EE 381 Probability & Statistic with Applications to Computing (Fall 2020)

Lecture 10
Statistics: Sampling Theory

Duc H. Tran, PhD

Sampling Theory’s topics

- Sample Statistics
- Sampling distribution for the mean

Introduction about Statistics

- Statistics is a branch of science concerning data analysis.
- Statistics uses probability, just as physics uses calculus.
- Reason to learn: we use statistics to learn things about the world, to infer what can become known from data.

Example: John and David are running for elected office and we want to predict who will win. Asking each voter who they will vote for is impractical and expensive → ask a subset of voters (chosen randomly) → use that subset to predict the outcome of the election.

Statistical Inference

- *Population*: the entire of a large group.
- *Sample*: small part of the *population*.
- *Statistical inference*: a process of inferring certain facts about the *population* from results found in the *sample*.
- *Sampling*: a process of obtaining samples.

Example: We want to draw conclusions about the heights (or weights) of 12,000 adult students (the *population*) by examining only 100 students (a *sample*) selected from this population.

Example: We want to draw conclusions about the percentage of defective bolts produced in a factory during a given 6-day week by examining 20 bolts each day produced at various times during the day. Population: all bolts produced during the week. Sample: the 120 selected bolts.

Note:

❖ *Population size* can be finite or infinite (denoted by N), and *sample size* is finite (denoted by n).

Sampling With and Without Replacement

- *Sampling with replacement*: sampling where each member of a population maybe chosen more than once.
- *Sampling without replacement*: sampling where each member of a population cannot be chosen more than once.

Note:

- ❖A finite population that is sampled with replacement can theoretically be considered infinite.
- ❖Sampling from a finite population that is very large can be considered as sampling from an infinite population.

Random Samples and Random Numbers

Question: how to properly choose a sample to represent the population sufficiently well?

- One possible way: make sure that each member of the population has the **same chance** of being in the sample → called *random sample*.
- Random sampling can be accomplished for small populations by drawing lots or by using a table of *random numbers* (Appendix H).

Note: inference from sample to population cannot be certain → must use language of probability in statement of conclusions.

Population Parameters

- A population is considered to be known when we know the probability function *f(x)* of the associated random variable *X*.

Example: If *X* is a random variable whose values are the heights of the 12,000 students → *X* has a probability function *f(x)*.

If *X* is *normally distributed* → we have *normal population*.

If *X* is *binomial distributed* → we have *binomial population*.

- *Population parameters*: quantities that appear in *f(x)* such μ, σ, p (in binomial distribution), moments.
- In case that *f(x)* is not known precisely, we might wish to draw statistical inferences about its quantities.

Sample Statistics

- We can take random samples from the population, then use these samples to obtain values that estimate and test hypotheses about the *population parameters*.

Example: X is a random variable whose values are the heights of the 12,000 students. We want to obtain a sample of size 100.

First choose one individual at random from the population. Let x_1 is the value of X_1.

Similarly, for 2nd individual, we have x_2 is the value of X_2. Continue up to X_{100}.

- Assume that the sampling is with replacement, $X_1, X_2, \ldots, X_n$ would be independent, identically distributed random variables with *f(x)*. Their joint distribution would be:

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = f(x_1)f(x_2)\ldots\mathrm{f}(x_n)$$

- Any quantity obtained from a sample for the purpose of estimating a *population parameters* is called a *sample statistic*.

Sampling Distributions

- A *sample statistic* that is computed from $X_1, X_2, \ldots, X_n$ is a function of these random variables and is therefore a random variable itself.
- The probability distribution of a *sample statistic* is called the *sampling distribution* of the statistic.

Example: consider all possible samples of size *n* that can be drawn form the population, and we compute the statistic for each sample. In this manner, we obtain the distribution of the statistic, which is its sampling distribution.

- There are also *mean*, *variance*, *standard deviation*, *moments*, etc. for a sampling distribution.
- The *standard deviation* is also called *standard error*.

