# Economic Risk and Decision Analysis for Oil and Gas Industry
# CE81.9008

**School of Engineering and Technology**
**Asian Institute of Technology**

**January Semester**

**Presented by**
**Dr. Thitisak Boonpramote**

**Department of Mining and Petroleum Engineering, Chulalongkorn University**

# Descriptive Statistics

**Part B:**
**Working With Grouped Data**

## Grouping Data

- Condensing large data sets into groups simplifies calculations of parameters
- Steps in grouping
  1. Define **classes** for data set to be analyzed
  2. Determine **frequency** of data element appearances in each class
  3. Calculate **absolute and relative frequency**
  4. Calculate **class mark (CM**), or **midpoint**, for each class
  5. Proceed with calculation of **parameters**

## Guidelines for Defining Classes

- Number of classes should be between 5 and 20
- Define classes so that every element in data set falls into one and only one class
- No class should be empty

## Guidelines for Defining Classes

- Approximate **number of classes ($N_c$)** for data set with $N$ elements

$$N_c = 1 + 3.322 \log N$$

  - **Class interval (CI)** should be same for entire data set

$$\text{CI} = \frac{X_{\max} - X_{\min}}{N_c}$$

## Determining Mean

Frequency of individual class

Class mark (mid value of individual class)

$$\overline{X} = \frac{\sum_{i=1}^{n} f_i (\text{CM}_i)}{\sum_{i=1}^{n} f_i}$$

## Determining Median, Mode

Lower boundary of median class

Number of elements in sample

Cumulative frequency of class preceding median class

$$\text{Median} = L + \text{CI}\left(\frac{n/2 - a}{b}\right)$$

Class interval

Number of elements in median class

- Mode taken as CM of class with highest frequency of data elements

## Determining Geometric, Harmonic Mean

- Geometric mean

$$G_m = \exp\left[\frac{\sum_{i=1}^{n} f_i \ln(\text{CM}_i)}{\sum_{i=1}^{n} f_i}\right]$$

- Harmonic mean

$$H_m = \exp\left[\frac{\sum_{i=1}^{n} f_i}{\sum_{i=1}^{n} (f_i/\text{CM}_i)}\right]$$

## Determining Standard Deviation, Variance

- Standard deviation

$$s = \sqrt{\frac{\sum_{i=1}^{n} f_i(\mathrm{CM}_i)^2 - \left[\sum_{i=1}^{n} f_i(\mathrm{CM}_i)\right]^2 / n}{n}}$$

- Variance = $s^2$

## Calculate Parameters From Grouped Data

- Calculate **measures of central tendency** for grouped drill-bit data

| Bit number | Ft. Drilled |
|---|---|
| 1 | 53 |
| 2 | 69 |
| 3 | 72 |
| 4 | 76 |
| 5 | 80 |
| 6 | 89 |
| … | … |
| 20 | 139 |

- Group data

$$N_c = 1 + 3.322 \log N$$
$$= 1 + 3.322 \log(20)$$
$$= 5.322 \cong 5$$

- Calculate class interval

$$\mathrm{CI} = \frac{X_{\max} - X_{\min}}{N}$$
$$= \frac{139 - 53}{5} = 17.2 \cong 17$$

## Calculate Parameters From Grouped Data

- Calculate measures of central tendency for grouped drill-bit data

| Class | Class Mark (CM) | Freq. (*f*) | *f* (CM) | *f* (CM)$^2$ | *f* ln(CM) | *f* /(CM) |
|---|---|---|---|---|---|---|
| 52-69 | 60.5 | 2 | 121.0 | 7.320.50 | 8.2053 | 0.0331 |
| 70-87 | 78.5 | 3 | 235.5 | 18,486.75 | 12.0893 | 0.0382 |
| 88-105 | 96.5 | *b* = 6 | 579.0 | 55,873.50 | 27.4173 | 0.0622 |
| 106-123 | 114.5 | 6 | 697.0 | 78,661.50 | 28.4434 | 0.0524 |
| 124-141 | 132.5 | 3 | 397.5 | 52,668.75 | 14.6597 | 0.0226 |
| | | $\Sigma f$ = 20 | 2,020.0 | 213,011.00 | 91.8150 | 0.2085 |

