Part 1
Lecture 2a The Central Limit Theorem

1

Medical Imaging
UNIVERSITY OF TORONTO

MiDATA

Who I am…

Pascal Tyrrell, PhD

Associate Professor

Department of Medical Imaging, Faculty of Medicine

Institute of Medical Science, Faculty of Medicine

Department of Statistical Sciences, Faculty of Arts and Science

QUESTIONS ABOUT A STUDY COMPARING TWO BLOOD PRESSURE DRUGS

Let’s suppose your friend is excited about a study in which the mean decrease in blood pressure was greater in a group of patients on a new drug compared to a group on an old drug. Before you also get excited what questions would you ask her about the study?

THE SEVEN QUESTIONS

1. How long was the follow up period ?
2. Did the new drug have any serious side effects ?
3. How were patients allocated to the groups ?
4. How large is the mean DBP difference between groups?
5. How many patients were in each group ?
6. How large is the *Variation* in response among patients?
7. Are the two groups comparable ?

NOTE: DBP = Decrease in blood pressure.

CENTRAL LIMIT THEOREM

CANADIAN GOVERNMENT POPULATION HEALTH SURVEY OF 6,000 CANADIAN WOMEN

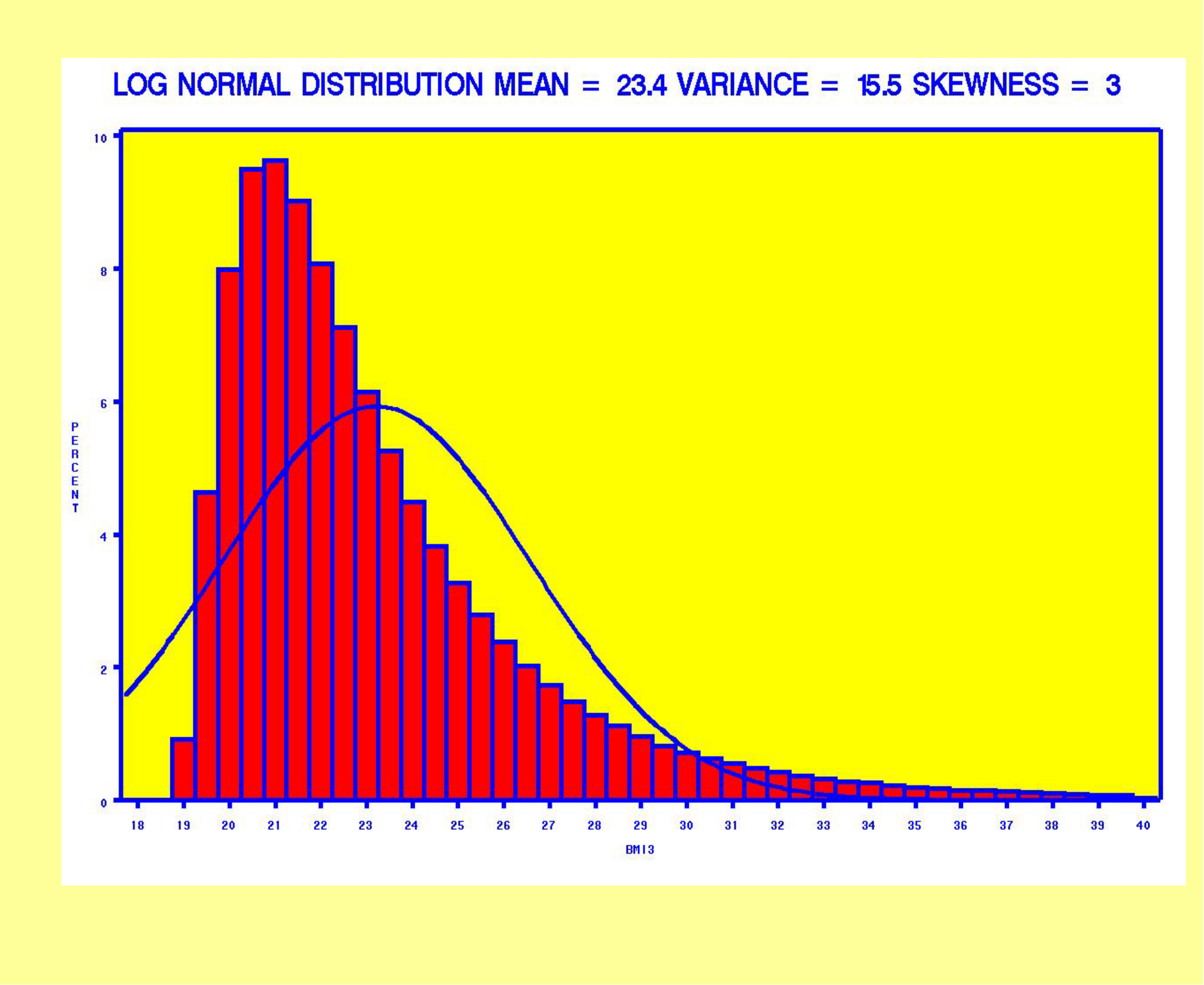
One measurement made was the Body Mass Index (BMI):