Sample Mean

- Let $X_1, X_2, \ldots, X_n$ denote the independent, identically distributed random variables for a random sample of size *n*. Then the *mean of the sample* or *sample mean* is a random variable defined by:

$$\bar{X} = \frac{X_1 + X_2 + \ldots + X_n}{n}$$

- If $x_1, x_2, \ldots, x_n$ denote values obtained in a particular sample size of *n*, then the *sample mean* is:

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n}$$

Example: If a sample of size 5 results in the sample values 7, 9, 1, 6, 2, then the sample mean is:

$$\bar{x} = \frac{7 + 9 + 1 + 6 + 2}{n} = 5$$

Sample Variance

- Let $X_1, X_2, \ldots, X_n$ denote the random variables for a random sample of size *n*. Then the *variance of the sample* or *sample variance* is defined by:

$$S^2 = \frac{(X_1-\bar{X})^2+(X_2-\bar{X})^2+\cdots+(X_n-\bar{X})^2}{n}$$

- The expected value of the sample variance is:

$$E(S^2) = \mu_{S^2} = \frac{n-1}{n}\sigma^2$$

which is very nearly σ^2 only for large n ($n \geq 30$).

The unbiased estimator is defined by:

$$\hat{S}^2 = \frac{n-1}{n}S^2 = \frac{(X_1-\bar{X})^2+(X_2-\bar{X})^2+\cdots+(X_n-\bar{X})^2}{n-1}$$

So that $$E\left(\hat{S}^2\right) = \sigma^2$$

Sample Variance

- **Example:** If a sample of size 5 results in the sample values 7, 9, 1, 6, 2, then the sample variance is:

$$S^2 = \frac{(7-5)^2 + (9-5)^2 + (1-5)^2 + (6-5)^2 + (2-5)^2}{5} = 9.2$$

An unbiased estimate will be:

$$\hat{S}^2 = \frac{5}{5-1} S^2 = 11.5$$

Sampling Distribution of Means

- Let f(x) be the probability distribution of some given population from which we draw a sample of size *n*. Then the probability function of the sample statistic $\bar{X}$ is called the *sampling distribution for the sample mean*, or the *sampling distribution of means*.

Theorem 5-1: The mean of the sampling distribution of means, denoted by $\mu_{\bar{X}}$, is given by:

$$E(\bar{X}) = \mu_{\bar{X}} = \mu$$

where μ is the mean of the population.

Theorem 5-2: If a population is infinite and the sampling is random or if the population is finite and sampling is with replacement, then the variance of the sampling distribution of means, denoted by$\sigma_{\bar{X}}^2$, is given by:

$$E[(\bar{X} - \mu)^2] = \sigma_{\bar{X}}^2 = \frac{\sigma^2}{n}$$

where σ^2 is the variance of the population.

Sampling Distribution of Means

Theorem 5-3: If the population is of size N, if sampling is without replacement, and if the sample size is $n \leq N$, then theorem 5-2's formula is replaced by:

$$\sigma_{\bar{X}}^2 = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$$

where $\mu_{\bar{X}}$ is still given by Theorem 5-1.

Note that theorem 5-3's formula reduces to theorem 5-2's formula as $N \to \infty$.

Theorem 5-4: If the population from which samples are taken is normally distributed with mean μ and variance σ^2, then the sample mean is normally distributed with mean μ and variance $\frac{\sigma^2}{n}$

Sampling Distribution of Means

Example:

A population consists of the five numbers 2, 3, 6, 8, 11. Consider all possible *samples of size two* which can be *drawn with replacement* from this population. Find:

a) Mean and standard deviation of the population

b) Mean and standard deviation of the sampling distribution of means.

Solution:

a)

$$\mu = \frac{2+3+6+8+11}{5} = 6$$

$$\sigma^2 = \frac{(2-6)^2+(3-6)^2+(6-6)^2+(8-6)^2+(11-6)^2}{5} = 10.8$$

So $\sigma = 3.29$

Sampling Distribution of Means

Example (ctn.):

b) Mean and standard deviation of the sampling distribution of means.

There are 5*5=25 samples of size two which can be drawn with replacement.

