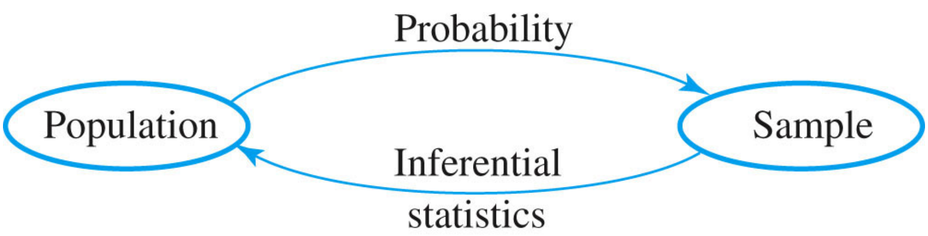
Sample Distribution

Today's Class

- Sample Distribution
- Central Limit Theorem

Statistics and Their Distributions

- A **statistic** is any quantity whose value can be calculated from sample data
- Prior to obtaining data, there is uncertainty as to what value of any particular statistic will result
- Therefore, a statistic is a random variable and will be denoted by an uppercase letter; a lowercase letter is used to represent the calculated or observed value of the statistic



Random Samples

- The probability distribution of any particular statistic depends not only on the population distribution and the sample size *n* but also on the method of sampling
- Frequently, we make the simplifying assumption that our data constitute a random sample $X_1, X_2, \ldots, X_n$ from a distribution. This means that
  - The X is are independent
  - All the X is have the same probability distribution

Deriving a Sampling Distribution

- Given such a sample, we can evaluate a statistic of interest
- The values of the statistic calculated from these samples allow us to examine the distribution of the statistic
- By changing settings, we can examine how the distribution of the statistic changes

Example

- A large automobile service center charges \$40, \$45, and \$50 for a tune-up of four-, six, and eight-cylinder cars, respectively. If 20% of its tune-ups are done on four-cylinder cars, 30% on six-cylinder cars, and 50% on eight-cylinder cars, then the probability distribution of revenue from a single randomly selected tune-up is given by

x	40	45	50
P(x)	0.2	0.3	0.5

- Suppose on a particular day, only two servicing jobs involve tune-ups. Let X_1 = the revenue from the first tune-up and X_2 = the revenue from the second

Example, cont'd

x_1	x_2	$p(x_1, x_2)$	$\bar{x}$	s^2
40	40	.04	40	0
40	45	.06	42.5	12.5
40	50	.10	45	50
45	40	.06	42.5	12.5
45	45	.09	45	0
45	50	.15	47.5	12.5
50	40	.10	45	50
50	45	.15	47.5	12.5
50	50	.25	50	0

Example cont'd

- To obtain the probability distribution of $\bar{X}$, look at the column:

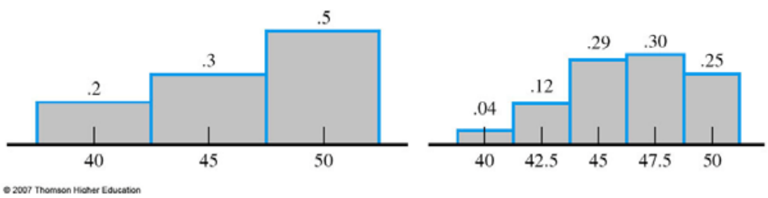
$\bar{X}$	40	42.5	45	47.5	50
$p_{\bar{x}}(\bar{x})$	0.04	0.12	0.29	0.30	0.25

- To obtain the probability distribution of S^2, need to look at the column:

S^2	0	12.5	50
$P_{s^2}(s^2)$	0.38	0.42	0.20

Example cont'd

- Spread appears to be reduced
  - Can calculate using expected values



$$\mu_X = E[X] = \sum x \cdot p_x(x) = 46.5 \quad \mu_X = E[X] = \sum \bar{x} \cdot p_{\bar{x}}(\bar{x}) = 46.5$$

$$\sigma_X^2 = V[X] = E[X^2] - \mu_X^2 = 15.25 \quad \sigma_{\bar{X}}^2 = V[\bar{X}] = E[\bar{X}^2] - \mu_{\bar{X}}^2 = 7.625$$

