Practical Applications of Statistical Analysis

M. Mokhtari, Ph.D.

Statistics and Types Of Statistics

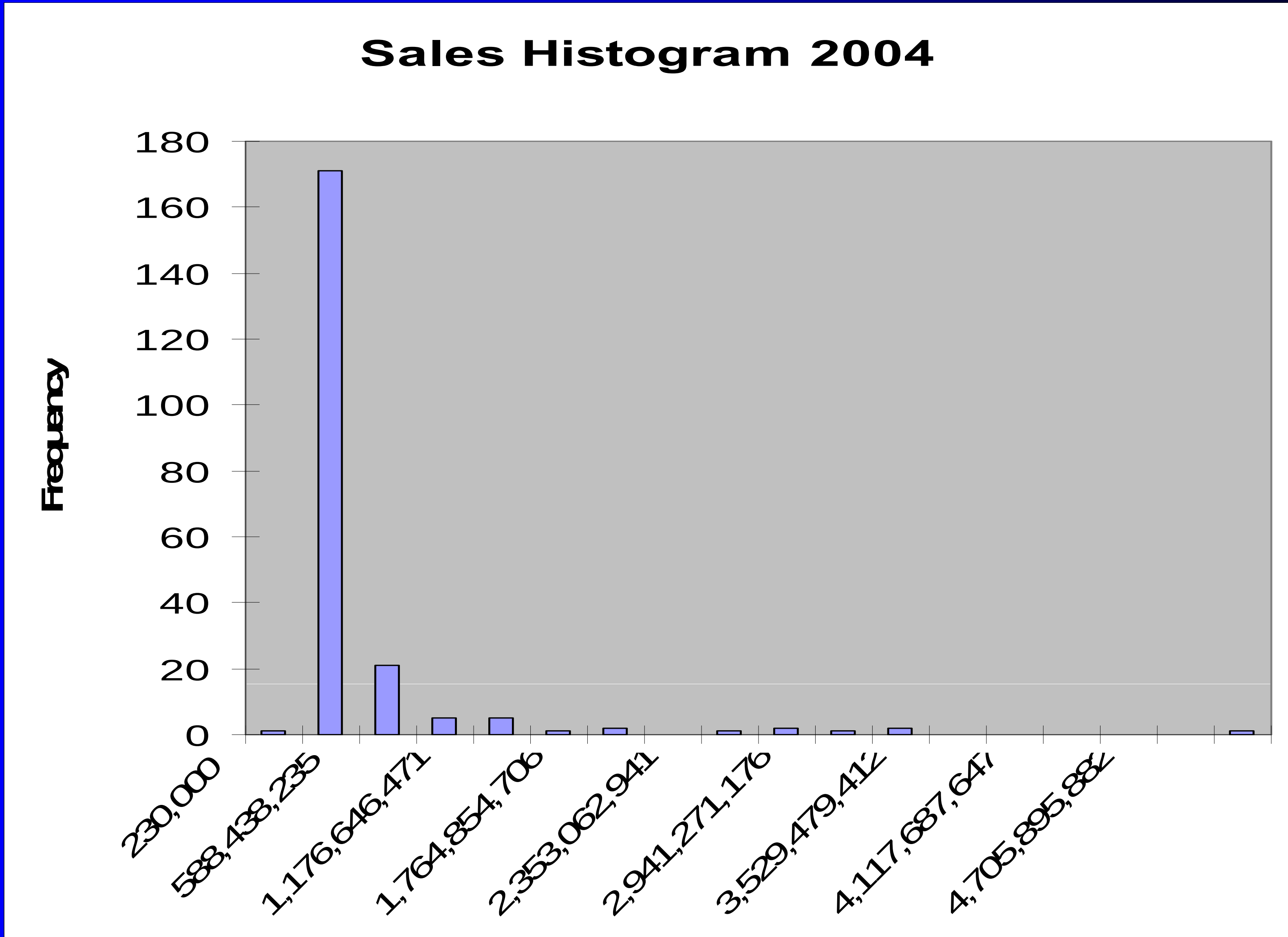
- Statistics is the science of collecting, organizing, presenting, analyzing, and interpreting numerical data for the purpose of better problem solving
- **Descriptive statistics** is the procedures used to organize and summarize data; for example, ARCS
- **Inferential statistics** is the method used to find out something about the population (*all enterprises in Armenia*) using a sample (*300 enterprises in ARCS*).

How To Organize Data & Peresnt Data (Descriptive Statistics)

- Arranging the data in ascending or descending order.
- Frequency Distribution
- Cumulative Frequency Distribution

- Bar Charts
- Pie Charts

Histogram of Sales (ARCS-2004)



What do we know and how we can use this knowledge:

Mean
(center of mass)

Variance and Standard Deviation

Confidence Intervals
Coefficient of Variation
Hypothesis Testing
Correlation
OLS

Probability Distribution

Others
(Covariance)

Quick Review: Mean Value

- Mean of the data is what you expect the data to be. *Mean could be a simple average value for the data: Life expectancy at birth = 72.7*

  GDP per capita (PPP) = $5,900
- Mean usually estimate by adding up the individual elements of a variable and dividing the sum by the total number of elements, i.e. average of 2 and 6 is then (2+6)/2
- Sometimes we hear a phrase "weighted average". It is a mean but instead of weighing each individual element equally we weigh them by specified weights (for example: food cannot be consumed in equal amount by a child and an adult, but we have to estimate an average consumption per family.)