(2, 2)	(2, 3)	(2, 6)	(2, 8)	(2, 11)
(3, 2)	(3, 3)	(3, 6)	(3, 8)	(3, 11)
(6, 2)	(6, 3)	(6, 6)	(6, 8)	(6, 11)
(8, 2)	(8, 3)	(8, 6)	(8, 8)	(8, 11)
(11, 2)	(11, 3)	(11, 6)	(11, 8)	(11, 11)

The corresponding sample means are:

2.0	2.5	4.0	5.0	6.5
2.5	3.0	4.5	5.5	7.0
4.0	4.5	6.0	7.0	8.5
5.0	5.5	7.0	8.0	9.5
6.5	7.0	8.5	9.5	11.0

$\mu_{\bar{X}} = \frac{150}{25} = 6$ $\sigma_{\bar{X}}^2 = \frac{135}{25} = 5.4$ and $\sigma_{\bar{X}} = 2.32 = \frac{\sigma}{\sqrt{n}}$

This illustrate theorem 5-2, where $\sigma_{\bar{X}}^2 = \frac{\sigma^2}{n} = \frac{10.8}{2} = 5.4$

Sampling Distribution of Means

Example: What if now the sampling is without replacement. Recalculate those variables.

a) Mean and standard deviation of the population

Still the same.

$$\mu = 6, \qquad \sigma^2 = 10.8, \qquad \text{and } \sigma = 3.29$$

b) There are ${}^{5}_{2}C = 10$ sample of size two which can be drawn without replacement, which are:

(2,3) (2,6) (2,8) (2,11) (3,6) (3,8) (3,11) (6,8) (6,11) (8,11)

Note: (2, 3) now is considered the same as (3, 2).

The corresponding sample means are: 2.5 4 5 6.5 4.5 5.5 7 7 8.5 9.5

$$\mu_{\bar{X}} = \frac{2.5 + 4 + 5 + \cdots + 9.5}{10} = 6$$

$$\sigma_{\bar{X}}^2 = \frac{(2.5-6)^2 + (4-6)^2 + \cdots + (9.5-6)^2}{10} = 4.05 \quad \text{and} \quad \sigma_{\bar{X}} = 2.01$$

This illustrate theorem 5-3, where $\sigma_{\bar{X}}^2 = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right) = \frac{10.8}{2}\left(\frac{5-2}{5-1}\right) = 4.05$

Sampling Distribution of Means

Theorem 5-5: Suppose that the population from which samples are taken has a probability distribution with mean μ and variance σ^2 that is not necessarily a normal distribution. Then the standardized variable associated with $\bar{X}$, given by:

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

is asymptotically normal, i.e.,

$$\lim_{n \to \infty} P(Z \leq z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{u^2}{2}}\, du$$

This theorem is a consequence of the central limit theorem (assuming the population is infinite or that sampling is with replacement.

Otherwise, it is correct if we replace $\sigma/\sqrt{n}$ by $\sigma_{\bar{X}} = \sqrt{\frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)}$ (given in theorem 5-3).

Sampling Distribution of Means

Example:

A population of 29-year-olds males has a *mean salary* of \$29,321 with a *standard deviation* of \$2,120. If a sample of 100 men is taken, what is the probability that their *mean salaries* will be less than \$29,000?

Solution:

Standardizing: $z = \frac{29000-29321}{2120*\sqrt{100}} = -1.51$

Use standard normal distribution table, $P(z < -1.51) = 1 - 0.9345 = 0.0655$

Hence, the probability that the mean salaries of 100 randomly selected men is less than \$29,000 is 6.55%.

Sampling Distribution of Proportions

- Suppose that a population is infinite and binomially distributed, with *p* and *q=1*-*p* being the respective probabilities that any given member exhibits or does not exhibit a certain property.
- Consider all possible samples of size *n* drawn from this population, and for each sample determine the statistic that is the proportion *P* of successes. Then we obtain a sampling distribution of proportions whose:

$$\mu_p = p \qquad \sigma_p = \sqrt{\frac{pq}{n}} = \sqrt{\frac{p(1-p)}{n}}$$

- For large values of *n* ($n \geq 30$), the sampling distribution is very nearly a normal distribution (Theorem 5-5).
- For finite populations in which sampling is without replacement, σ_p will be replaced by $\sigma_P^2 = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$ with $\sigma = \sqrt{pq}$.

