PENNY ANTICS

STATISTICS WORKSHOP

WL / TQ

WORKSHOP NOTES

FOR

PENNY ANTICS

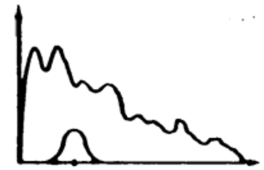
NOTES PREPARED BY

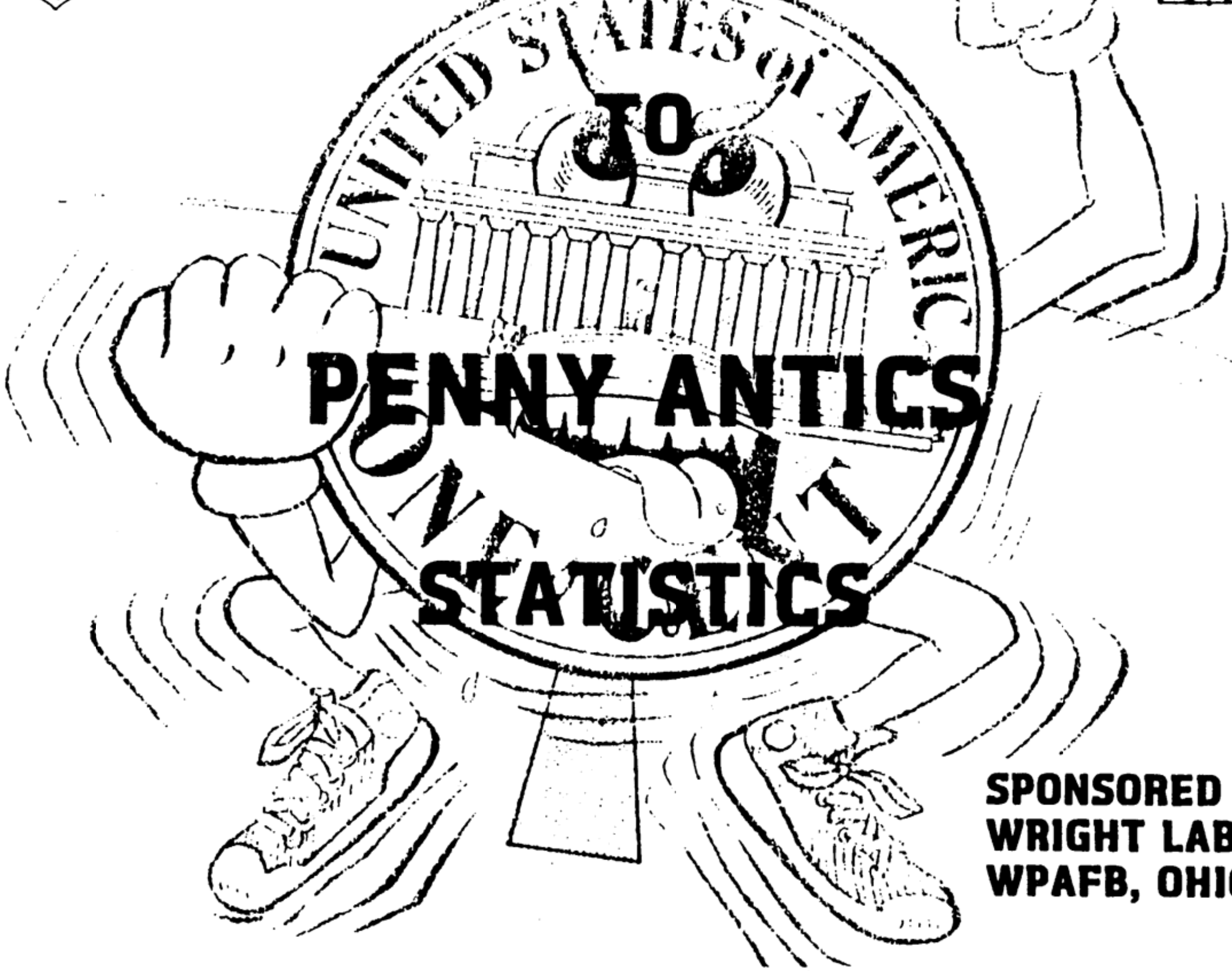
JOHN C. SPARKS
WRIGHT LABORATORY
WPAFB, OHIO 45433

WELCOME

TO

PENNY ANTICS STATISTICS





SPONSORED BY:
WRIGHT LABORATORY
WPAFB, OHIO

PENNY ANTICS AGENDA

(DAY 1)

I) Welcome and Introductions
   - A Lonely Night at China Lake
   - Birthday Party
   - Quality Diamond
   - Statistical Triad

II) Exploring Data is Child's Play
   - Gathering Data by Inches and Miles
   - Stem and Leaf
   - Box and Whisker

III) A Footprint in Time
   - Let the Data Speak for Itself
   - Histogram Hysteria
   - Box it Up

IV) Basic Training
   - Measures of Central Tendency
   - Measures of Dispersion
   - Decoding the Distribution

(LUNCH BREAK)

V) First Day Jitters at the Census Bureau
   - Your Goverment Training Program
   - The Boss Wants the Numbers by Two
   - You Gotta Learn the Ropes in Order to Survive

VI) The Founding Fathers of Pennyville
   - The Great American Melting Pot
   - Rules in the Pool
   - Demographics

VII) Chebyshev Comes to Town
   - Distribution Free Wares
   - Try it for Yourself
   - Putting Teeth on the Box and Whisker

VIII) JENGA Time
   - Building a Data Base
   - A Mountaintop Experience
   - Setting Up the Goal Post
   - Quality Improvement Cycle

(END DAY 1)

PENNY ANTICS (CONT)

(DAY 2)

I) The Great Law
- Intuitive Central Limit Theorem
- Ringing the Bells
- Z-score and More
- Empirical Rule

II) The Heart of it All
- Say Hello to the Mean Machine
- The Central Limit Theorem
- What is Pythagoras Doing Here

III) A Massive Demonstration at Pennyville
- Underlying Distribution
- Order from Chaos
- I can See Clearly Now

(LUNCH)

IV) Fishin Ain't Bad in Pennyville
- Baiting our Hooks
- Confidence Building in Town
- Stalking More Elusive Game