## Calculate Parameters From Grouped Data

- Calculate measures of central tendency for grouped drill-bit data

| Class | CM | *f* | *f* (CM) | *f* (CM)$^2$ | *f* ln(CM) | *f* /(CM) |
|---|---|---|---|---|---|---|
| | | $\Sigma f$ = 20 | 2,020.0 | 213,011.00 | 91.8150 | 0.2085 |

- Mean

$$\overline{X} = \frac{\sum_{i=1}^{n} f_i(CM_i)}{\sum_{i=1}^{n} f_i} = \frac{2{,}020.0}{20} = 101\,\text{ft}$$

## Calculate Parameters From Grouped Data

- Calculate measures of central tendency for grouped drill-bit data

| Class | CM | *f* | *f* (CM) | *f* (CM)$^2$ | *f* ln(CM) | *f* /(CM) |
|---|---|---|---|---|---|---|
| | | $\Sigma f = 20$ | 2,020.0 | 213,011.00 | 91.8150 | 0.2085 |

- Median

$$M = L + CL\left(\frac{n/2 - a}{b}\right)88 + 17\left(\frac{20/2 - 5}{6}\right)$$

$$= 102.17 \text{ ft}$$

## Calculate Parameters From Grouped Data

- Calculate measures of central tendency for grouped drill-bit data

| Class | CM | *f* | *f* (CM) | *f* (CM)$^2$ | *f* ln(CM) | *f* /(CM) |
|---|---|---|---|---|---|---|
| | | $\Sigma f = 20$ | 2,020.0 | 213,011.00 | 91.8150 | 0.2085 |

- Geometric mean

$$G_m = \exp\left[\frac{\sum_{i=1}^{n} f_i \ln(CM_i)}{\sum_{i=1}^{n} f_i}\right] = \exp\left(\frac{91.8150}{20}\right) = e^{4.5980} = 98.57 \text{ ft}$$

## Calculate Parameters From Grouped Data

- Calculate measures of central tendency for grouped drill-bit data

| Class | CM | *f* | *f* (CM) | *f* (CM)$^2$ | *f* ln(CM) | *f* /(CM) |
|---|---|---|---|---|---|---|
| | | $\Sigma f = 20$ | 2,020.0 | 213,011.00 | 91.8150 | 0.2085 |

- Harmonic mean

$$H_m = \left[ \frac{\sum_{i=1}^{n} f_i}{\sum_{i=1}^{n} f_i \left( f_i / CM_i \right)} \right] = \frac{20}{0.2085} = 95.92 \text{ ft}$$

## Calculate Parameters From Grouped Data

- Calculate measures of central tendency for grouped drill-bit data

| Class | CM | *f* | *f* (CM) | *f* (CM)$^2$ | *f* ln(CM) | *f* /(CM) |
|---|---|---|---|---|---|---|
| | | $\Sigma f = 20$ | 2,020.0 | 213,011.00 | 91.8150 | 0.2085 |

- Standard deviation

$$s = \sqrt{\frac{\sum_{i=1}^{n} f_i (\text{CM}_i)^2 - \left[ \sum_{i=1}^{n} f_i (\text{CM}_i) \right]^2 \Big/ n}{n}} = \sqrt{\frac{213{,}011 - \frac{2{,}020^2}{20}}{20}}$$

$$= \sqrt{\frac{213{,}011 - 204{,}020}{20}} = \sqrt{449.55} = 21.20 \text{ ft}$$

## Calculate Parameters From Grouped Data

- Calculate measures of central tendency for grouped drill-bit data

| Class | CM | *f* | *f* (CM) | *f* (CM)$^2$ | *f* ln(CM) | *f* /(CM) |
|---|---|---|---|---|---|---|
| | | $\Sigma f = 20$ | 2,020.0 | 213,011.00 | 91.8150 | 0.2085 |

