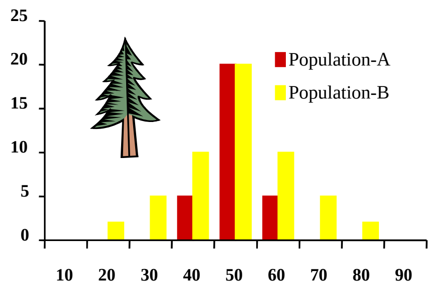
OUTLINE

Standard Measures of Dispersion
- Range
- Variance
- Standard deviation
- Coefficient of variation

The Normal Curve
- Normal curve areas
- Departures from normality
- Graphic methods of evaluation

Example: Larch Population

Pop-A & Pop-B Tree Height
Mean = 50, spread is 40-60 or 20-80



Standard Measures of Dispersion

<u>Range</u>

Range = Y_{max} - Y_{min}

Using Larch Tree Example:

$Range_{pop-a}$ = 60 - 40 = 20
$Range_{pop-b}$ = 80 - 20 = 60

Note: VERY sensitive to outliers!

Standard Measures of Dispersion

Variance

$$S^2 = \frac{\sum (Y_i - \bar{Y})^2}{n-1}$$

Takes in to account all values,
not just largest and smallest.

Computationally more intensive…

Important Concept

Sum of the squared deviations is known as:
SUM-OF-SQUARES (or SS)

SS / N is known as:
MEAN-SQUARES (or MS)

* Hold this thought for 5 weeks, this will become a central notion of ANOVA

Standard Measures of Dispersion

Variance

Simplified "machine formula":

$$S^2 = \frac{\sum Y_i^2 - \frac{(\sum Y_i)^2}{n}}{n-1}$$

Example: Variance Calculation

Y	**Y^2**
10	100
8	64
6	36
5	25
5	25
3	9
2	4
1	1
$\Sigma Y = 40$	$\Sigma Y^2 = 264$

n = 8

Using "machine formula":

$$S^2 = \frac{264 - (40)^2 / 8}{7}$$

$S^2 = 9.14$

Standard Measures of Dispersion

Standard Deviation

$$S = \sqrt{S^2}$$

S equals the square root of the variance.

From previous example, S = 3.0237

Standard Measures of Dispersion

Coefficient of Variation

"*It is well established that distributions with larger means have greater standard deviations*."

Q: How can we then compare the dispersion of one group with the dispersion of another?
A: Compare relative amounts of variation (0-100%).

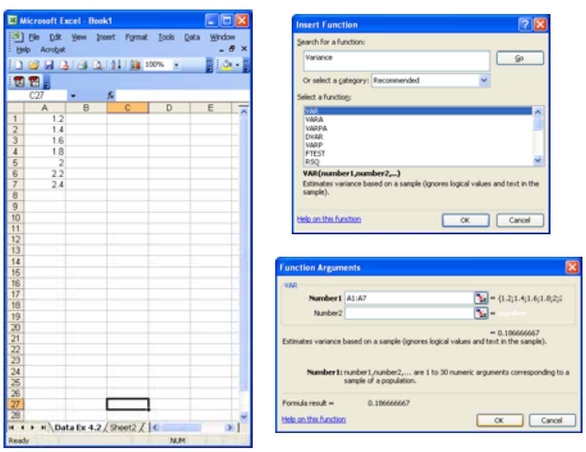
$$V = CV = \frac{S}{\bar{Y}}(100)$$

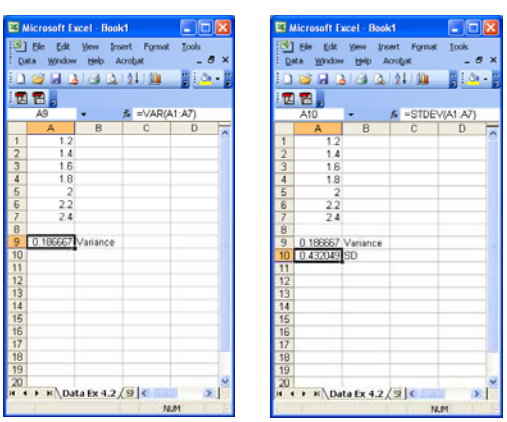
Example 4.2 (p.38) Using Excel

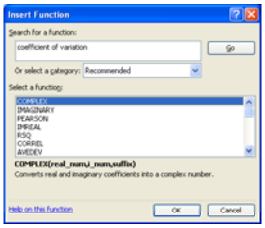
Variates are X_i:

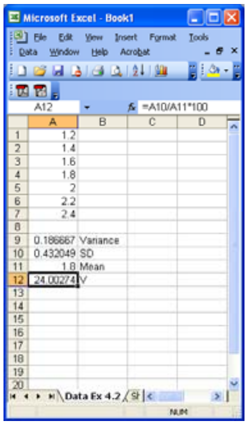
1.2
1.4
1.6
1.8
2.0
2.2
2.4

n = 7
Mean = 1.8
Median = 1.8
Mode = N/A









The “Normal” Curve

Basic Properties

- Unimodal
- Symmetrical around mean
- Assymptotic to axis (± ∞)
- Bell-shaped
- Area under the curve = 1
- Inflection points at μ ± σ
- 99% of area defined by ± 3 σ

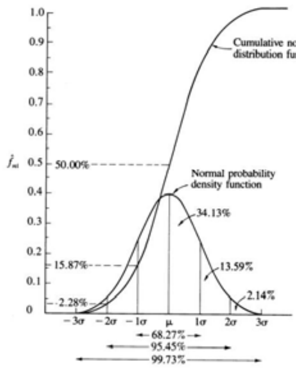
The Normal Distribution

Normal Probability Density Function:

$$Y_i = \frac{1}{\sigma\sqrt{2\pi}} e^{-(X_i-\mu)^2/2\sigma^2}$$

Y_i = height of ordinate or density
μ = mean of distribution
σ = SD of distribution

Thus, there are an infinite number of such NPDF’s.



Cumulative Normal Distribution Function

and

Normal Probability Density Function

Standard Normal Curve

<u>Mathematical properties make this one</u>
<u>of the most significant advances in all of statistics</u> !

Deviation from the mean is measured in:
Standard Deviates

Expressing distance from the mean in units of σ:
Standard Normal Deviates

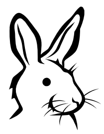
Standard Normal Deviates

$$Z = \frac{\bar{Y} - \mu}{\sigma}$$

Because of this relationship,
if the mean and variance
of a population is known, one can calculate a
probability associated with an observation Y.

Using SND to Determine Probabilities
- Example -

A biologist needs to build a trap to catch rabbits.



From years of morphometric analysis, it is well established that cottontail rabbits are known to have a:

mean shoulder width = 3.80 in
variance around mean = 0.36 in

Q: If the trap door is made to be 5.00 in wide, what percentage of rabbits will be able to make it through the door?

Example (continued):

First determine the standard normal deviate (SND).

$$Z = \frac{\bar{Y} - \mu}{\sigma} = \frac{5.0 - 3.8}{0.6} = 2.00$$

Go to Table B.2 and find the area under the curve at Z = 2.0 (defined as that point and to the right).

The area is 0.0228, thus the area to the left (rabbits) is 1 - 0.0228 = 0.9772 (97.72%)

Departures from Normality

A true NDF is largely a theoretical construct.

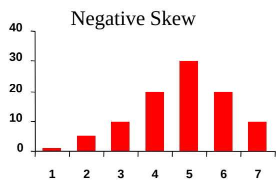
Extremely large samples may approach a true NDF.

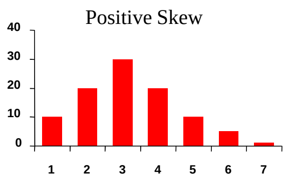
Practically, most statistical samples are not normal.

There are two primary measures of “shape-departure”:

Skewness is asymmetry about the abscissa.

Kurtosis is vertical shape deflection.



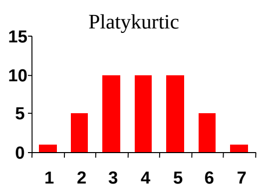


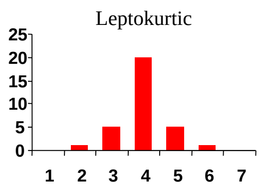
Skewness

Skewness measures the amount of "tailage" or asymmetry.

A long tail to the left is Negative Skew.

A long tail to the right is Positive Skew.





Kurtosis

Measures the proportion of the distribution in the middle and tails relative to the shoulder.

A "flat-topped" distrib. is Platykurtic.

A "pointy" distribution is Leptokurtic.

Moment Statistics

Both skewness and kurtosis can be quantified using central moment statistics (same as in physics).

The general form of a central moment is the average of the deviations of all items from the mean, each raised to the power of r:

$$\left(\frac{1}{n}\right)\sum\left(Y-\bar{Y}\right)^r$$

Moment Statistics

The <u>first</u> central moment is zero [mean] by definition (μ).

The <u>second</u> central moment is the variance (σ^2).

The <u>third</u> central moment is the skewness (g_1).

The <u>fourth</u> central moment is the kurtosis (g_2).