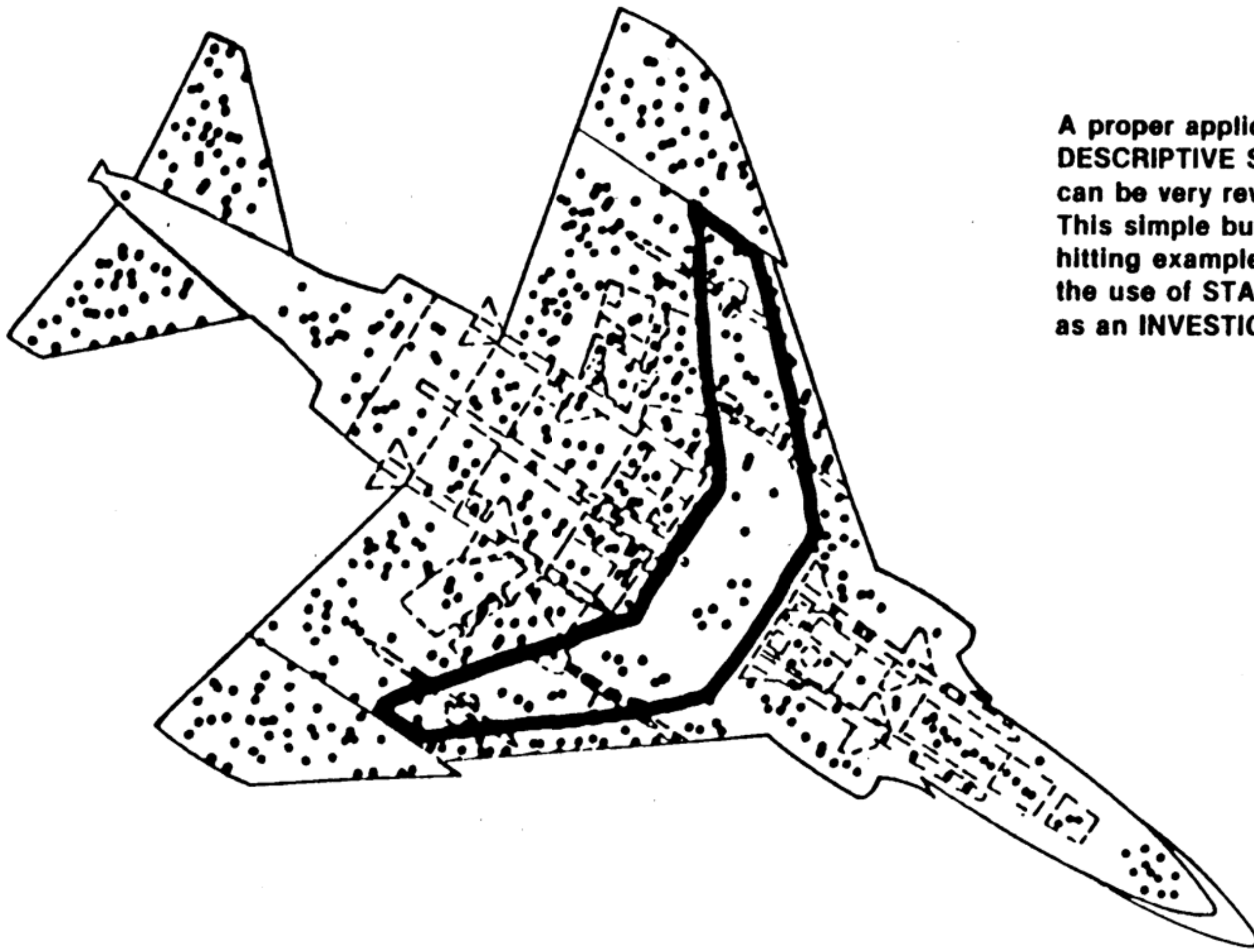
V) The Life and Times of Pennyville
- Setting the Stage
- Peaceful Beginnings
- Trouble is Brewing
- Justice is Served
- Restoration
- A Bum Rap and Lawsuit
- Peace Returns

VI) Wrapping it Up
- The Lesson of Pennyville
- Love it or Leave it
- Sadistics or Statistics

VII) Graduation and Adjournment
- Workshop Evaluation
- Certificate and 1556 Credit
- Preview of Upcoming Programs

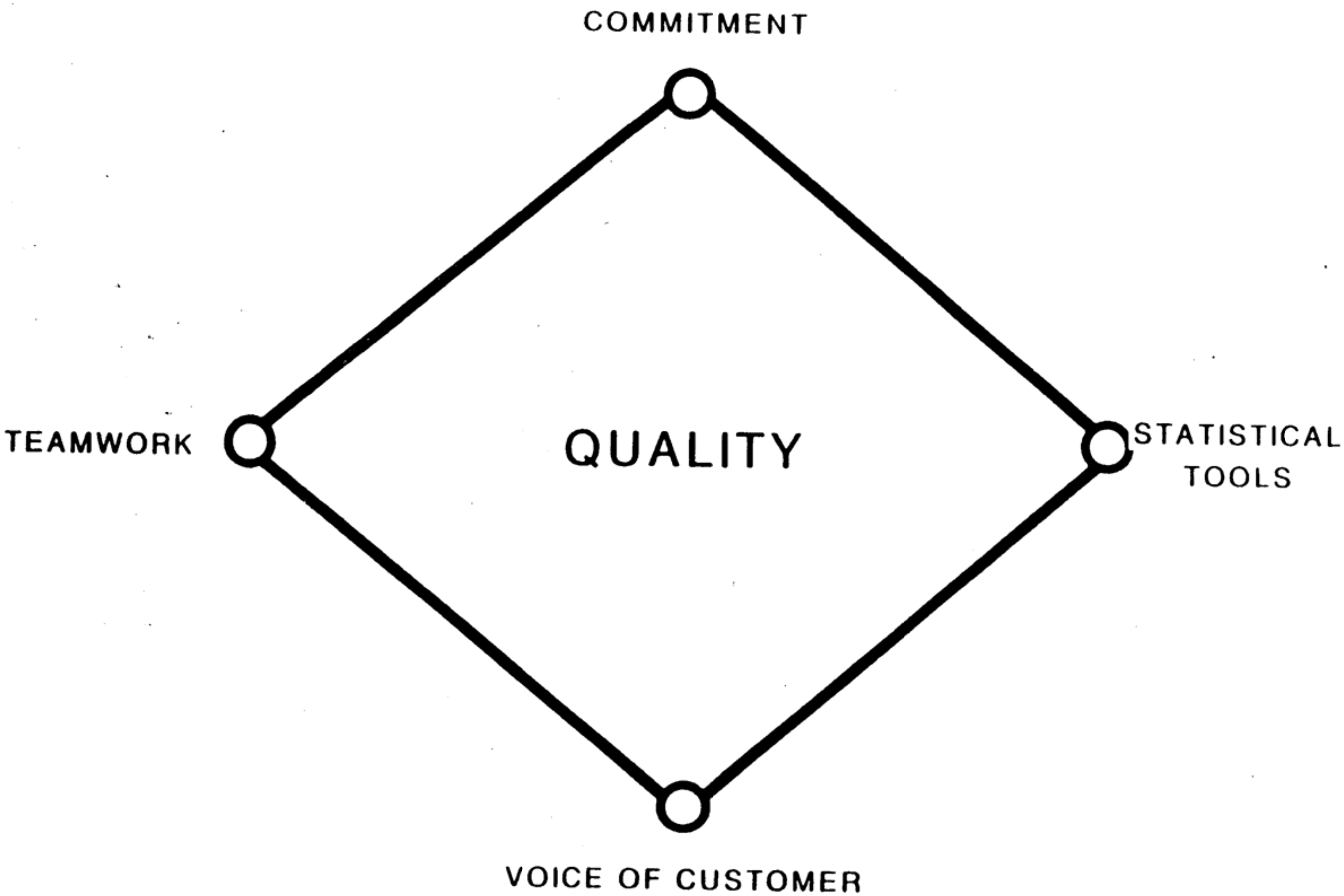
(END DAY 2)