$$\text{BMI} = \text{WEIGHT(kg)} / \text{HEIGHT (m)}^2$$

CENTRAL LIMIT THEOREM

POPULATION OF 6,000 FEMALES

*NOTE THE POSITIVE SKEWNESS



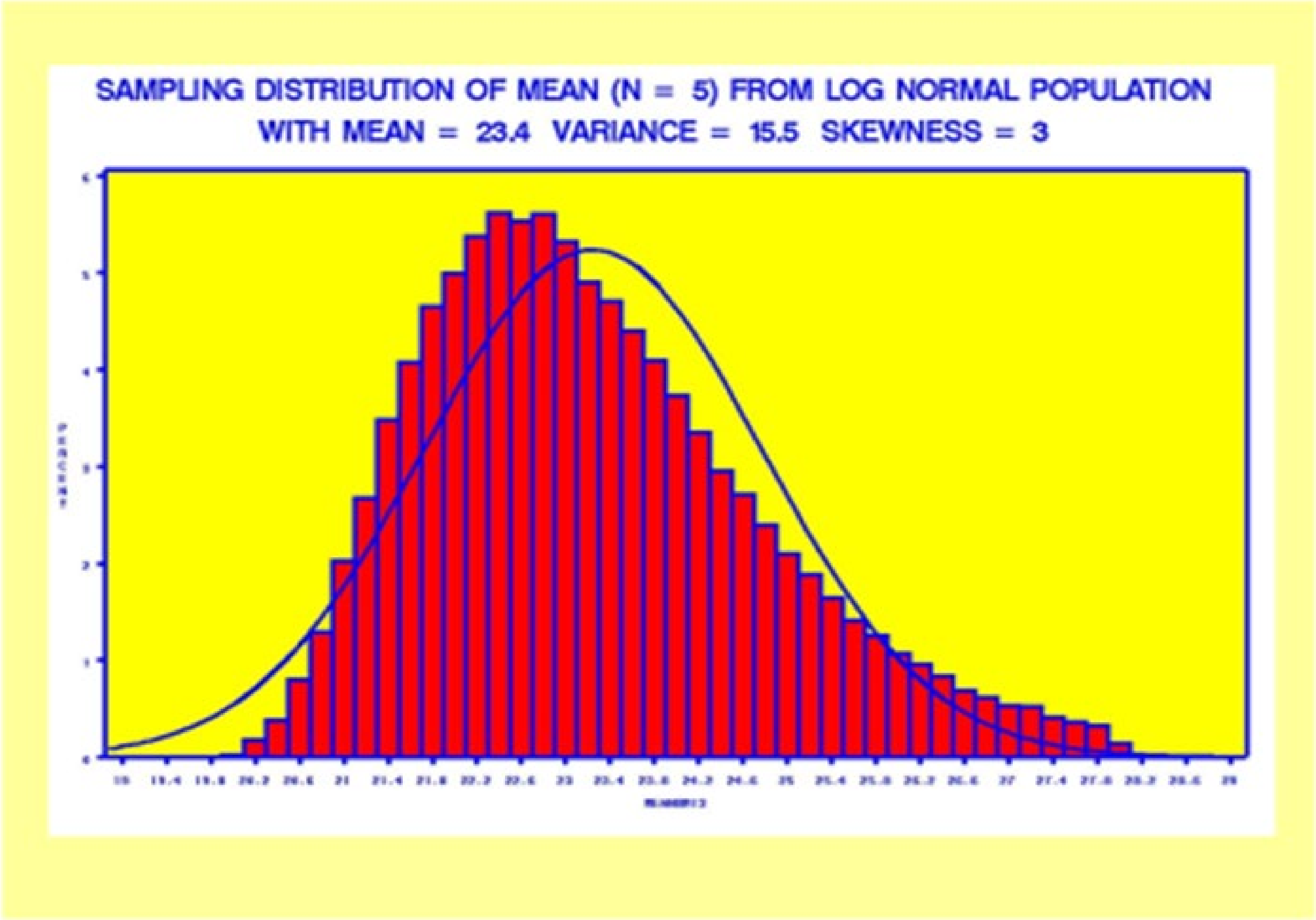
CENTRAL LIMIT THEOREM

Paul randomly selected a sample of 5 women from this database and calculated their mean BMI. He then placed these women back into the population and again randomly selected a sample of five women and calculated their mean BMI. He did this 1,000 times and produced the histogram of these 1,000 means. He did this 2 more times with each sample having 10 women and then again with each sample having 25 women.