Moment Statistics

- Skewness -

$$g_1 = \frac{k_3}{\sqrt{\left(s^2\right)^3}} \quad \text{where,}$$

$$k_3 = \frac{n\sum X_i^3 - 3\sum X_i \sum X_i^2 + 2\left(\sum X_i\right)^3 / n}{(n-1)(n-2)}$$

For a true N distribution, g_1 should be zero.

Negative values of g_1 are attained for left-skewed distributions.

Positive values of g_1 are attained for right-skewed distributions.

$|g_1| > 1.0$ is problematic.

Moment Statistics

- Kurtosis -

$$g_2 = \frac{k_4}{s^4}, \text{ where}$$

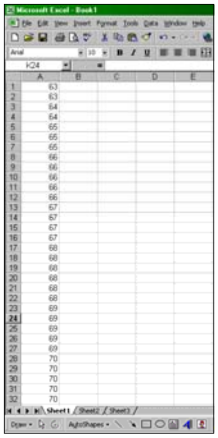
$$k_4 = \frac{\left(n^3+n^2\right)\sum X^4 - 4\left(n^2+n\right)\sum X^3 \sum X - 3\left(n^2-n\right)\left(\sum X^2\right)^2 + 12n\sum X^2\left(\sum X\right)^2 - 6\left(\sum X\right)^4}{n(n-1)(n-2)(n-3)}$$

For a true N distribution, g_2 should be zero.

Negative values of g_2 are attained for platykurtic distributions.

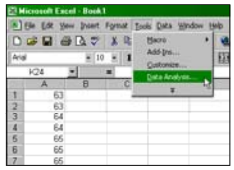
Positive values of g_2 are attained for leptokurtic distributions.

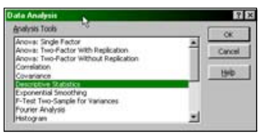
$|g_2| > 1.0$ is problematic.



Moment Statistics

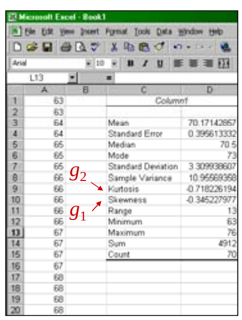
- MS-Excel, Ex. 6.1 (Zar) -





Moment Statistics

- MS-Excel -



Beta Measures of Symmetry & Kurtosis

Some authors (e.g., D'Agostino et al. 1990) speak of a population parameter called $\sqrt{\beta_1}$ as a measure of symmetry (where $\sqrt{\beta_1} = 0$ indicates symmetry).

A parameter designated β_2 is often used as a measure of kurtosis (with $\beta_2 = 3$ indicating non-kurtosis).

Beta Measures of Symmetry & Kurtosis

$$\sqrt{b_1} = \frac{(N-2)\,g_1}{\sqrt{N(N-1)}}$$

and

$$b_2 = \frac{(N-2)(N-3)\,g_2}{(N+1)(N-1)} + \frac{3(N-1)}{N+1}$$

More on this later. See also P. 69-70 & Ex. 6.1.

Graphical Methods of Distribution Evaluation

Plot the Data, plot the data, plot the data !!!

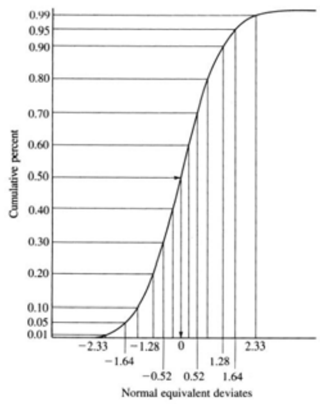
1. Histograms
2. Cumulative Normal Probability Plots (CNPPs)

Cumulative Normal Probability Plots (CNPPs)

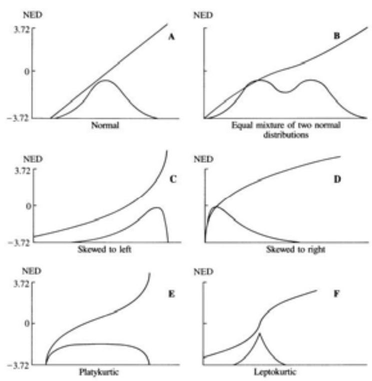
RECALL...
A normal frequency distribution graphed in a cumulative fashion produces a sigmoidal curve.

If we drop perpendiculars to the abscissa from the cumulative normal curve, we obtain corresponding quantiles.

Normal Equivalent Deviates (NEDs) are cumulative proportions transformed on to a standard deviation scale.



Transformation of cumulative percentages into NEDs



Frequency distributions with their corresponding cumulative distributions plotted as:

<u>Normal Quantile Plots</u>