- Variance ($s^2$)

$$s^2 = 21.20^2 = 449.55$$

## Frequency Distribution (Histogram)

- Presents distribution of frequencies of values of variables
  - Sometimes used for **ungrouped data**, usually discrete variables
  - More commonly used for **grouped data**, either discrete or continuous variables
- **Absolute frequency distribution** shows actual number of data elements in each class
- **Relative frequency distribution** shows proportion of data items in each class

## Kinds of Distributions

- **Cumulative** include number or proportion of data elements **in and below** each class
- **Decumulative** include number or proportion of data elements **in and above** each class

## Frequency Curve

- Frequency polygon in which data points are fitted to smooth curve
- To fit normal curve to histogram

Number of observations ($n$ = 1 if drawn on proportional frequency)

Class interval used to draw histogram

$$f(x) = \frac{n \times \mathrm{CI}}{s\sqrt{2\pi}} e^{-0.5\left(\frac{X-\overline{X}}{s}\right)^2}$$

## Ogive

- Presents cumulative frequencies (relative or absolute) vs. class boundaries
  - **Cumulative** (frequencies equal to or less than) **ogive** has frequencies plotted at **upper boundary** of each class
  - **Decumulative** (frequencies equal to or greater than) **ogive** has frequencies plotted at **lower boundary** of each class

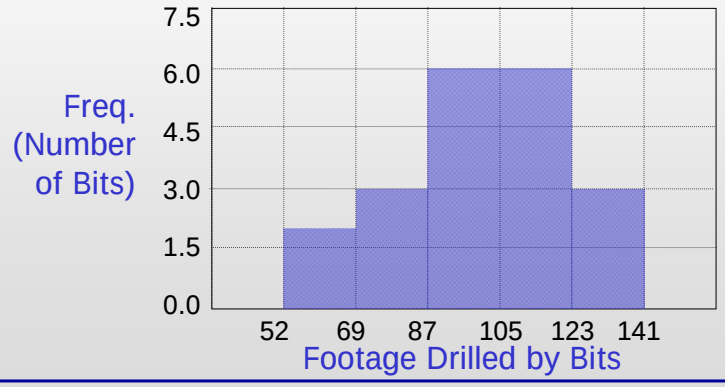
## Plot Bit Records

- Construct histogram and ogives (both kinds)

| Class | Class Mark (CM) | Freq. (*f* ) | Cum. Freq (CF) | Cum. Rel. Freq. | Decum. Rel. Freq. |
|---|---|---|---|---|---|
| 52-69 | 60.5 | 2 | 2 | 2/20=0.10 | 20/20=1.00 |
| 70-87 | 78.5 | 3 | 3+2=5 | 5/20=0.25 | 18/20=0.90 |
| 88-105 | 96.5 | 6 | 6+5=11 | 11/20=0.55 | 15/20=0.75 |
| 106-123 | 114.5 | 6 | 6+11=17 | 17/20=0.85 | 9.20=0.45 |
| 124-141 | 132.5 | 3 | 3+17=20 | 20/20=0.00 | 3/20=0.15 |
| | | 20 | | | |

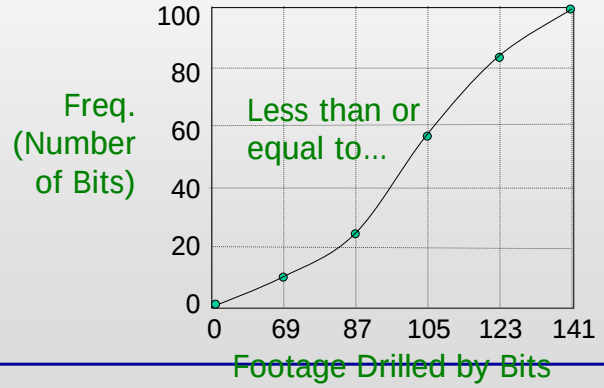
## Plot Bit Records

- Construct histogram and ogives (both kinds)
- Histogram, frequency curve



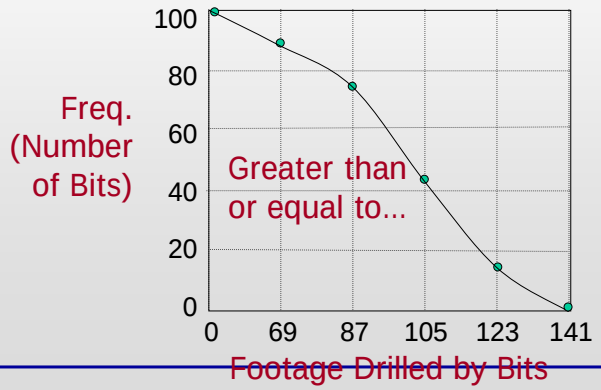

## Plot Bit Records

- Construct histogram and ogives (both kinds)
- Ogives (cumulative)