F-4 HIT PLOT
SEA EXPERIENCE

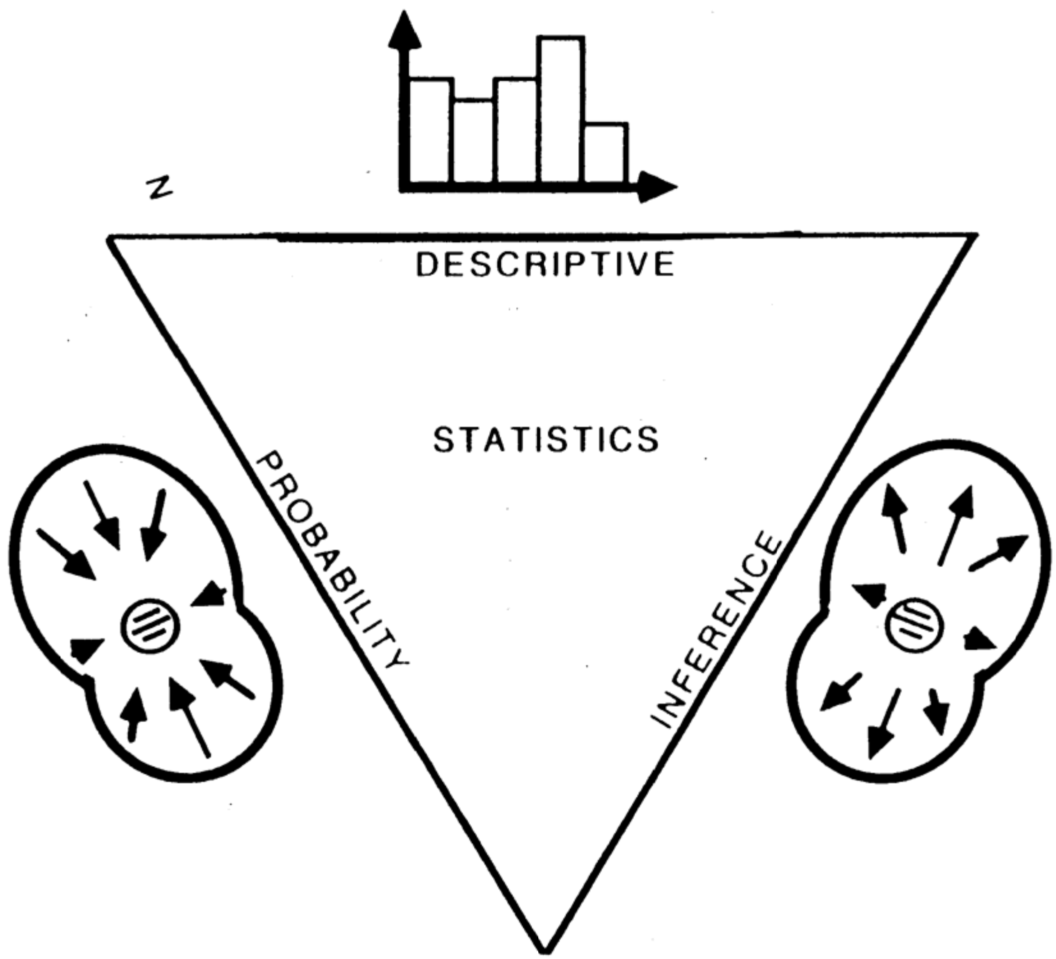


A proper application of DESCRIPTIVE STATISTICS can be very revealing ! This simple but hard hitting example illustrates the use of STATISTICS as an INVESTIGATIVE TOOL.

QUALITY DIAMOND



STATISTICAL TRIAD



HEIGHT DATA SUPPLIED BY THE
STATS 122 CLASS
-SUMMER 1991-

MALE				
71	75	72	72	65
70	69	68	72	72
72	73	75	70	64
76	72	69	72	73
74	67	67	75	74
74	73	68	78	75
68	78	69	73	76
68	73	71	74	71
73	71	70	77	74
68	73	69	71	

FEMALE			
68	66	65	64
66	62	67	67
62	60	65	66
66	64	66	
65	68	66	
63	65	60	
65	57	68	
68	63	61	
67	67	69	
64	68	63	

MIXED STEM AND LEAF PLOT

5:7
5:
6:0,0,1
6:2,2,3,3,3
6:**4**,4,4,4,**5**,5,5,5,5,5
6:6,6,6,6,6,6,7,7,7,7,7,**7,7**
6:8,8,8,8,8,**8,8,8,8,8**,9,**9,9,9,9**
7:**0,0,0,1,1,1,1,1**
7:**2,2,2,2,2,2,2,3,3,3,3,3,3,3**
7:**4,4,4,4,4,5,5,5**
7:**6,6,7**
7:**8**
7:

UNMIXED STEM AND LEAF PLOT

5:7	
5:	
6:0,0,1	
6:2,2,3,3,3	
6:4,4,4,5,5,5,5,5	**4,5**
6:6,6,6,6,6,6,7,7,7,7,7	**7,7**
6:8,8,8,8,8,9	**8,8,8,8,8,9,9,9,9**
7:	**0,0,0,1,1,1,1,1**
7:	**2,2,2,2,2,2,2,3,3,3,3,3,3,3**
7:	**4,4,4,4,4,5,5,5**
7:	**6,6,7**
7:	**8**

NOTES

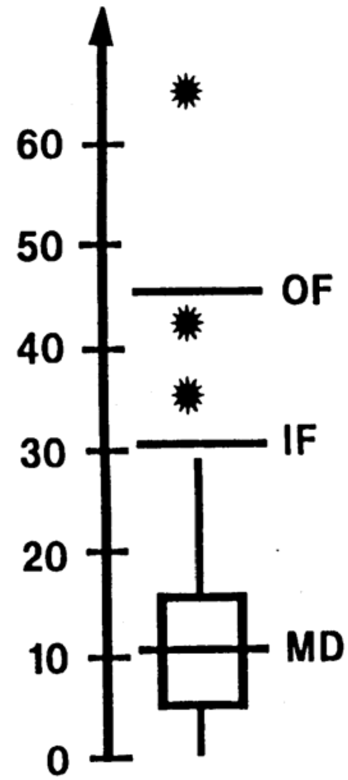
1) The use of **BOLD** and the splitting apart of the male and female heights is yet another way to MAKE DATA TALK.

MILEAGE DATA
SEPT 1991

STEM AND LEAF

6:5
6:
5:
5:
4:
4:
3:6
3:0
2:
2:3
1:5,6,7
1:0,0,0,1,2,3,3,3,4
0:5,5,5,6,6,8
0:1,3,3

BOX AND WHISKER

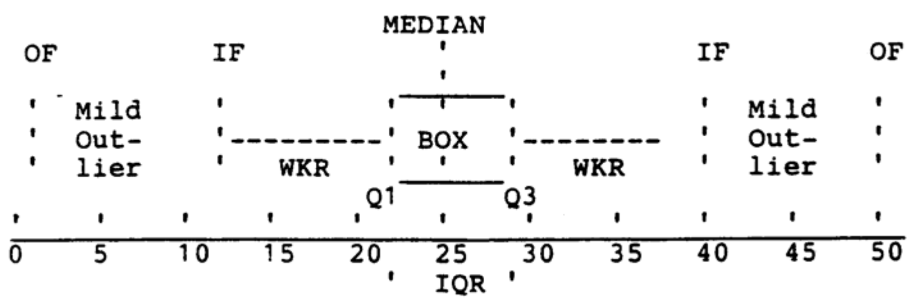


QENI 10-1-22

EXPLORING A SMALL DATA SET - 26 OBSERVATIONS

```
RANK ORDERING                    STEM AND LEAF
  13                               1 : 3
  17                               1 : 7,7
  17                               2 : 0,0,1,2,3,3,4
  20                               2 : 5,5,5,6,6,7,8,8,9,9
  20                               3 : 2,2,3,4
  21                               3 : 6,8
  22 - Q1 = 22
  23
  23                                   NOTES
  24
  25                             STEM - HORIZONTAL ROW
  25                             LEAF - A SINGLE NUMBER
  25                                    ON THE ROW
  26 - MEDIAN = 25.5             TRUNK- VERTICAL COLUMN
  26                                    LEFT OF :
  27                             MEDIAN IS UNDERLINED
  28
  28
  29
  29 - Q3 = 29
  32
  32              IQR = Q3 - Q1 = 7
  33
  34
  36
  38
```

BOX AND WHISKER PLOT



NOTES

1) Fence Distances are measured from the ends of the box.
2) The distance to the Inner Fence (IF) is 1.5 x IQR.
3) The distance to the Outer Fence (OF) is 3.0 x IQR.
4) Extreme Outlier Zones lie beyond the Outer Fences.
5) Wiskers (WKR) extend from the ends of the box to the largest observations inside the Inner Fences.
6) The box extends from Q1 to Q3.
7) Locate an Outlier by an * in the appropriate Zone.