Sampling Distribution of Proportions

Example: It has been found that 2% of the tools produced by a certain machine are defective. What is the probability that in a shipment of 400 such tools, 3% or more will prove defective?

Solution:

$$\mu_p = p = 0.02 \text{ and } \sigma_p = \sqrt{\frac{pq}{n}} = \sqrt{\frac{0.02*0.98}{400}} = 0.007$$

Using the correction for discrete variables, *1/2n=1/800=0.00125*.

(0.003 – 0.00125) in standard units = *(0.03-0.00125-0.02)/0.007 = 1.25*

Required probability = (area under normal curve to right z=1.25) = **0.1056**.

If we had not used the correction, we would have obtained 0.0764.

Sampling Distribution of Differences and Sums

Suppose we are given 2 populations. We compute a statistic S_1 for each sample size n_1, this yields sampling distribution for S_1 with μ_{S_1} and σ_{S_1}.

Similarly, we have sampling distribution for S_2 with μ_{S_2} and σ_{S_2} for each sample size n_2.

- Sampling distribution of differences: $S_1 - S_2$. Its mean and variance are:

$$\mu_{S_1 - S_2} = \mu_{S_1} - \mu_{S_2} \qquad \sigma_{S_1 - S_2} = \sqrt{\sigma_{S_1}^2 + \sigma_{S_2}^2}$$

Its standardized variable:

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

- Sampling distribution of sums: $S_1 + S_2$. Its mean and variance are:

$$\mu_{S_1 + S_2} = \mu_{S_1} + \mu_{S_2} \qquad \sigma_{S_1 + S_2} = \sqrt{\sigma_{S_1}^2 + \sigma_{S_2}^2}$$

Sampling Distribution of Differences and Sums

Example: The electric light bulbs of manufacturer A have a mean lifetime of 1400 hours with a standard deviation of 200 hours, while those of manufacturer *B* have a mean lifetime of 1200 hours with a standard deviation of 100 hours. If random samples of 125 bulbs of each brand are tested, what is the probability that the brand *A* bulbs will have a mean lifetime that is at least 160 hours more than the brand B bulbs?

Solution:

Let $\bar{X}_A$ and $\bar{X}_B$ denote the mean lifetimes of samples A and B, respectively.

$$\mu_{\bar{X}_A - \bar{X}_B} = \mu_{\bar{X}_A} - \mu_{\bar{X}_A} = 1400 - 1200 = 200\ hours$$

And $$\sigma_{\bar{X}_A - \bar{X}_B} = \sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}} = \sqrt{\frac{100^2}{125} + \frac{200^2}{125}} = 20\ hours$$

The difference 160 hours in standard units = (160-200)/20 = -2

Required probability = (area under normal curve to right of z=-2)

= 0.5 + 0.4772 = **0.9772**

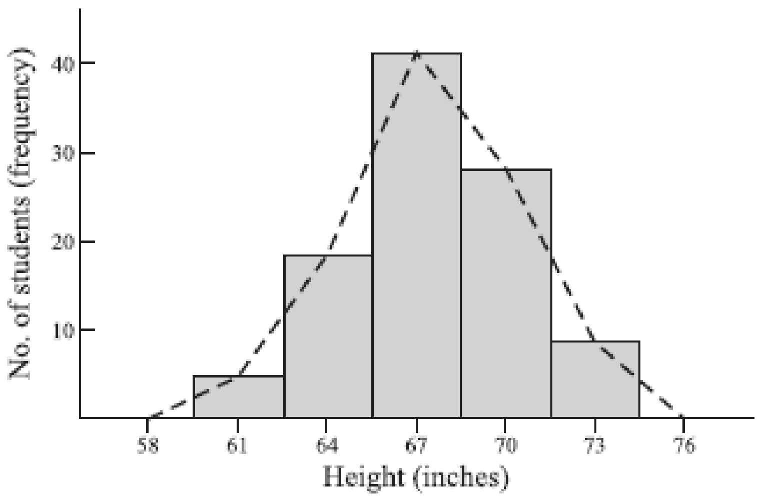
Frequency Distributions

It is useful to organize or group the *raw data* when a sample size is large and is difficult to observe the characteristics (or compute statistics).