DISTRIBUTION OF 1,000 SAMPLE MEANS WITH N = 5

SAMPLE OF 1,000 RANDOMLY SELECTED MEANS OF 5 FEMALES

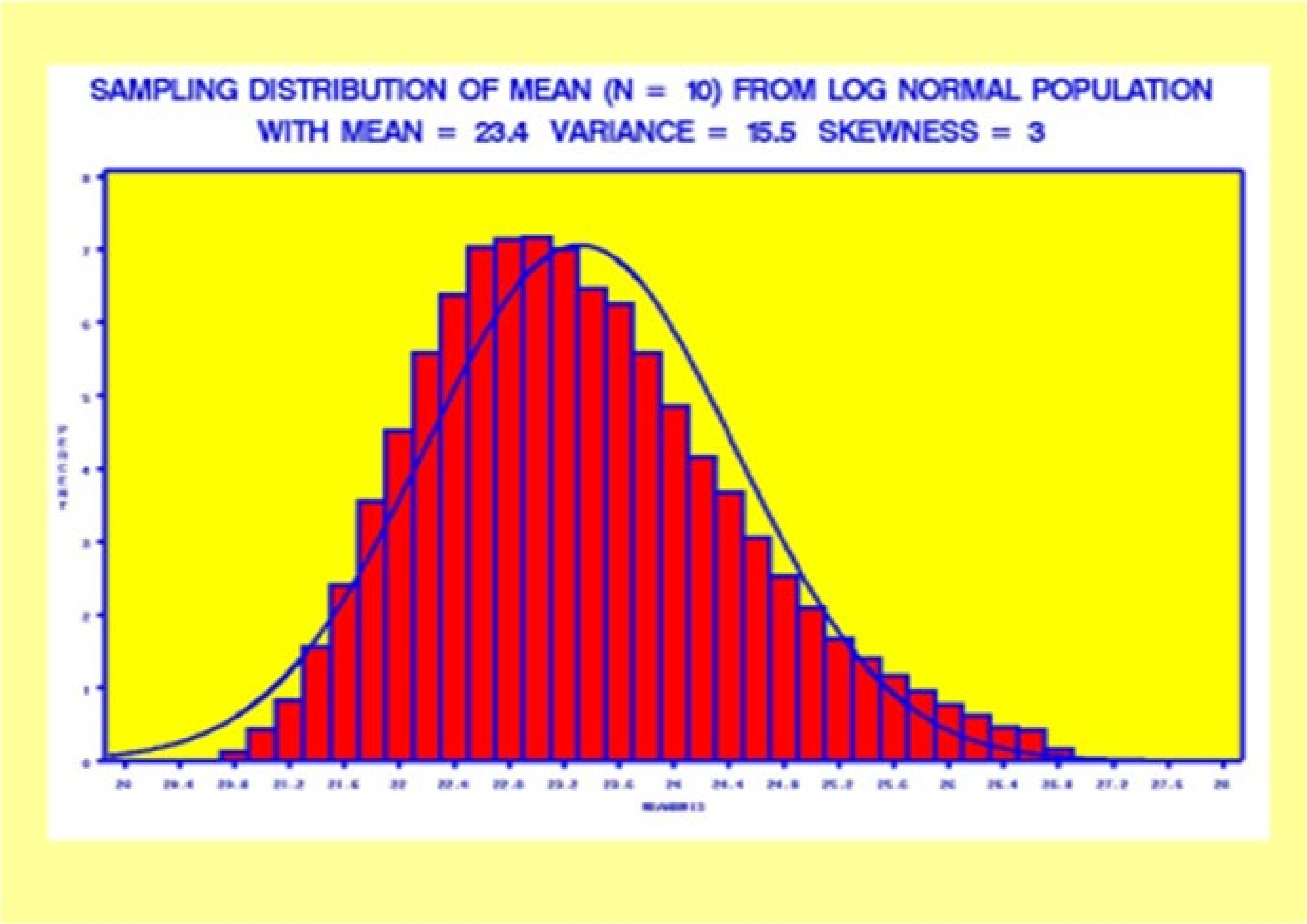
AND OVERLAPPING GAUSSIAN PROBABILITY DISTRIBUTION



DISTRIBUTION OF 1,000 MEANS N = 10

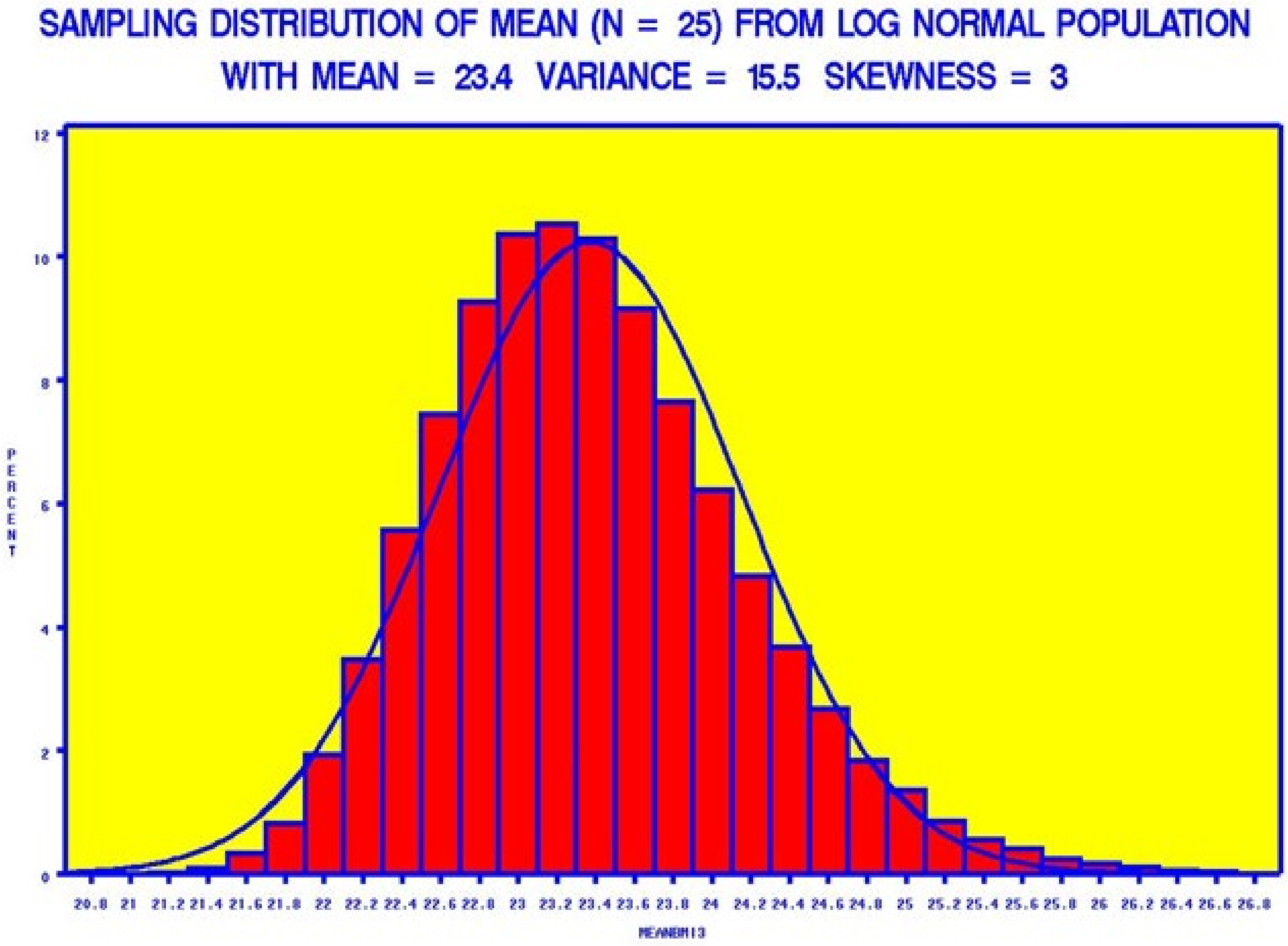
SAMPLE OF 1,000 RANDOMLY SELECTED MEANS OF 10 FEMALES

AND OVERLAPPING GAUSSIAN PROBABILITY DISTRIBUTION



SAMPLE OF 1,000 RANDOMLY SELECTED MEANS OF 25 FEMALES

AND OVERLAPPING GAUSSIAN PROBABILITY DISTRIBUTION



CENTRAL LIMIT THEOREM

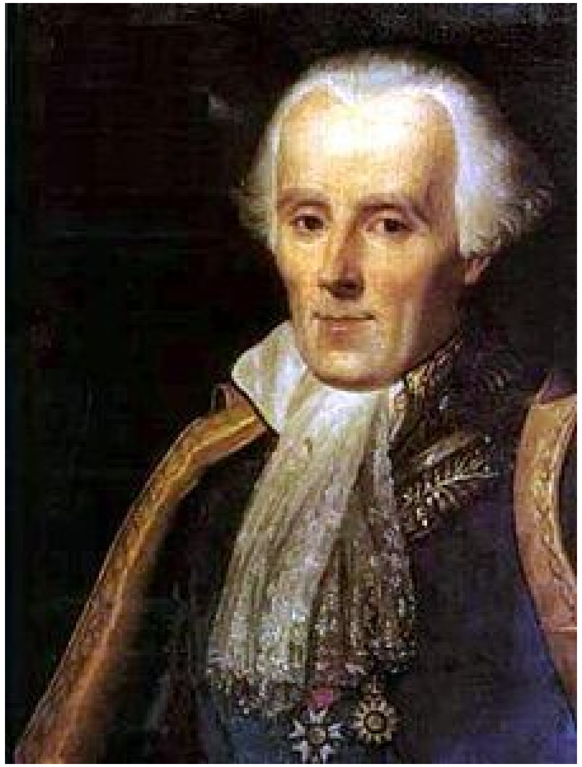
The probability distribution of a sample mean of N observed values randomly selected from a population approaches the Gaussian (Normal) probability distribution as N approaches infinity.