MINTAGE QUANTITIES IN BILLIONS FOR LINCOLN MEMORIAL CENTS

YEAR	NUMBER
1989	12.6
1988	11.3
1987	9.6
1986	8.9
1985	10.9
1984	13.7
1983	14.2
1982	16.7
1981	12.9
1980	12.6
1979	10.1
1978	9.8
1977	8.7
1976	8.9
1975	10.0
1974	8.9
1973	7.6
1972	6.0
1971	5.3
1970	5.5
1969	5.5
1968	4.7
1967	3.0
1966	2.2
1965	1.5
1964	6.4
1963	2.5
1962	2.4
1961	2.5
1960	2.2
1959	1.9

**MY PENNY SAMPLE TAKEN FROM CIRCULATION
IN MID 1990 - 200 OBSERVATIONS**

```
1990 : OOOOO OOOOO OOOOO OOOOO OOOOO OOOOO OOOO
1989 : OOOOO OOOOO OOOOO QQOOO OOOOO OOO
1988 : OOOOO OOOOO OOOOO OOOO
1987 : OOOOO OOOOO OOO
1986 : OOOOO MM
1985 : OOOOO OOOOO O
1984 : OOOOO OO
1983 : OOOOO OOOOO
1982 : OOOOO OOOOO OOO
1981 : OOOOO OO
1980 : QQOO
1979 : OOOOO OO
1978 : OOO
1977 : OOOOO
1976 : OOO
1975 : OOOOO
1974 : OO
1973 : OOO
1972 : OO
1971 : O
1970 : OOO
1969 : OOO
1968 : OO
1967 :
1966 :
1965 : OO
1964 : OOOOO
1963 :
1962 :
1961 :
1960 :
1959 :
```

"SMOOTHING" THE DISTRIBUTION BY ADDING OBSERVATIONS

```
1990:00000000000000000000000000000000000000000000000000000
1989:00000000000000000000000000000000000000000000000000000000000
1988:000000000000000000000000000000000000000
1987:0000000000000000000000
1986:0000000000000000
1985:00000000000000000000
1984:000000000000000
1983:00000000000000
1982:0000000000000000000000000
1981:00000000000000
1980:00000000000000000
1979:000000000000000
1978:000000
1977:0000000000000
1976:000000
1975:00000000000
1974:0000
1973:0000000
1972:0000
1971:00000
1970:000000
1969:0000000
1968:0000
1967:
1966:0
1965:00
1964:000000
1963:0
1962:0
1961:0
1960:0
1959:0
```

NOTES

1) This data set was formed by combining two data sets of 200 coins each for a total of 400 coins.
2) Both data sets were retrieved from the same coffee can.
3) One of the data sets is J.S.
4) Note the 1964 "spike" is still present. What do you think happened in that year? Use the **RED BOOK!**

TOM TAKES HIS SAMPLE FROM A GRAVE
200 OBSERVATIONS

```
  1990 :
  1989 : OOOOO O
  1988 : OOOOO OOO
  1987 : OOOOO OOOOO OOOOO O
  1986 : OOOOO OOOOO O
  1985 : OOOOO OOOQQ OO
  1984 : OOOOO OOOOO
  1983 : OOOOO OOOOO OOOO
* 1982 : OOOOO OOOOO OOOOO OOOOO OOMMO OO
  1981 : OOOOO OOOOO OOOOO
  1980 : OOOOO OOOOO OO
  1979 : OOOOO OOO
  1978 : OOOOO OOOOO
  1977 : QQOOO
  1976 : OOOO
  1975 : OOOOO
  1974 : OOOOO O
  1973 : OOOOO O
  1972 : OOOOO
  1971 :
  1970 : OO
  1969 : OOO
  1968 : O
  1967 :
  1966 :
  1965 :
  1964 : OOO
  1963 : O
  1962 : OO
  1961 :
  1960 : OO
  1959 :
```

A POPULATION OF 493 WHEAT PENNIES LIVING IN AN OLD COFFEE CAN

1958 :	00000 00000 00000 00000 00000 0000	29
1957 :	00000 00000 00000 00000 00000 00000 000	33
1956 :	00000 00000 00000 00000 00000 000	28
1955 :	00000 00000 00000 00000	20
1954 :	000	3
1953 :	00000 00000 00000 00000 000	23
1952 :	00000 00000 00000 00000 00000 0000	29
1951 :	00000 00000 00000 00000 00000 00000 0000	34
1950 :	00000 00000 00000 000	18
		217

1949 :	00000 0000	9
1948 :	00000 0000	9
1947 :	00000 000	8
1946 :	00000 00000 00000 00000 00	22
1945 :	00000 00000 00000 00000 00000	25
1944 :	00000 00000 00000 00000 00000 00000 00420 00000 00	
1943 :	00	2
1942 :	00000 00	7
1941 :	00000 00000 00000 0	16
1940 :	00000 00000 0000	14
		154

1939 :	00000 00000	10
1938 :	00000 000	8
1937 :	00000 0000	9
1936 :	00000 000	8
1935 :	00000 00	7
1934 :	00000	5
1933 :		0
1932 :		0
1931 :		0
1930 :	0000	4
		51

In this example, the pennies are ELEMENTS and the date is the the OBSERVATION. What other OBSERVATIONS can we make using these ELEMENTS?

1929 :	00000 00	7
1928 :	00000	5
1927 :	00000	5
1926 :	00000 0	6
1925 :	00000	5
1924 :	00	2
1923 :	0000	4
1922 :		0
1921 :	0	1
1920 :	0000	4
		39

1919 :	00000 00000 00	12
1918 :	000	3
1917 :	00000 000	8
1916 :	0000	4
1915 :	00	2
1914 :		0
1913 :		1
1912 :		0
1911 :		0
1910 :	0	1
1909 :	0	1
		32

NOTES

1) Notice the steady decline by decade.
2) What was going on in 1943? In 1944? Use the RED BOOK!