We arrange data to *classes* or *categories*, and number of individuals belonging to each class is called *class frequency*. The resulting arrangement is called *frequency distribution*.

Heights of 100 Male Students at XYZ University

Height (inches)	Number of Students
60–62	5
63–65	18
66–68	42
69–71	27
72–74	8
TOTAL	100

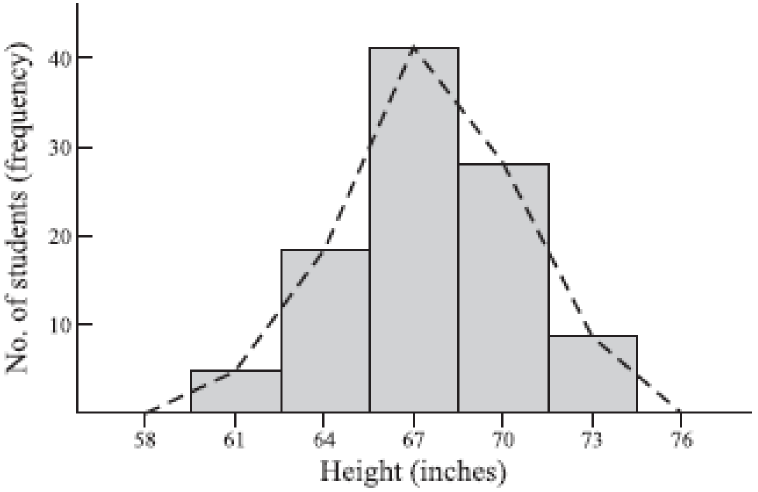


Frequency Distributions

- Heights from 60 to 62 inches: *class interval*. Height recorded as 60 inches is between 59.5 to 60.5, while one as 62 is between 61.5 to 62.5. Then 59.5 and 62.5 are called *class boundaries*.
- The width of the class interval, denoted by c_j, is the different between upper and lower class boundaries.
- The midpoint of the class interval is called the *class mark* (the representative of the class).

Heights of 100 Male Students at XYZ University

Height (inches)	Number of Students
60–62	5
63–65	18
66–68	42
69–71	27
72–74	8
TOTAL	100



Computation of Mean, Variance, and Moments for Grouped Data

Given the following table of class mark and class frequency:

Class Mark	Class Frequency
x_1	f_1
x_2	f_2
$\vdots$	$\vdots$
x_k	f_k
TOTAL	n

- The total frequency is: $n = f_1 + f_2 + \cdots + f_k = \sum f$
- There are f_1 numbers equal to $x_1, \ldots, f_k$ numbers equal to x_k. The mean is:

$$\bar{x} = \frac{f_1 x_1 + f_2 x_2 + \cdots + f_k x_k}{n} = \frac{\sum f x}{n}$$

- The variance is:

$$s^2 = \frac{f_1(x_1 - \bar{x})^2 + f_2(x_2 - \bar{x})^2 + \cdots + f_k(x_k - \bar{x})^2}{n} = \frac{\sum f(x - \bar{x})^2}{n}$$

Computation of Mean, Variance, and Moments for Grouped Data

Example: Find the mean and standard error of heights of 100 male students of XYZ University.

The mean height of the 100 male students:

$$\bar{x} = \frac{\sum fx}{n} = \frac{61 * 5 + 64 * 18 + 67 * 42 + 70 * 27 + 73 * 8}{100} = 67.45\ inches$$

The standard deviation of the heights of 100 male students:

$$s = \sqrt{\frac{\sum f(x-\bar{x})^2}{n}} = \sqrt{\frac{852.75}{100}} = 2.92\ inches$$

Height (inches)	Class Mark (x)	$x - \bar{x} = x - 67.45$	$(x - \bar{x})^2$	Frequency (f)	$f(x - \bar{x})^2$
60–62	61	−6.45	41.6025	5	208.0125
63–65	64	−3.45	11.9025	18	214.2450
66–68	67	−0.45	0.2025	42	8.5050
69–71	70	2.55	6.5025	27	175.5675
72–74	73	5.55	30.8025	8	246.4200
				$n = \sum f = 100$	$\sum f(x - \bar{x})^2 = 852.7500$

Computation of Mean, Variance, and Moments for Grouped Data

In cases where class intervals all have equal size *c* (the width of the class interval), we can use coding methods for computing the mean and variance quickly.