Example cont'd

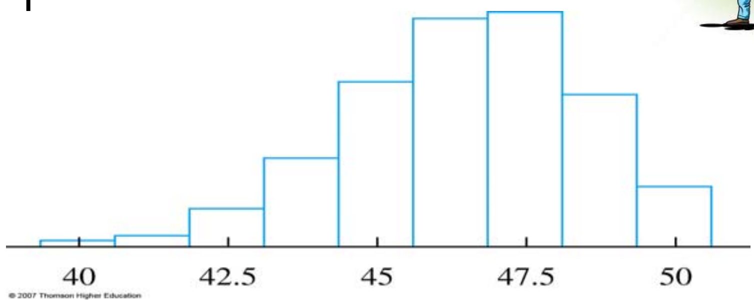
- Consider when n = 4

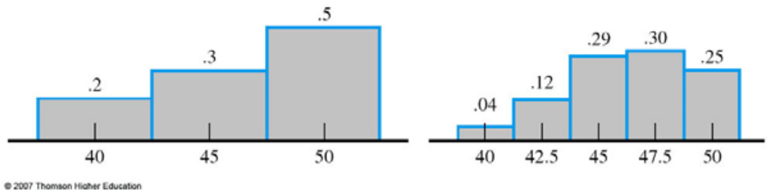
$\bar{X}$	40	41.25	42.5	43.75	45	46.25	47.5	48.75	50
$p_{\bar{X}}(\bar{x})$	0.0016	0.0096	0.376	0.936	0.1761	0.2340	0.2350	0.1500	0.0625

$$\mu_{\bar{X}} = E[\bar{X}] = \sum \bar{x} \cdot p_{\bar{X}}(\bar{x}) = 46.5$$

$$\sigma_{\bar{X}}^2 = V[\bar{X}] = E[\bar{X}^2] - \mu_{\bar{X}}^2 = 3.8125$$

Example, cont'd





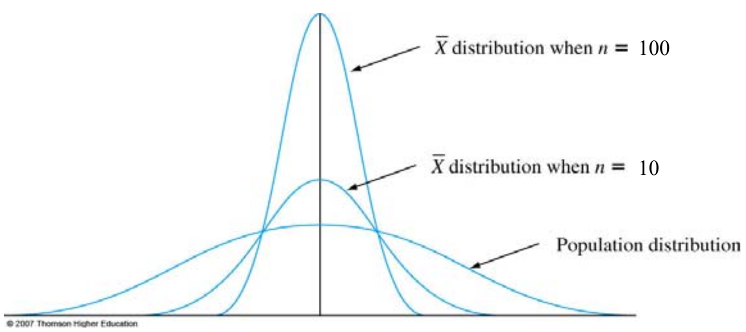
Distribution of the Sample Mean

- Let X_1, X_2, .. X_n be a random sample from a distribution with mean value μ and standard deviation σ. Then the expected value of sample mean is

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

and the variance of sample mean is

$$V(\bar{X}) = \sigma^2_{\bar{X}} = \frac{\sigma^2_X}{n}$$



Example 5.24

- In a notched tensile fatigue test on a titanium specimen, the expected number of cycles to first acoustic emission (used to indicate crack initiation) is μ = 28,000, and the standard deviation of the number of cycles is σ = 5000. Let $X_1, X_2, \ldots, X_{25}$ be a random sample of size 25, where each X_i is the number of cycles on a different randomly selected specimen.
  - What are the expected value and STD of the sample mean number of cycles until first emission?

Solution

- The expected value of the sample mean number of cycles until first emission:

$$E(\bar{X}) = \mu$$
$$= 28{,}000$$

- The standard deviation of the sample mean :

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$
$$= \frac{5000}{\sqrt{25}}$$
$$= 1000$$

Example 5.24

- In a notched tensile fatigue test on a titanium specimen, the expected number of cycles to first acoustic emission (used to indicate crack initiation) is $\mu = 28{,}000$, and the standard deviation of the number of cycles is $\sigma = 5000$. Let $X_1, X_2, \ldots, X_{25}$ be a random sample of size 25, where each X_i is the number of cycles on a different randomly selected specimen.
  - What are the expected value and STD when n=100?