## Plot Bit Records

- Construct histogram and ogives (both kinds)
- Ogives (decumulative)



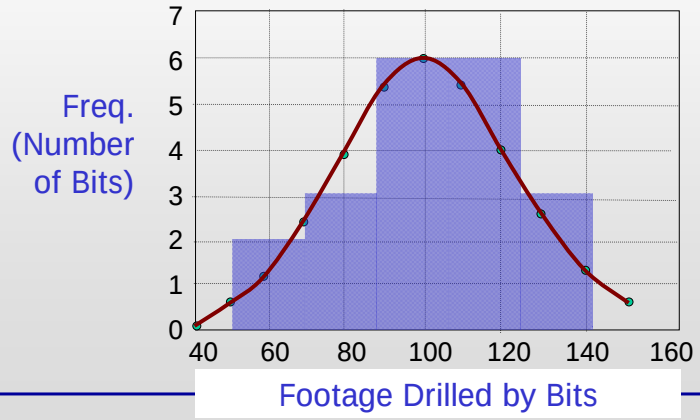

## Plot Bit Records

- Fit a normal curve to the histogram

| $x$ | $p=-0.5\left(\frac{x-\overline{X}}{s}\right)^s$ | $e^p$ | $f(x)=e^p\frac{20\mathrm{CI}}{s\sqrt{2\pi}}$ |
|---|---|---|---|
| 40 | -3.5567 | 0.0285 | 0.1702 |
| 50 | -2.4805 | 0.0837 | 0.4992 |
| 60 | -1.5978 | 0.2024 | 1.2070 |
| ... | ... | ... | ... |
| 140 | -1.4972 | 0.2238 | 1.3347 |
| 150 | -2.3549 | 0.0949 | 0.5661 |

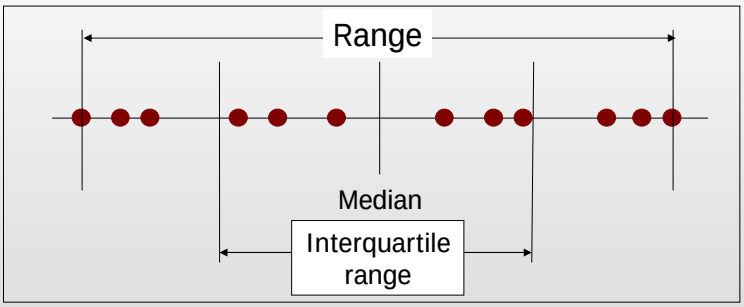
## Plot Bit Records

- Fit a normal curve to the histogram




## Quartiles

- Quartiles (symbol *Q*) divide data set into four equal parts



## Calculating Quartiles

$$Q_1 = L_1 + \text{CI}\left(\frac{n/4 - a_1}{b_1}\right)$$

$$Q_2 = L_2 + \text{CI}\left(\frac{n/4 - a_2}{b_2}\right)$$

$$Q_3 = L_3 + \text{CI}\left(\frac{n/4 - a_3}{b_3}\right)$$

Analogous to *a* in equation for computational median

Lower limits of classes containing quartiles

Analogous to *b* in equation for computational median

## Other Subdivisions of Data Sets

- **Deciles (symbol *D*)** divide data set into 10 equal parts
- **Percentiles (symbol *P*)** divide data set into 100 equal parts

Median, second quartile ($Q_2$), fifth decile ($D_5$), and 50th percentile ($P_{50}$) are identical.