Assuming the transformation from the class mark *x* to a corresponding integer *u* given by:

$$x = a + cu$$

Where *a* is an arbitrary chosen class mark corresponding to *u=0*.

The coding formulas for the mean and variance are:

$$\bar{x} = a + \frac{c}{n}\sum fu = a + c\bar{u}$$

$$s^2 = c^2\left[\frac{\sum fu^2}{n} - \left(\frac{\sum fu}{n}\right)^2\right] = c^2(\overline{u^2} - \bar{u}^2)$$

Computation of Mean, Variance, and Moments for Grouped Data

Example: Using coding method , find the mean and standard deviation of heights of the 100 male students at XYZ University.

Mean:

$$\bar{x} = a + \left(\frac{\sum fu}{n}\right)c = 67 + \left(\frac{15}{100}\right)(3) = 67.45 \text{ inches}$$

Standard deviation:

$$s^2 = c^2\left[\frac{\sum fu^2}{n} - \left(\frac{\sum fu}{n}\right)^2\right] = c^2(\overline{u^2} - \bar{u}^2)$$

$$= (3)^2\left[\frac{97}{100} - \left(\frac{15}{100}\right)^2\right] = 8.5275$$

→ $s = 2.92$

	x	u	f	fu	fu^2
	61	−2	5	−10	20
	64	−1	18	−18	18
$a \rightarrow$	67	0	42	0	0
	70	1	27	27	27
	73	2	8	8	32
			$n = \sum f = 100$	$\sum fu = 15$	$\sum fu^2 = 97$

Computation of Mean, Variance, and Moments for Grouped Data

Similar formulas are available for higher moments.

The *rth* moments about the mean and the origin, respectively, are given by:

$$m_r = \frac{f_1(x_1 - \bar{x})^r + \cdots + f_k(x_k - \bar{x})^r}{n} = \frac{\sum f(x - \bar{x})^r}{n}$$

$$m_r' = \frac{f_1(x_1)^r + \cdots + f_k(x_k)^r}{n} = \frac{\sum f(x)^r}{n}$$

Computation of Mean, Variance, and Moments for Grouped Data

Example: Given the table of heights of 100 students.

a) Using random numbers, select 30 random samples of 4 students each (with replacement).

b) Find the mean and standard deviation of the sampling distribution of mean in (a)

c) Compare the results of (b) with theoretical values.

Heights of 100 Male Students at XYZ University

Height (inches)	Number of Students
60–62	5
63–65	18
66–68	42
69–71	27
72–74	8
TOTAL	100

Computation of Mean, Variance, and Moments for Grouped Data

Example (ctn.):

a) Use two digits to number each of the 100 students: 00, 01, 02, …99. Each student number is called a sampling number.

Height (inches)	Frequency	Sampling Number
60–62	5	00–04
63–65	18	05–22
66–68	42	23–64
69–71	27	65–91
72–74	8	92–99

The first line of random number table, we find the sequence 51, 77, 27, 46, 40, etc., which we take as random sampling numbers, each of which yields the height of a particular student.

For instance, 51 corresponds to a student having height 66-68 inches, which we take as 67 inches (the *class mark*).

Computation of Mean, Variance, and Moments for Grouped Data

Example (ctn.):

a) Note that we could start anywhere on the random number table.

