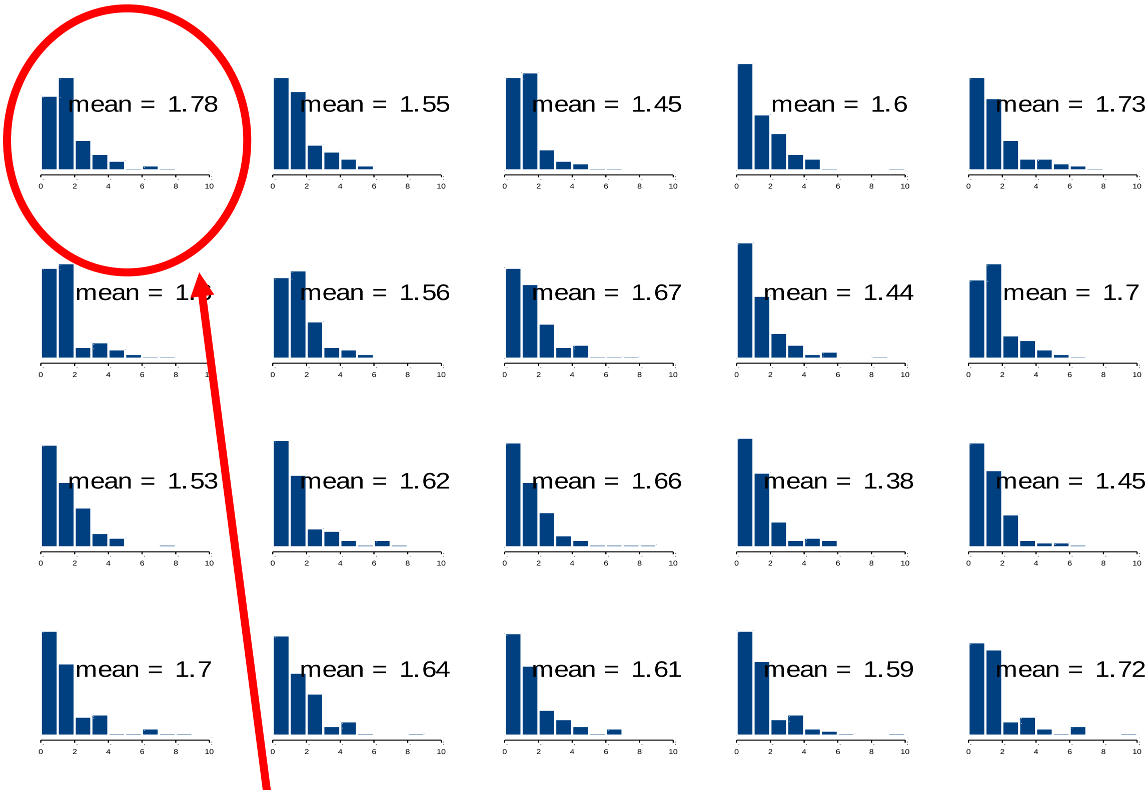
Exercise

1. If $X \sim \mathrm{N}(65,28)$ and $Y \sim \mathrm{N}(85,36)$ are 2 independent r.v, Find (a) $P(X+Y \leq 142)$
   (b) $P(134 \leq X+Y \leq 166)$ (c) $P(Y-X>4)$ (d) $P(12 \leq Y-X \leq 24)$
2. Each day Mr. Njoroge walks to the library bto read a newspaper. Total time spent walking is normally distributed with mean 15 minutes and standard deviation 2 minutes. Total time spent in the library is also normally distributed with mean 25 minutes and standard deviation $\sqrt{12}$ minutes. Find the probability that on one day;
   a) he is away from his home for more than 45 minutes.
   b) he spends more time walking than in the library

4.6 Sampling Distributions

In many investigations the data of interest can take on many possible values and it is often of interest to estimate the population mean, μ. A common estimator for μ is the sample mean $\bar{x}$. Consider the following set up: We observe a sample of size *n* from some population and compute the mean $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ .Since the particular individuals included in our sample are *random*, we would observe a different value of $\bar{x}$ if we repeated the procedure. That is, $\bar{x}$ *is also a random quantity*. Its value is determined partly by which people are randomly chosen to be in the sample. If we repeatedly drew samples of size *n* and calculated $\bar{x}$, we could ascertain the sampling distribution of $\bar{x}$.

Many possible samples, many possible $\bar{x}$'s



We only see one!

We will have a better idea of how good our one estimate is if we have good knowledge of how $\bar{x}$ behaves; that is, if we know the probability distribution of $\bar{x}$.