RAW DATA FOR A RANDOM SAMPLE OF 300 PENNIES

	C1	C2	C3	C4	C5	C6
1)	1975	1980	1990	1978	1982	1980
2)	1967	1987	1988	1969	1974	1965
3)	1987	1988	1982	1988	1970	1986
4)	1989	1974	1977	1991	1988	1977
5)	1990	1975	1981	1989	1989	1976
6)	1990	1973	1979	1990	1972	1978
7)	1989	1990	1982	1991	1972	1982
8)	1960	1980	1982	1990	1960	1990
9)	1970	1985	1971	1985	1987	1981
10)	1987	1990	1969	1982	1984	1970
11)	1989	1967	1989	1962	1983	1970
12)	1990	1976	1982	1971	1990	1977
13)	1972	1985	1986	1960	1971	1989
14)	1981	1990	1982	1989	1982	1990
15)	1990	1978	1989	1990	1990	1976
16)	1978	1989	1985	1989	1982	1987
17)	1975	1990	1976	1989	1984	1982
18)	1976	1989	1990	1979	1981	1988
19)	1990	1986	1990	1988	1985	1990
20)	1983	1979	1984	1981	1990	1985
21)	1970	1974	1986	1986	1967	1985
22)	1967	1959	1982	1980	1989	1979
23)	1989	1988	1988	1974	1990	1975
24)	1987	1989	1963	1989	1989	1974
25)	1990	1990	1987	1989	1990	1980
26)	1984	1984	1984	1980	1960	1972
27)	1990	1988	1989	1969	1990	1984
28)	1985	1964	1979	1969	1960	1977
29)	1988	1972	1988	1989	1975	1975
30)	1981	1971	1982	1959	1969	1980
31)	1985	1988	1989	1971	1964	1985
32)	1984	1987	1986	1991	1977	1978
33)	1976	1989	1984	1990	1989	1990
34)	1990	1984	1968	1984	1967	1986
35)	1985	1984	1960	1986	1970	1984
36)	1990	1984	1990	1982	1990	1989
37)	1979	1973	1978	1971	1988	1983
38)	1981	1990	1990	1981	1979	1979
39)	1989	1987	1989	1988	1982	1987
40)	1981	1975	1987	1984	1990	1990
41)	1990	1982	1991	1981	1970	1990
42)	1981	1981	1988	1985	1984	1990
43)	1982	1971	1975	1967	1988	1985
44)	1990	1984	1990	1982	1988	1990
45)	1987	1988	1985	1981	1982	1988
46)	1979	1990	1989	1990	1980	1986
47)	19[illegible]8	1985	1983	1978	1982	1976
48)	1988	1977	1990	1973	1981	1983
49)	1990	1990	1986	1981	1976	1990
50)	1973	1985	1979	1982	1982	1980

A RANDOM SAMPLE OF 300 PENNIES COLLECTED BETWEEN JAN AND JUNE OF 1991

								CUMULATIVE TOTAL
1991:OOOO							4	4
1990:OOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOO							49	53
1989:OOOOOOOOOOOOOOOOOOOOQQOOOOOOO							29	82
1988:OOOOOOOOOOOOOOOOOOOOO							21	103
1987:OOOOOOOOOOOO							12	115
1986:OOOOOOOOOO						10	115	125
1985:OOOOOOOOOOOOOOOO						16		141
1984:OOOOOOOOMMOOOOOO						16		157
1983:OOOO						4		161
1982:OOOOOOOOOOOOOOOOOOOOOOO						23		184
1981:OOOOOOOOOOOOOOO					15	69		199
1980:OOOOOOOO					8			207
1979:OOOOOOOOOO					10			217
1978:OOOOOOOQ					8			225
1977:QOOOOO					6			231
1976:OOOOOOOO				8	47			239
1975:OOOOOOO				7				246
1974:OOOOO				5				251
1973:OOOOO				5				256
1972:OOOOO				5				261
1971:OOOOOOO			7	30				268
1970:OOOOOOO			7					275
1969:OOOO			4					279
1968:O			1					280
1967:OOOOOO			6					286
1966:		0	25					286
1965:O		1						287
1964:OO		2						289
1963:O		1						290
1962:OO		2						292
1961:	0	6						292
1960:OOOOOO	6							298
1959:OO	2							300
1958:	0							300
1957:	0							300
	8							

NOTES

1) DESCRIPTIVE STATISTICS is letting data "TALK TO YOU" using innovative arrangements. MM are the two data points strangling the MEDIAN and the QQ's are the data points strangling the QUARTILES. Replacing OO with MM or QQ is a creative way of letting data SPEAK FOR ITSELF!
2) The steady dying out of pennies with time is still with us!

HISTOGRAM HYSTERIA USING MY 300 PENNIES

```
     FIVE YEAR CELL WIDTH                  TEN YEAR CELL WIDTH
                               B                                 B
1991:00000000000000000000000   1  :00000000000000000000          1
1990:00000000000000000000000   1  :0000000000000000000F          1
1989:00000000000000000000000   1  :0000000000000000000F          1
1988:00000000000000000000000   1  :00000000000000000000          1
1987:00000000000000000000000   1  :0000000000000000000F          1
1986:00000000000000            2  :0000000000000000000F          1
1985:00000000000000            2  :00000000000000000000          1
1984:0000000000000F            2  :0000000000000000000F          1
1983:00000000000000            2  :0000000000000000000F          1
1982:00000000000000            2  :00000000000000000000          1
1981:000000000F                3  :000000000                     2
1980:0000000000                3  :000000000                     2
1979:000000000F                3  :00000000F                     2
1978:0000000000                3  :000000000                     2
1977:000000000F                3  :000000000                     2
1976:000000                    4  :00000000F                     2
1975:000000                    4  :000000000                     2
1974:000000                    4  :00000000F                     2
1973:000000                    4  :000000000                     2
1972:000000                    4  :000000000                     2
1971:00000                     5  :000                           3
1970:00000                     5  :000                           3
1969:00000                     5  :000                           3
1968:00000                     5  :000*                          3
1967:00000                     5  :000                           3
1966:0                         6  :000                           3
1965:0                         6  :000                           3
1964:0*                        6  :000                           3
1963:0                         6  :000                           3
1962:0                         6  :000                           3
1961:00                        7  :0                             4
1960:0F                        7  :0                             4
1959:00                        7  :0                             4
1958:0F                        7  :0                             4
1957:00                        7  :F                             4
1956:                          8  :F                             4
1955:                          8  :0                             4
1954:                          8  :0                             4
1953:                          8  :0                             4
1952:                          8  :0                             4
```

NOTES

1) HISTOGRAMS smooth data over BOXES which allows pattern recognition. The height of a box is simply the TOTAL NUMBER of observations in a five or ten year period divided by the box width. B is the box number.
2) The judicious use of filler (F) or shaving (*) is called "fitting the data" which helps the pattern recognition process. Note the LOSS OF YEAR BY YEAR DETAILS!

