr>
<td>8</td>
<td>21.501</td>
<td>11.638</td>
</tr>
<tr>
<td>9</td>
<td>19.575</td>
<td>7.559</td>
</tr>
<tr>
<td>10</td>
<td>19.304</td>
<td>4.861</td>
</tr>
</tbody>
</table>