As “N approaches infinity” is mathematician talk. We saw that the distribution of sample means was approximately Gaussian for N as small as 25.

CENTRAL LIMIT THEOREM

Pierre Simon La Place

1749 – 1827



Carl Friedrich Gauss

1777 – 1855

GAUSSIAN PROBABILITY DENSITY FUNCTION

$$f\left(\overline{BMI}\right) = \frac{e^{-\frac{\left(\overline{BMI}-\mu\right)^2}{2\times\frac{\sigma^2}{n}}}}{\sqrt{2\pi \times \frac{\sigma^2}{n}}} \qquad with \quad \pi = 3.1416 \qquad e = 2.7163$$

$Sample\ mean\ \overline{BMI}\ is\ unbiased\ estimator\ of\ theoretical\ (population) mean\ \mu.$
$Many\ sample\ means\ each\ with\ many\ BMI\ values\ selected\ randomly\ from$
$population\ has\ a\ bell\ shaped\ Normal\ probability\ distribution\ with\ mean\ \mu$
$and\ variance\ \sigma^2.\ The\ parameter\ \sigma, standard\ deviation\ of\ variable\ BMI,$
$is\ an\ important\ measure\ of\ variation. The\ variance\ of\ the\ mean\ \overline{BMI}$
$variable\ is\ \frac{\sigma^2}{n}\ with\ standard\ error\ \sqrt{\frac{\sigma^2}{n}}$

Standard Gaussian Probability Density Function

$$If\ Z = \frac{\left(\overline{BMI_1} - \overline{BMI_2}\right) - (\mu_1 - \mu_2)}{\sqrt{2 \times \frac{\sigma^2}{n}}} \qquad then\ f(Z) = \frac{e^{-Z^2}}{\sqrt{2\pi}}$$

$and\ Probability\ (-1.96 < Z < 1.96) = 0.95$

$Mean\ of\ the\ standard\ Gaussian\ variable\ Z\ is\ 0.$

$Variance\ of\ Z\ is\ 1\ and\ Standard\ deviation\ of\ Z\ is\ 1.$

$\sqrt{2 \times \frac{\sigma^2}{n}}\ is\ the\ standard\ error\ of\ the\ difference\ of$

$two\ INDEPENDENT\ sample\ means.$