4.6.1 Properties of the Sampling Distribution of the Sample Means (Summary)

When all of the possible sample means are computed, then the following properties are true:

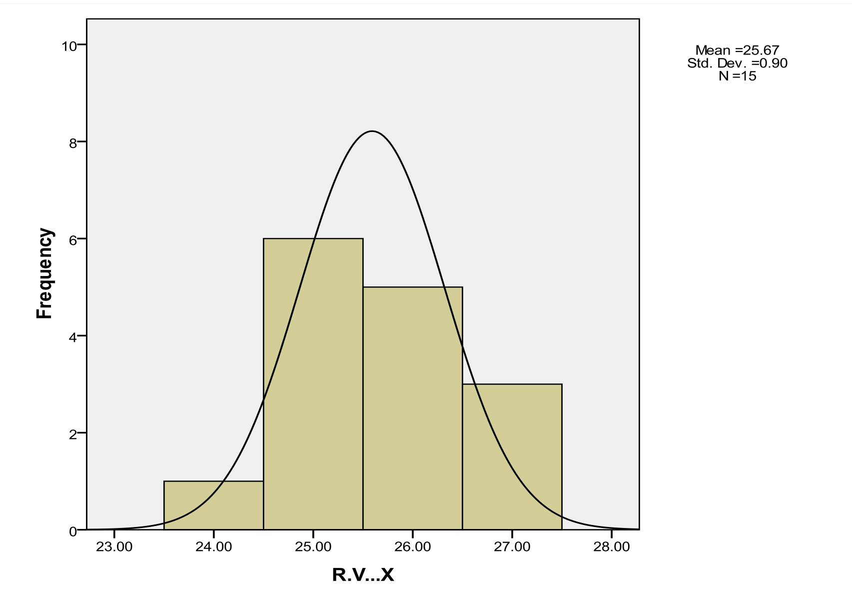
1. The mean of the sample means will be the mean of the population
2. The variance of the sample means will be the variance of the population divided by the sample size.
3. The standard deviation of the sample means (known as the standard error of the mean) will be smaller than the population mean and will be equal to the standard deviation of the population divided by the square root of the sample size.
4. If the population has a normal distribution, then the sample means will have a normal distribution.
5. If the population is not normally distributed, but the sample size is sufficiently large, then the sample means will have an approximately normal distribution. Some books define sufficiently large as at least 30 and others as at leas t25.

Example

The law firm of Hoya and Associates has six partners (A, B, C, D, E, F). At their weekly partners meeting each reported the number of hours they charged clients for their services last week. A 24, B 26, C 28, D 26, E 24. F26 (eg, Mr. E charged 24 hrs)
If n=2, (ie two partners) are selected randomly, how many different samples are possible?
This is the combination of 6 objects taken 2 at a time. Ie there are ${}^6c_2 = 15$ possible samples.
15 sample means are given below: (e.g. if the sample has A and B, sample mean is 24)
AB 25, AC 26, AD 25, AE 24, AF 25, BC 27, BD 26, BE 25, BF 26, CD 27, CE 26, CF 27, DE 25, DF 26, EF 25 putting this in a frequency table we get

$\bar{x}$	24	25	26	27
f	1	6	5	3

This is almost the sampling distribution of means. If we divide individual frequencies by total frequency (ie 15) we get "relative frequency" or probability. These probabilities add up to one, so we have a prob. distribution. The above information says that the probability that sample mean is 24 is 1 out of 15 or 0.066667. Now draw a histogram for sampling distribution of means and fit a normal curve on it.



Note the shape is similar to Normal distribution

The sampling distribution is simply this probability distribution defined over all possible samples of size n from the population of size N. In the real world problems N will be large (e.g. 200 million US population) and n will be also be large (e.g., 1000 people surveyed) and ${}^{N}c_n$ will be astronomical number. Then the sampling distribution can only be imagined. We have chosen a simple example of N=6, n=2 so that the entire sampling distribution can be explicitly computed and visualized. Now the random variable is $\bar{x}$, it is no longer just X.

Definitions
Central Limit Theorem:- Stats that as the sample size increases, the sampling distribution of the sample means will become approximately normally distributed.
Sampling Distribution of the Sample Means:- Distribution obtained by using the means computed from random samples of a specific size.
Sampling Error :- Difference which occurs between the sample statistic and the population parameter due to the fact that the sample isn't a perfect representation of the population.
Standard Error or the Mean:- The standard deviation of the sampling distribution of the sample means. It is equal to the standard deviation of the population divided by the square root of the sample size.