## Position Indicator for Fractile

Fractile indicator ($P_i$, $D_i$, or $Q_i$) at which fractile is read from ungrouped data

Magnitude of desired fractile

If $D_7$ is desired, $i = 7$

$$F = \frac{i}{k}(n+1)$$

Maximum number of divisions

## Determine Footage Drilled at Fractiles

- Use ungrouped drilling records
- Determine footage drilled for $P_{50}$, $D_7$, $Q_3$, and $D_5$

| Bit number | Ft. Drilled |
|---|---|
| 1 | 53 |
| 2 | 69 |
| 3 | 72 |
| … | … |
| 20 | 139 |

$$F(P_{50}) = \frac{50}{100}(20+1) = 10.5$$

10th value = 102; 11th value = 105
$\therefore P_{50}$ = (102+105)/2 = 103.5

## Determined Footage Drilled at Fractiles

- Use ungrouped drilling records
- Determine footage drilled for $P_{50}$, $D_7$, $Q_3$, and $D_5$

| Bit number | Ft. Drilled |
|---|---|
| 1 | 53 |
| 2 | 69 |
| 3 | 72 |
| … | … |
| 20 | 139 |

$$F(D_7) = \frac{7}{10}(20+1) = 14.7$$

14th value = 110; 15th value = 115
$\therefore D_7$ = 110+0.7(115-110)=113.5

## Determined Footage Drilled at Fractiles

- Use ungrouped drilling records
- Determine footage drilled for $P_{50}$, $D_7$, $Q_3$, and $D_5$

| Bit number | Ft. Drilled |
|---|---|
| 1 | 53 |
| 2 | 69 |
| 3 | 72 |
| … | … |
| 20 | 139 |

$$F(Q_3) = \frac{3}{4}(20+1) = 15.75$$

15th value = 115; 16th value = 116
$\therefore Q_3$ = 115+0.75(116-115)=115.75

## Determined Footage Drilled at Fractiles

- Use ungrouped drilling records
- Determine footage drilled for $P_{50}$, $D_7$, $Q_3$, and $D_5$

| Bit number | Ft. Drilled |
|---|---|
| 1 | 53 |
| 2 | 69 |
| 3 | 72 |
| … | … |
| 20 | 139 |

$$F(D_5) = \frac{5}{10}(20+1) = 10.5$$

Since $D_5 = P_{50}$,
$D_5 = 103.5$

## Coefficient of Peakedness, $a_4$

- Dimensionless value

$$a_4 = \frac{\text{fourth central moment}}{s^4} = \frac{m_4}{s_4}$$

$$= \frac{\sum_{i=1}^{n} f_i \left(X_i - \overline{X}\right)^4}{n s^4}$$

## Coefficient of Skewness, $a_3$

$$a_3 = \frac{\text{third central moment}}{s^3} = \frac{m_3}{s_3}$$

$$= \frac{\sum_{i=1}^{n} f_i \left(X_i - \bar{X}\right)^3}{\frac{n}{s^3}}$$

When $a_3$ = 0, curve is symmetric or bell shaped
When $a_3$ < 0, curve is skewed to right
When $a_3$ > 0, curve is skewed to left

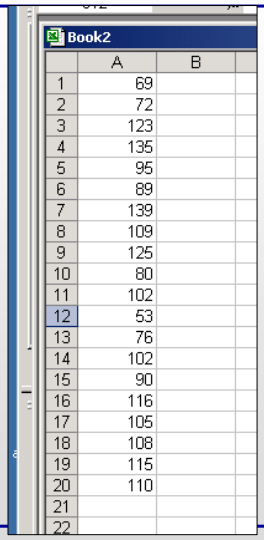
## Using Spreadsheets (Excel)

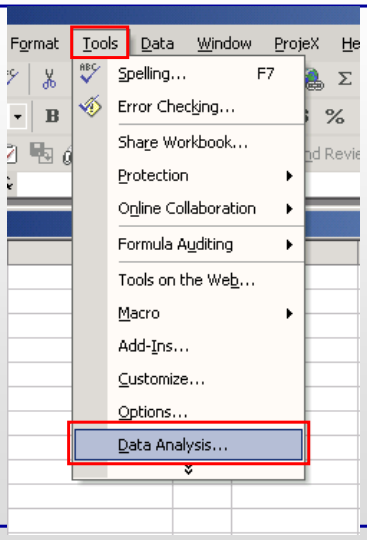
- May need to install plug-in(s)
- May enter data in random order
  - No need to sort
  - May need to enter data in single column or row
- Excel uses '**kurtosis**' for '**peakedness**'