Mean

- Average = (Sum of all values) / (Number of values)

For example: 3, 4, 7, 3, 8
Average = (3 + 4 + 7 + 3 + 8) / 5 = 5

In this example, there were 5 observations, that is, *i=1, 2, …,5.* Hence, the total number of observations are 5 (*n=5*)

$$Mean = \frac{1}{n}\sum_{i}^{n} X_i$$

Variance

- ***Variance*** is a measure of dispersion of the data around the mean
- Variance also serves as a measurement of risk, i.e. greater the variability around the mean greater the risk
- Since in the real world we rarely deal with entire population, we get to work with the ***sample variance formula***:

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

- ***Standard Deviation*** (S) is a standardized measure of dispersion around the mean. The value of standard deviation is just a square root of variance

An Example: Computation Of Variance (n=4)

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

$$S^2 = \frac{26}{3} = 8.67 \rightarrow S = 2.94$$

X_i	$\bar{X}$	$X_i - \bar{X}$	$(X_i - \bar{X})^2$
2	5	-3	9
8	5	3	9
3	5	-2	4
7	5	2	4

$\sum X_i$	$\bar{X}$	$\sum (X_i - \bar{X})$	$\sum (X_i - \bar{X})^2$
20	5	0	26

Covariance

- A measure of co-movement
- It is just the average of the cross products of two random variables

$$Covariance = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{N-2}$$

X-Average	Y-Average	(X-Average)(Y-Average)
1	-7.75	-7.25
11	8.75	96.25
-4	-2.25	9.00
-8	0.75	-6.00

Sum / (4-2) = 92/2 = 46 = Covariance

$$Covariance = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{N-2}$$

Application of Variance and Covariance: Regression Model

Modeling Y: Assume that theory tells us that Y is influenced by X:

$$Y_t = f(X_t)$$

Also, assume that , the relationship between Y and X is linear; and that, "e" is a portion of Y that is not explained by X. Then, we have the following model:

$$Y_t = \beta_0 + \beta_1 X_t + e_t$$

Application of Variance and Covariance :

Modeling Y: Assume Y is influenced by X; also, assume that "e" is a portion of Y that is not explained by X)

- Example:

$$Y_t = \beta_0 + \beta_1 X_t + e_t$$

Note : Y_t = β_0 if $X_t = 0$ or $\beta_1 = 0$

Note: $\frac{\Delta Y}{\Delta X} = \beta_1$ = Change in Y, if X changes by a unit

Note: *Elasticity* = *E = %ΔY / %ΔX =* $E = \beta_1 \frac{\overline{X}}{\overline{Y}}$

Example: Modeling Y = f (X)

Year	Y	X
1994	2	12
1995	3	16
1996	3	23
1997	5	30
1998	6	34
1999	8	48
2000	9	65
2001	12	74
Average or Mean =	6	37.75
Variance =	12	512.7

Covariance(Y, X) = 90

Application - continued:

- Slope or parameter β_1 is estimated using what we have already learned:

$$\hat{\beta}_1 = \frac{\text{Covariance}(Y, \text{X})}{Variance(\text{X})}$$

- This slope tells us the following: For one unit increase in explanatory variable X, how many units will the predicted variable Y.
- And β_0 is estimated by:

$$\beta_0 = \overline{Y} - \beta_1 \overline{X}$$

Application - continued:

- Slope or parameter β_1 is estimated by:

$$\hat{\beta}_1 = \frac{\text{Covariance}(Y, \text{X})}{\textit{Variance}(\text{X})}$$

$$\hat{\beta}_1 = \frac{90}{512.7} = 0.176$$

- This slope tells us the following: For one unit increase in X, we find 0.176 unit increase in Y. Or for 100 unit increase in X, then Y will go up by 17.6 unit.

- β_0 is estimated by:

$$\beta_0 = \overline{Y} - \beta_1 \overline{X}$$

$$\beta_0 = 6 - (0.176)(37.75) = -0.64$$

This shows that if X =0, then Y = -0.64

Application of Variance and Covariance - continue

$$Y_t = \beta_0 + \beta_1 X_t + e_t$$

- *Elasticity* or *E= %ΔY / %ΔX is* :

$$E = \beta_1 \frac{\bar{X}}{\bar{Y}}$$

$$E = (0.176)\frac{37.75}{6} = 1.107$$

Therefore, "E= 1.107" shows that for every one percent rise in X, we will see Y to rise by 1.107 percent.

Univariate Descriptive Statistics - ARCS 2009

	VI=Employment	*V39=Exports*	*V69=Sales*
Mean	17	127,235,294	98,237,481
Standard Error	3	60,639,880	22,467,923
Median	7	30,000,000	27,000,000
Mode	3	30,000,000	7,000,000
Standard Deviation	49	250,024,631	187,980,132
Sample Variance	2,447	62,512,316,176,470,600	35,336,530,138,429,900
Kurtosis	212	5	9
Skewness	14	2	3
Range	799	800,000,000	879,500,000
Minimum	1	0	500,000
Maximum	800	800,000,000	880,000,000
Sum	5,038	2,163,000,000	6,876,623,700
Count	300	17	70

Covariance - ARCS 2009

	VI	*V39*	*V69*
VI=Employment	2,439		
V39=Export	30,762,508,651	58,835,121,107,266,400	
V69=Sales	1,319,889,425	5,641,795,918,367,350	34,831,722,565,023,800

Correlation Coefficients = Cov (Y, X) / (S_y S_x) - ARCS 2009

	VI	*V39*	*V69*
VI=Employment	1		
V39=Export	0.69	1.00	
V69=Sales	0.42	0.21	1