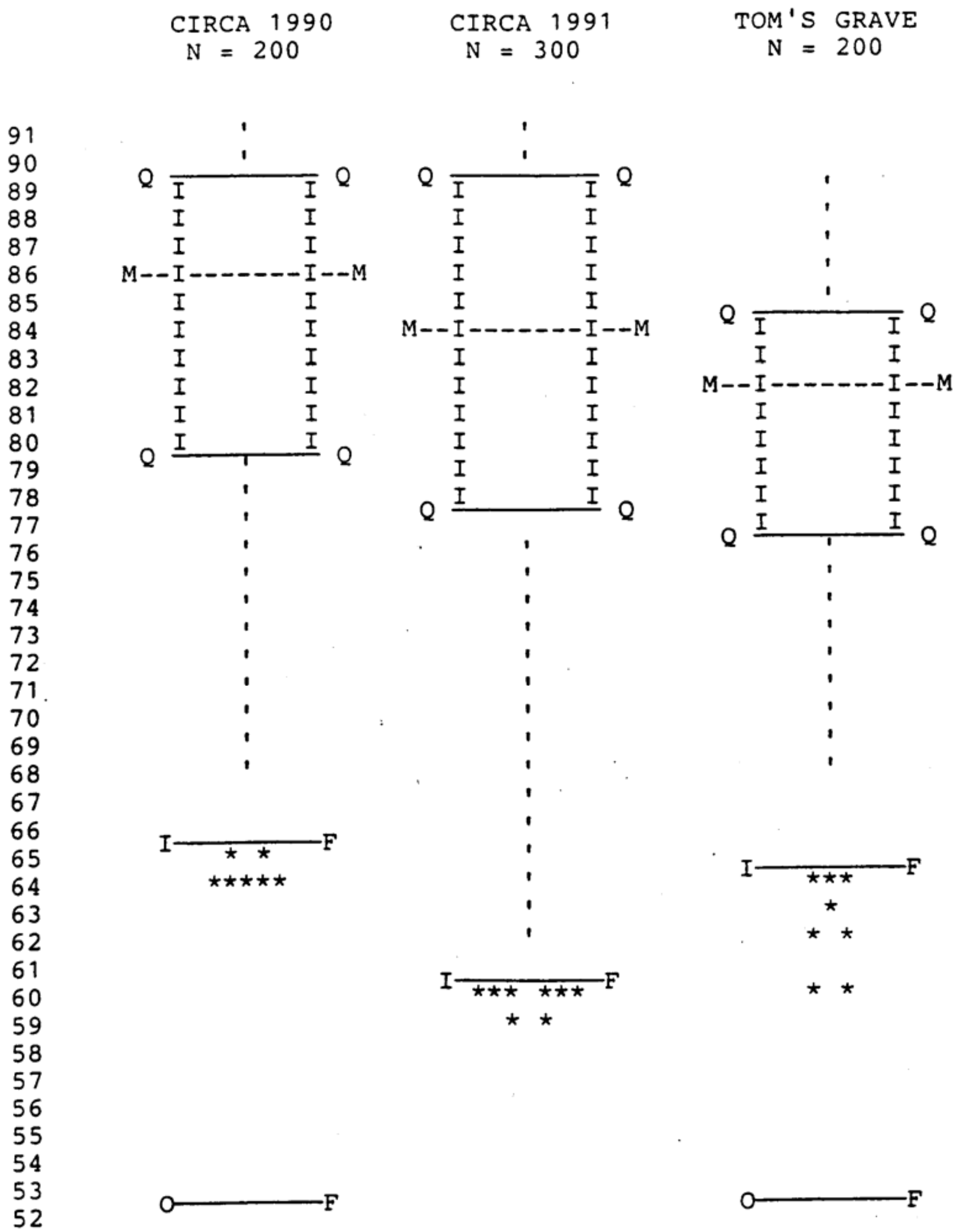
THE LIFE OF A PENNY - SAMPLE OF 300 OBSERVATIONS

DATE	AGE	NUMB	CUM	OLDR	%CUM	%OLDR
1991	0	4	4	296	1	99
1990	1	49	53	247	18	82
1989	2	29	82	218	27	73
1988	3	21	103	197	34	66
1987	4	12	115	185	38	62
1986	5	10	125	175	42	58
1985	6	16	141	159	47	53
1984	7	16	157	143	52	48
1983	8	4	161	139	54	46
1982	9	23	184	116	61	39
1981	10	15	199	101	66	34
1980	11	8	207	93	69	31
1979	12	10	217	83	72	28
1978	13	8	225	75	75	25
1977	14	6	231	69	77	23
1976	15	8	239	61	80	20
1975	16	7	246	54	82	18
1974	17	5	251	49	84	16
1973	18	5	256	44	85	15
1972	19	5	261	39	87	13
1971	20	7	268	32	89	11
1970	21	7	275	25	92	8
1969	22	4	279	21	93	7
1968	23	1	280	20	94	6
1967	24	6	286	14	95	5
1966	25	0	286	14	95	5
1965	26	1	287	13	96	4
1964	27	2	289	11	97	3
1963	28	1	290	10	97	3
1962	29	2	292	8	98	2
1961	30	0	292	8	98	2
1960	31	6	298	2	99	1
1959	32	2	300	0	100	0

NOTES

1) The above table is called a "MORTALITY TABLE".
2) Columns 4 and 5 always add to 300.
3) Columns 6 and 7 always add to 100%.
4) Percentages are obtained by dividing by 300.
5) This table can be used to predict penny survival chances versus penny age. This use of our SAMPLE is getting into the area of INFERENCE which is using SAMPLE data to make a statement about the "corresponding" POPULATION. We are always walking on ICE when we are doing an INFERENCE study. The wise statistician conducts INFERENCE studies that makes the ICE SAFE ---- but it is still ICE !

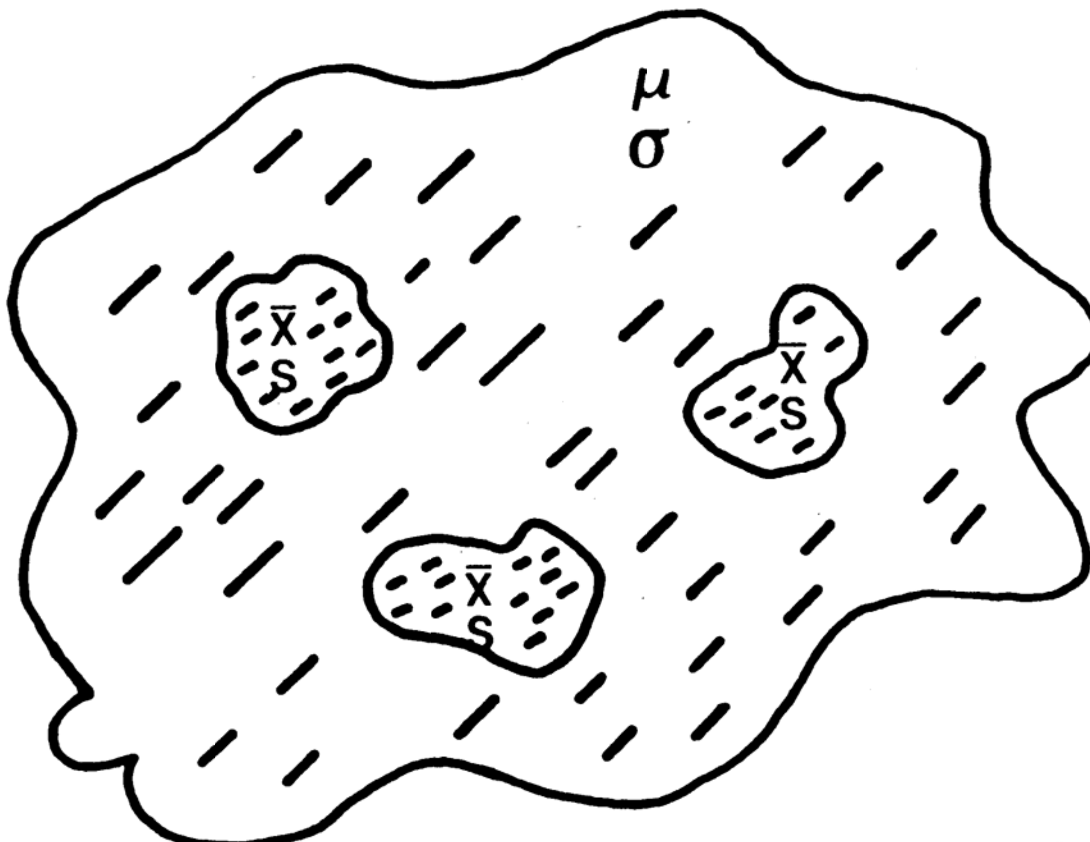
THREE LITTLE BOX AND WHISKERS



NOTES

1) N is SAMPLE SIZE. The Outer Fence for the middle Box and Whisker is at 1943 which is off the chart!
2) Grouping Box and Whisker plots is yet another way of making the DATA TALK TO YOU!