4.6.2 The Mean and Standard Deviation of $\bar{x}$

What are the mean and standard deviation of $\bar{x}$?
Let's be more specific about what we mean by a sample of size *n*. We consider the sample to be a collection of *n independent and identically distributed (*or *iid)*random variables $X_1, X_2, ..., X_n$ with common mean μ and common standard deviation σ.

Thus, $E(\overline{X}) = E\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(x_i) = \frac{1}{n}\sum_{i=1}^{n}\mu = \frac{1}{n}(n\mu) = \mu$

$$Var(\overline{X}) = Var\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} Var(x_i) = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{1}{n^2}(n\sigma^2) = \frac{\sigma^2}{n} \quad \Rightarrow SD(\overline{X}) = \frac{\sigma}{\sqrt{n}}$$

4.6.3 The Central Limit Theorem

Now we know that $\bar{x}$ has mean μ and standard deviation $\sigma/\sqrt{n}$, but what is its distribution?
If $X_1, X_2, ..., X_n$ are *normally distributed,* then $\bar{x}$ is also normally distributed. Thus, $X_i \sim N(\mu, \sigma^2) \;\Rightarrow\; \overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$**.** If $X_1, X_2, ..., X_n$ are *not* normally distributed, then the Central Limit Theorem tells us that $\bar{x}$ is *approximately N*ormal.
In brief if $X_1, X_2, ..., X_n$ are *iid* random variables with mean μ and finite standard deviation σ. Then for a sufficiently large n, the sampling distribution of $\overline{X}$ is approximately Normal with mean μ and variance $\frac{\sigma^2}{n}$.

Remarks

- Central limit theorem involves two different distributions: the distribution of the original population and the distribution of the sample means
- The formula for a z-score when working with the sample means is: $z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$ ~ N(0,1)

Example

Intelligence Quotient (IQ) is normally distributed with mean 110 and standard deviation of 10. A moron is a person with IQ less than 80. Find the probability that a randomly chosen person is a moron. Let idiot be defined as one with an IQ less than 90. Find the probability that a randomly chosen person is an idiot. (Hint this random variable is for a single person X) If a sample of 25 students is available, what is the probability that the **average** IQ exceeds 105? What is the probability that the average IQ exceeds 115 (Hint this random variable is for an average over 25 persons or $\overline{X}$)

Solution

$IQ = X \sim \mathrm{N}\left(110_1, 10^2\right)$, and therefore for a sample of 25 people average $IQ = \overline{X} \sim \mathrm{N}\left(110_1, 4\right)$

The probability that a randomly chosen person is a moron is given by

$P(X < 80) = P\left(Z < \frac{80-110}{10}\right) = \Phi(-3) = 0.0013$

The probability that a randomly chosen person is an idiot is given by

$P(X < 90) = P\left(Z < \frac{90-110}{10}\right) = \Phi(-2) = 0.0228$

The probability that the **average** IQ exceeds 105 is $P(\overline{X} > 105)$

The random variable under consideration here is the average. Hence, a sampling distribution is relevant when we consider average IQ as the variable of interest, not he IQ of an individual student, but the average over 25 students. Standard deviation of the sampling distribution = Standard Error SE = $\sigma/\sqrt{\mathrm{n}} = 10/5 = 2$. Now

$P(\overline{X} > 105) = P(Z > \frac{105-110}{2}) = \underbrace{P(Z \leq 2.5)}_{due\,to\,symmetry} = \Phi(2.5) = 0.9938$

We now find probability that the average IQ exceeds 115 ie

$P(\overline{X} > 115) = P(Z > \frac{115-110}{2}) = \underbrace{P(Z \leq -2.5)}_{due\,to\,symmetry} = \Phi(-2.5) = 0.0062$

Exercise

1) The annual salaries of employees in a large company are approximately normally distributed with a mean of \$50,000 and a standard deviation of \$10,000. If a random sample of 50 employees is taken, what is the probability that their average salary is;
   a) less than \$45,000?
   b) between \$45,000 and \$65,000?
   c) more than \$70,000
2) Library usually has 13% of its books checked out. Find the probability that in a sample of 588 books greater than 14% are checked out. ANS= 0.2358
3) The length of similar components produced by a company are approximated by a normal distribution model with a mean of 5 cm and a standard deviation of 0.02 cm.
   a) If a component is chosen at random what is the probability that the length of this component is between 4.98 and 5.02 cm?
   b) what is the probability that the average length of of a sample of 25 component is between 4.96 and 5.04 cm?
4) The length of life of an instrument produced by a machine has a normal distribution with a mean of 12 months and standard deviation of 2 months. Find the probability that in a random sample of 4 instrument produced by this machine, the average length of life
   a) less than 10.5 months. b) between 11and 13 months.