## Find Means, Median, Mode

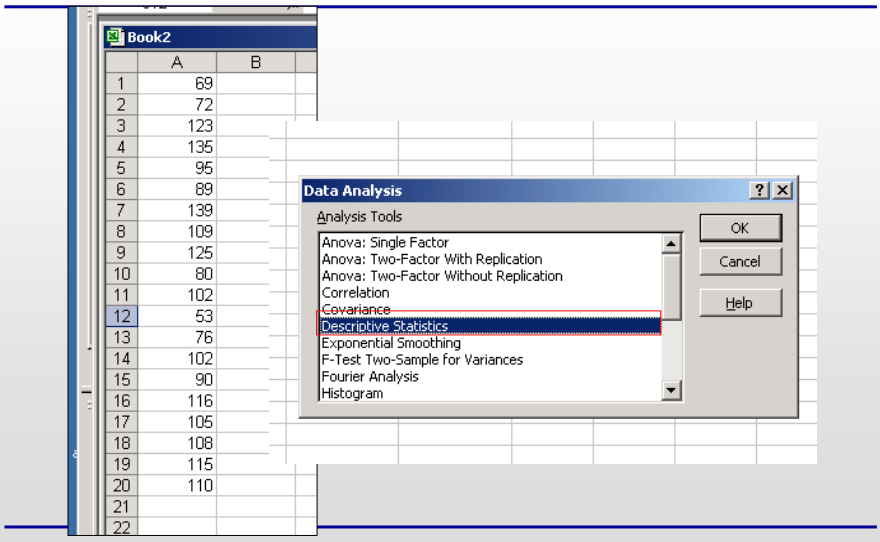
- Enter data in single column
- Use Tools > Data Analysis > Descriptive Analysis
  - Enter range of data
  - Click ‘Grouped by columns’
  - Click ‘Summary Statistics’