POPULATION AND SAMPLES



μ, σ ARE POPULATION PARAMETERS

X,S ARE SAMPLE STATISTICS

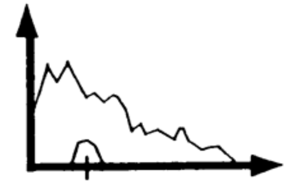
STATISTICS MAY BE USED TO ESTIMATE PARAMETERS

CENTRAL TENDENCY MEASURES

	MEDIAN	MEAN	MODE
COMPUTABILITY	EASY	HARD	FUZZY
PRO	• NOT INFLUENCED BY EXTREMES	• USES ALL DATA • CLT ($N \geq 30$)	• QUICKLY INDICATES WHERE ACTION IS
CON	• INSENSITIVE TO CHANGES IN DATA SET	• INFLUENCED BY EXTREMES	• MOB RULE

GENI 10-1-24

DISPERSION MEASURES



	RANGE	VARIANCE	STANDARD DEVIATION
COMPUTABILITY	EASY	HARD	HARD
PRO	• EASY TO GET A QUICK HANDLE ON DISPERSION	• USES ALL DATA • ANOVA	• USES ALL DATA • CLT ($N \geq 30$)
CON	• EASILY INFLUENCED BY EXTREMES!	• CONCEPTUALLY HARD TO GRASP	• REQUIRES ADJUSTMENT FACTORS IN ADVANCED STATISTICS

GENI 10-1-23

30

DEMING, STATISTICS AND QUALITY

An early proponent of the statistical Design of Experiment (DOX) philosophy in the United States was William Edwards Deming. A CENSUS BUREAU STATISTICIAN, Deming was a recognized expert in sampling and experimental design prior to World War II. One of his early masterworks, Some Theory of Sampling, was first published by John Wiley and Sons in 1950 and is still available as a Dover reprint. Soon after World War II, Deming took his expertise to Japan in order to help the Japanese people rebuild their devastated economy. As early as 1955, the Japanese were using DOX to solve industrial process problems. By 1960, Deming was an established name throughout Japan and the rest has become all too evident to the American public. Deming is now in his nineties and is rightfully revered by many as a sage. Still active through his writtings, Deming is a major force behind the current quality drive in our country. Deming's famous Forteen Points are stated underneath the Quality Diamond. Notice that all of these points address the human side of quality which must be in place before our statistical tools can work!

DEMING'S FAMOUS FORTEEN POINTS

Commitment
0

Teamwork 0 **QUALITY** 0 Statistical Tools

0
Voice of Customer

CREATING AN ENVIRONMENT FOR QUALITY IMPROVEMENT

1) Create constancy of purpose
2) Adopt the new philosophy
3) Cease dependence on inspection to achieve quality
4) End the practice of awarding business on price tag alone
5) Improve constantly the system of productivity and service
6) Institute training
7) Institute leadership
8) Drive out fear
9) Break down barriers between departments
10) Eliminate slogans and substitute "how to"
11) Eliminate quotas
12) Remove barriers that rob people of pride of workmanship
13) Encourage education and self improvement for everyone
14) Put everyone in the company to work to accomplish the transformation

Without the above environment,

the use of statistical techniques

IS FUTILE !

Employees collect ELEMENTS and make OBSERVATIONS for each and every one of them. Write down the NUMERICAL VALUES corresponding to your OBSERVATIONS. The numbers are RANDOM in the sense that you should have no control over them! What you see is what you get! There are several things you must do with the numbers. Find the MODE which is the number which occurs most often. RANK ORDER which means lining the numbers up one by one starting from the smallest and ending with the largest. Calculate the RANGE which is the difference between the largest and smallest numbers. Calculate the MEDIAN which is the "middle" number in your lineup. Also calculate a number's "PERCENTILE RANK" within the lineup. For example, if a number is in the fourth position in a lineup of twelve numbers, give this number a PERCENTAGE RANK of 33 %. Other things you must do are given by the formulas that follow. In each formula, N is the number of observations and $X_1, X_2, X_3, - - - X_N$ are the individual numbers.

CALCULATE THE MEAN

$$\frac{X_1 + X_2 + X_3 + --- + X_N}{N}$$

CALCULATE THE VARIANCE

$$\frac{(X_1 - MEAN)^2 + (X_2 - MEAN)^2 + --- + (X_N - MEAN)^2}{NUM}$$

CALCULATE THE STANDARD DEVIATION

The STANDARD DEVIATION or SD is simply SQR(VARIANCE)

EXPLANATORY NOTES MAY BE FOUND ON THE NEXT PAGE

IF YOU REQUIRE FURTHER EXPLANATION.

NOTES

1) NUM = N if you are using your numbers as a POPULATION. This means we are only interested in the numbers we have and no others.
2) NUM = N - 1 if you are using your numbers as a SAMPLE. SAMPLES are used to make PREDICTIONS about POPULATIONS. SAMPLE numbers are usually a very small percentage of all the possible numbers that can come from the PARENT POPULATION!

THIS TRAINING PROGRAM HAS BEEN CAREFULLY PREPARED BY THE CENSUS BUREAU. WHEN YOU ARE DONE READING THIS, REPORT TO DUTY!

CALCULATION OF MEAN AND STANDARD DEVIATION FOR MY 300 PENNIES

MEAN DATE CALCULATION

1959 +	32	x	4 =	128
	31	x	49 =	1519
	30	x	29 =	870
	29	x	21 =	609
	28	x	12 =	336
	27	x	10 =	270
	26	x	16 =	416
	25	x	16 =	400
	24	x	4 =	96
	23	x	23 =	529
	22	x	15 =	330
	21	x	8 =	168
	20	x	10 =	200
	19	x	8 =	152
	18	x	6 =	108
	17	x	8 =	136
	16	x	7 =	112
	15	x	5 =	75
	14	x	5 =	70
	13	x	5 =	65
	12	x	7 =	84
	11	x	7 =	77
	10	x	4 =	40
	9	x	1 =	9
	8	x	6 =	48
	7	x	0 =	0
	6	x	1 =	6
	5	x	2 =	10
	4	x	1 =	4
	3	x	2 =	6
	2	x	0 =	0
	1	x	6 =	6
	0	x	2 =	0
			300	6876

