The Central Limit Theorem

If the distribution of each observation in the population has mean μ and standard deviation σ *regardless of whether the distribution is normal or not*:

1. The distribution of the sample means $\overline{X}_n$ (from samples of size *n* taken from the population) has mean μ identical to that of the population
2. The standard deviation of this distribution is $\sigma/\sqrt{n}$.
3. As *n* gets large the shape of the sample distribution of the mean is approximately that of a *normal* distribution

The Central Limit Theorem (continued)

Example: Cholesterol level in U.S. males 20-74 years old

The serum cholesterol levels for all 20-74 year-old US males has a population mean μ=211mg/100 ml and standard deviation σ=46 mg/100 ml. Let's say that we take a sample size of size n = 25.

What is the probability that $\overline{X} \geq 217$ mg/100 ml?

What about $\overline{X} \geq 220$ mg/100 ml?

If $\overline{X} \geq 217$ mg/100 ml, from the Central Limit Theorem we have that

$$P(\overline{X} \geq 217) = P\left(\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \geq \frac{217 - 211}{9.2}\right) = P(Z \geq 0.65) = 0.258$$

Similarly $P(\overline{X} \geq 220) = P(Z \geq 0.98) = 0.164$.

Example (continued):

If we select samples of size n=25, what proportion of time will their means be over 230 mg/100 ml?

By the central limit theorem, the distribution of the sample means will have mean μ=211mg/100 ml, and std. deviation $\sigma_{\overline{x}} = 46/\sqrt{25} = 9.2$mg/100 ml. So the problem is

$$P(\overline{X}_{25} \geq 230) = P\left(\frac{\overline{X}_{25} - \mu}{\sigma_{\overline{x}}} \geq \frac{230 - \mu}{\sigma_{\overline{x}}}\right) = P\left(Z \geq \frac{230 - 211}{9.2}\right) = P(Z \geq 2.07)$$

Referring to Table A.3 in the textbook, this probability is 0.019. Thus, less than 2% of the time, these sample means expected to be larger than 230mg/100 ml.

Example (continued):

To calculate the upper and lower cutoff points enclosing 95% of the means of samples of size n=25 drawn from this population we do the following: The cutoff points in the standard normal distribution are –1.96 and +1.96. Through standardization we can translate it to a statement about serum cholesterol levels.

$$-1.96 \le Z \le 1.96 \Leftrightarrow -1.96 \le \frac{\overline{X}_{25} - 211}{9.2} \le 1.96$$

$$\Leftrightarrow 211 - 1.96(9.2) \le \overline{X}_{25} \le 211 + 1.96(9.2)$$

$$\Leftrightarrow 193.0 \le \overline{X}_{25} \le 229.0$$

Approximately 95% of the sample means will fall between 193 and 229 mg/100 ml. This is a general result, i.e., 95% of *any* normal distribution is contained within $\mu \pm 1.96\sigma$. Here, $\sigma = \sigma_{\overline{x}} = \sigma/\sqrt{n}$ and $\mu = \overline{x}\tilde{n}$