## Statistics From Excel

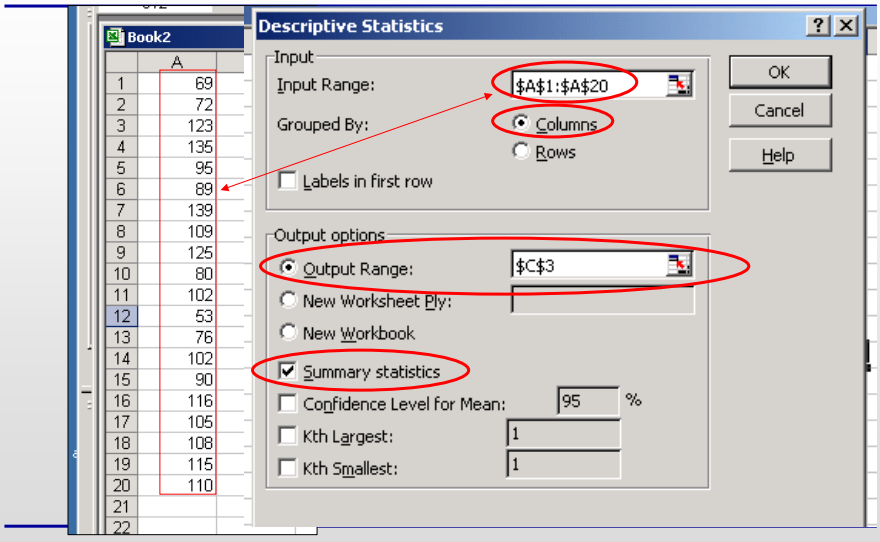


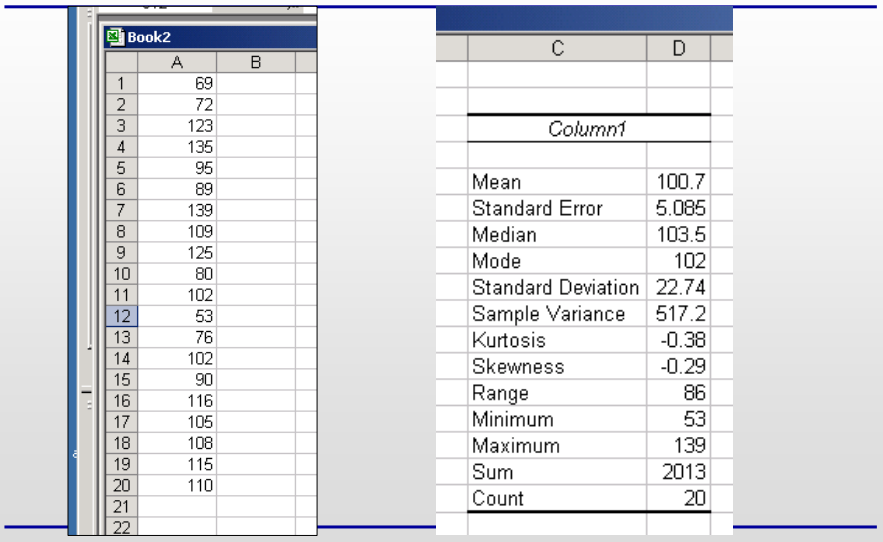
## Statistics From Excel




## Statistics From Excel



## Statistics From Excel

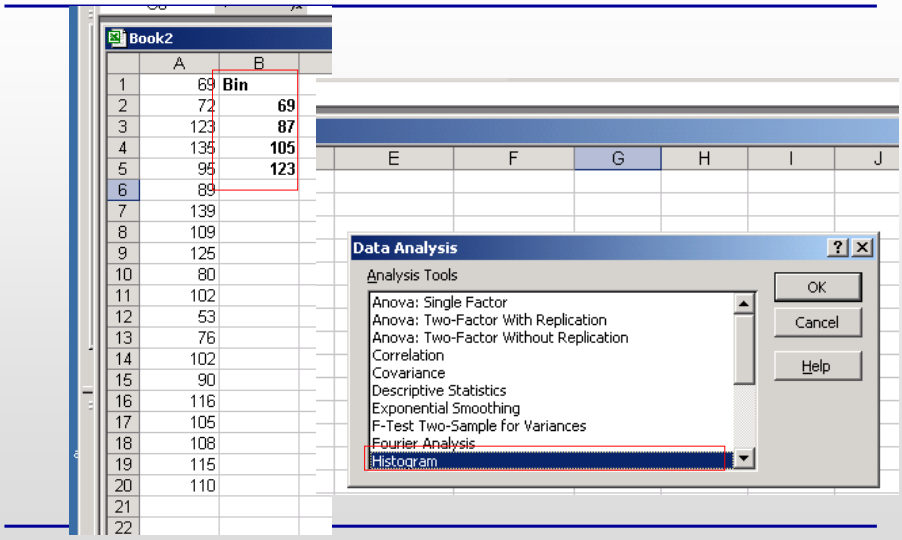


Book2

| | A | B |
|---|---|---|
| 1 | 69 | |
| 2 | 72 | |
| 3 | 123 | |
| 4 | 135 | |
| 5 | 95 | |
| 6 | 89 | |
| 7 | 139 | |
| 8 | 109 | |
| 9 | 125 | |
| 10 | 80 | |
| 11 | 102 | |
| 12 | 53 | |
| 13 | 76 | |
| 14 | 102 | |
| 15 | 90 | |
| 16 | 116 | |
| 17 | 105 | |
| 18 | 108 | |
| 19 | 115 | |
| 20 | 110 | |
| 21 | | |
| 22 | | |

| | C | D |
|---|---|---|
| | *Column1* | |
| | Mean | 100.7 |
| | Standard Error | 5.085 |
| | Median | 103.5 |
| | Mode | 102 |
| | Standard Deviation | 22.74 |
| | Sample Variance | 517.2 |
| | Kurtosis | -0.38 |
| | Skewness | -0.29 |
| | Range | 86 |
| | Minimum | 53 |
| | Maximum | 139 |
| | Sum | 2013 |
| | Count | 20 |