Solution

- The expected value of the sample mean number of cycles until first emission:

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

$$= 28{,}000$$

- The standard deviation of the sample mean :

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

$$= \frac{5000}{\sqrt{100}}$$

$$= 500$$

Example: Sample Mean

- The height of men can be represented by a normal distribution with mean 72 inches and standard deviation 2 inches. Suppose you measure the height of 10 randomly chosen men.
  - What is the probability that their mean height will be greater than 71 inches?

Example Solution

So, $\bar{X} \sim N(72, 0.63^2)$

$$P(\bar{X} > 71)$$
$$= 1 - F\left(\frac{71 - 72}{0.63}\right)$$
$$= 1 - F(-1.58) \quad \textit{from } \underline{\textit{Table}}$$
$$= 1 - 0.0571$$
$$= 0.9429$$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	$\boxed{.0571}$	.0559

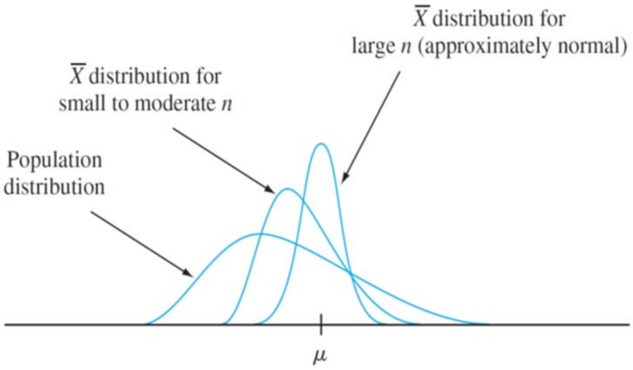
Central Limit Theorem (CLT)

- Let $X_1, X_2, .. X_n$ be a random sample from a distribution with mean value μ and variance σ^2
- Then if n is sufficiently large,

$$\overline{X} \sim N(\mu, \sigma^2/n)$$

- The larger the value of n, the better the approximation (n>30)

CLT



CLT Example

- Let X denote the number of flaws in copper wire. The pmf is as follows:

X	0	1	2	3
P(X=x)	0.48	0.39	0.12	0.01

  - If one hundred wires are sampled from this population, what is the probability that the average number of flaws per wire in this sample is less than 0.5?

Solution

X	0	1	2	3
P(X=x)	0.48	0.39	0.12	0.01

$E(X) = \mu$

$= 0 \times 0.48 + 1 \times 0.39 + 2 \times 0.12 + 3 \times 0.01$

$= 0.66$

$V(X) = \sigma^2 = E(X^2) - [E(X)]^2$

$= 0^2 \times 0.48 + 1^2 \times 0.39 + 2^2 \times 0.12 + 3^2 \times 0.01 - 0.66^2$

$= 0.5244$

Solution

$\mathrm{N} = 100, \overline{\mathrm{X}} \sim \mathrm{N}(0.66, \frac{0.5244}{100})$

$\mathrm{P}(\overline{\mathrm{X}} < 0.5)$

$= \mathrm{F}\left(\frac{0.5 - 0.66}{\sqrt{0.005244}}\right)$

$= \mathrm{F}(-2.21)$ from Table

$= 0.0136$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	$\boxed{.0136}$	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559

Example 5.26

- The amount of a particular impurity in a batch of a certain chemical product is a random variable with mean value 4.0 g and standard deviation 1.5 g.
  - If 50 batches are independently prepared, what is the (approximate) probability that the sample average amount of impurity $\overline{X}$ is between 3.5 and 3.8 g?

Solution

$$N = 50, \overline{X} \sim N(4.0, (\tfrac{1.5}{\sqrt{50}})^2) \therefore \overline{X} \sim N(4.0, 0.2121^2)$$

$$P(3.5 < \overline{X} < 3.8)$$

$$P\left(\frac{3.5-4}{0.2121} < Z < \frac{3.8-4}{0.2121}\right)$$

$$= \Phi\left(\frac{3.8-4}{0.2121}\right) - \Phi\left(\frac{3.5-4}{0.2121}\right)$$

$= \Phi(-0.94) - \Phi(-2.36)$ from Table

$$= .1736 - .0091$$

$$= .1645$$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	$\boxed{.0091}$	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−0.9	.1841	.1814	.1788	.1762	$\boxed{.1736}$	.1711	.1685	.1660	.1635	.1611