	Sampling Numbers Drawn	Corresponding Heights	Mean Height
1.	51, 77, 27, 46	67, 70, 67, 67	67.75
2.	40, 42, 33, 12	67, 67, 67, 64	66.25
3.	90, 44, 46, 62	70, 67, 67, 67	67.75
4.	16, 28, 98, 93	64, 67, 73, 73	69.25
5.	58, 20, 41, 86	67, 64, 67, 70	67.00
6.	19, 64, 08, 70	64, 67, 64, 70	66.25
7.	56, 24, 03, 32	67, 67, 61, 67	65.50
8.	34, 91, 83, 58	67, 70, 70, 67	68.50
9.	70, 65, 68, 21	70, 70, 70, 64	68.50
10.	96, 02, 13, 87	73, 61, 64, 70	67.00
11.	76, 10, 51, 08	70, 64, 67, 64	66.25
12.	63, 97, 45, 39	67, 73, 67, 67	68.50
13.	05, 81, 45, 93	64, 70, 67, 73	68.50
14.	96, 01, 73, 52	73, 61, 70, 67	67.75
15.	07, 82, 54, 24	64, 70, 67, 67	67.00

	Sampling Numbers Drawn	Corresponding Heights	Mean Height
16.	11, 64, 55, 58	64, 67, 67, 67	66.25
17.	70, 56, 97, 43	70, 67, 73, 67	69.25
18.	74, 28, 93, 50	70, 67, 73, 67	69.25
19.	79, 42, 71, 30	70, 67, 70, 67	68.50
20.	58, 60, 21, 33	67, 67, 64, 67	66.25
21.	75, 79, 74, 54	70, 70, 70, 67	69.25
22.	06, 31, 04, 18	64, 67, 61, 64	64.00
23.	67, 07, 12, 97	70, 64, 64, 73	67.75
24.	31, 71, 69, 88	67, 70, 70, 70	69.25
25.	11, 64, 21, 87	64, 67, 64, 70	66.25
26.	03, 58, 57, 93	61, 67, 67, 73	67.00
27.	53, 81, 93, 88	67, 70, 73, 70	70.00
28.	23, 22, 96, 79	67, 64, 73, 70	68.50
29.	98, 56, 59, 36	73, 67, 67, 67	68.50
30.	08, 15, 08, 84	64, 64, 64, 70	65.50

Computation of Mean, Variance, and Moments for Grouped Data

b) Now let create the frequency distribution of sample mean heights table.

Sample Mean	Tally	f	u	fu	fu^2
64.00	/	1	−4	−4	16
64.75		0	−3	0	0
65.50	//	2	−2	−4	8
66.25	~~////~~ /	6	−1	−6	6
$a \rightarrow$ 67.00	////	4	0	0	0
67.75	////	4	1	4	4
68.50	~~////~~ //	7	2	14	28
69.25	~~////~~	5	3	15	45
70.00	/	1	4	4	16
		$\sum f = n = 30$		$\sum fu = 23$	$\sum fu^2 = 123$

The mean and variance are obtained using the coding methods:

$$\text{Mean} = a + c\bar{u} = a + \frac{c\sum fu}{n} = 67.00 + \frac{(0.75)(23)}{30} = 67.58 \text{ inches}$$

$$\text{Standard deviation} = c\sqrt{\overline{u^2} - \bar{u}^2} = c\sqrt{\frac{\sum fu^2}{n} - \left(\frac{\sum fu}{n}\right)^2}$$

$$= (0.75)\sqrt{\frac{123}{30} - \left(\frac{23}{30}\right)^2} = 1.41 \text{ inches}$$

Computation of Mean, Variance, and Moments for Grouped Data

c)

- The theoretical mean and standard deviation of the sampling distribution of means are:
  - ❖The mean: $\mu_{\bar{X}} = \mu = 67.45\ inches$
  - ❖The standard deviation :

$$\sigma_{\bar{X}} = \frac{population\ standard\ deviation}{\sqrt{n}} = \frac{\sigma}{\sqrt{n}} = \frac{2.92}{\sqrt{4}} = 1.46\ inches$$

- The sampling distribution of means of 30 random samples of 4 student each
  - ❖The mean: 67.58 inches
  - ❖The standard deviation: 1.41 inches

We have close agreement with these values of mean and standard deviation.

Discrepancies are due to the fact that only 30 samples were selected, and the sample size was small.

Reference

Notes, equations, and figures in the lecture are based on or taken from materials in the course textbook:

“Probability and Statistics”, by Spiegel, Schiller and Srinivasan, ISBN 987-007-179557-9 (McGraw-Hill/Schaun’s)