## Histogram From Excel



Book2

| | A | B |
|---|---|---|
| 1 | 69 | **Bin** |
| 2 | 72 | **69** |
| 3 | 123 | **87** |
| 4 | 135 | **105** |
| 5 | 95 | **123** |
| 6 | 89 | |
| 7 | 139 | |
| 8 | 109 | |
| 9 | 125 | |
| 10 | 80 | |
| 11 | 102 | |
| 12 | 53 | |
| 13 | 76 | |
| 14 | 102 | |
| 15 | 90 | |
| 16 | 116 | |
| 17 | 105 | |
| 18 | 108 | |
| 19 | 115 | |
| 20 | 110 | |
| 21 | | |
| 22 | | |

E F G H I J

Data Analysis

Analysis Tools

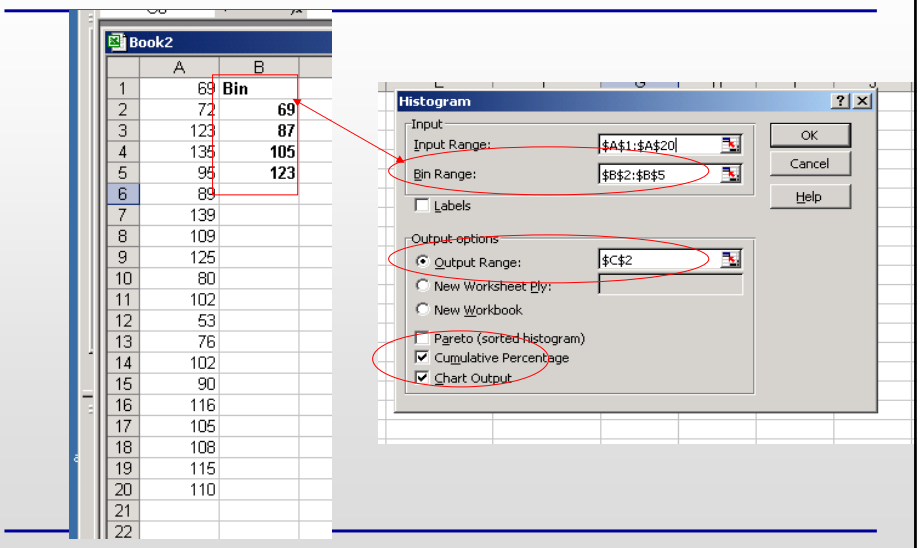
- Anova: Single Factor
- Anova: Two-Factor With Replication
- Anova: Two-Factor Without Replication
- Correlation
- Covariance
- Descriptive Statistics
- Exponential Smoothing
- F-Test Two-Sample for Variances
- Fourier Analysis
- Histogram

OK

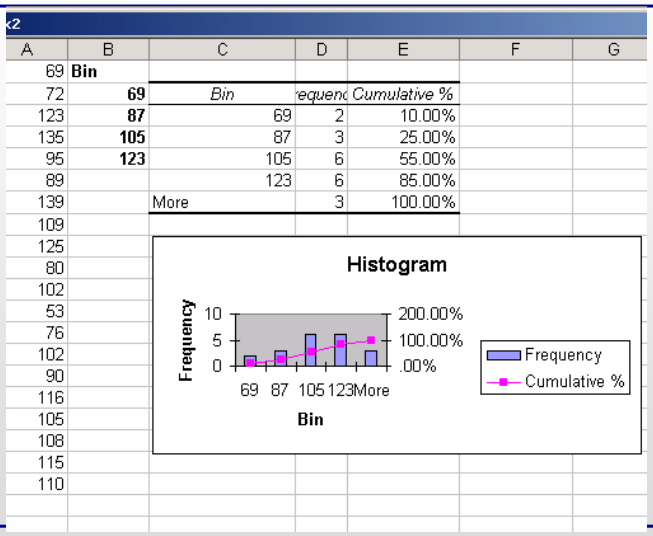
Cancel

Help

## Histogram From Excel




## Histogram From Excel



# Histogram From Excel