VARIANCE CALCULATION

$(32 - 22.93)^2 =$	82.26	x	4 =	329.04
$(31 - 22.93)^2 =$	65.12	x	49 =	3190.88
$(30 - 22.93)^2 =$	49.98	x	29 =	1449.42
$(29 - 22.93)^2 =$	36.84	x	21 =	773.64
$(28 - 22.93)^2 =$	25.70	x	12 =	308.40
$(27 - 22.93)^2 =$	16.56	x	10 =	165.60
$(26 - 22.93)^2 =$	9.42	x	16 =	150.72
$(25 - 22.93)^2 =$	4.28	x	16 =	68.48
$(24 - 22.93)^2 =$	1.14	x	4 =	4.56
$(23 - 22.93)^2 =$	.0*	x	23 =	.11
$(22 - 22.93)^2 =$	.86	x	15 =	12.90
$(21 - 22.93)^2 =$	3.72	x	8 =	29.76
$(20 - 22.93)^2 =$	8.58	x	10 =	85.80
$(19 - 22.93)^2 =$	15.44	x	8 =	123.52
$(18 - 22.93)^2 =$	24.30	x	6 =	145.80
$(17 - 22.93)^2 =$	35.16	x	8 =	281.28
$(16 - 22.93)^2 =$	48.02	x	7 =	336.14
$(15 - 22.93)^2 =$	62.88	x	5 =	314.40
$(14 - 22.93)^2 =$	79.74	x	5 =	398.70
$(13 - 22.93)^2 =$	98.60	x	5 =	493.00
$(12 - 22.93)^2 =$	119.46	x	7 =	836.22
$(11 - 22.93)^2 =$	142.32	x	7 =	996.24
$(10 - 22.93)^2 =$	167.18	x	4 =	668.72
$(9 - 22.93)^2 =$	194.04	x	1 =	194.04
$(8 - 22.93)^2 =$	222.90	x	6 =	1337.40
$(7 - 22.93)^2 =$	253.76	x	0 =	0
$(6 - 22.93)^2 =$	286.62	x	1 =	286.62
$(5 - 22.93)^2 =$	321.48	x	2 =	642.96
$(4 - 22.93)^2 =$	358.34	x	1 =	358.34
$(3 - 22.93)^2 =$	397.20	x	2 =	794.40
$(2 - 22.93)^2 =$	438.06	x	0 =	0
$(1 - 22.93)^2 =$	480.92	x	6 =	2885.52
$(0 - 22.93)^2 =$	525.78	x	2 =	1051.56
				18714.17

CENTRAL TENDENCY MEASURES

MEAN = 1959 + 6876/300 = 1981.93
MEDIAN = 1984
MODE = 1990

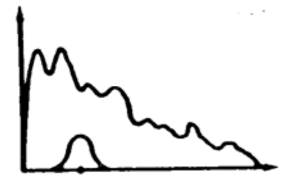
MEASURES OF DISPERSION

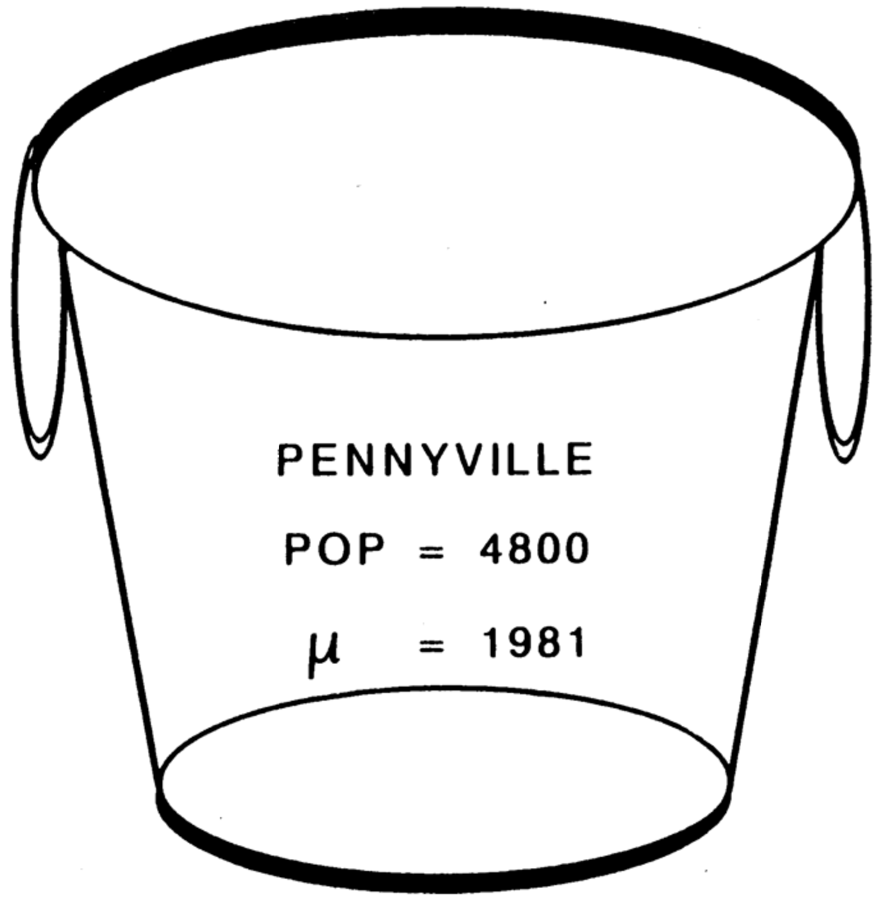
VARIANCE = 18714.17/300 = 62.38 (POPULATION)
VARIANCE = 18714.17/299 = 62.59 (SAMPLE)
STANDARD DEVIATION = SQR(VAR) = 7.91 (either)
RANGE = 1991 - 1959 = 32

STANDARD PENNY SAMPLES OF
200 OBSERVATIONS

NAME	MEAN	SD	MEDIAN	MODE
K.G	1982.8	6.8	1984.5	1990
S.C.	1981.7	6.3	1984	1989
M.S.	1981.2	6.3	1982	1984
R.A.	1981.3	6.1	1983	1988
B.A.	1980.4	-	-	-
R.R.	1981.7	6.9	1982	1988
K.C.	1980.3	-	1982	1988/84
N.L.	1982.7	6.6	1984	1989
S.C.	1981.2	7.3	1984	1989
T.M.	1982.1	8.4	1985	1989
V.G.	1981	4.5	1983	1989
B.K.	1981.3	7.4	1983	1989
N.S.	1980	7.5	1982	1988
B.S.	1983	6.3	1984	1989
M.M.	1981	7.7	-	1989
J.S.	**1983.5**	**6.8**	**1986**	**1990**
J.L.	1980.4	3.8	1982	1982
T.L.	**1979.7**	**6.3**	**1982**	**1982**

PENNYVILLE WELCOMES YOU





RULES IN THE POOL

We have just founded Pennyville by pooling our resources. Each of us poured exactly 200 pennies into the old washtub sitting on the floor. How many residents does Pennyville have? The answer is simply the number of samples (call this number K) times 200. Pennyville usually has over 5000 residents. Other demographics such as the village mean and standard deviation can be computed using the following pooling rules.

VILLAGE OR GRAND MEAN

This is simple. Simply add up all the sample means and divide by the number of samples. This formula assumes a common sample size. CAUTION: THIS FORMULA DOES NOT HOLD IF WE HAVE POOLED SAMPLES HAVING VARIOUS SIZES!

POOLED VARIANCE

Again, this is fairly simple since our samples are 1) very large and 2) have a common sample size. In this case, the pooled variance is simply the average of the sample variances. CAUTIONS: THIS FORMULA ASSUMES 1) LARGE SAMPLES, 2) SAMPLES HAVING A COMMON SIZE AND 3) RANDOM SAMPLES COMMING FROM THE SAME SOURCE. NUMBER 3) IS SAYING WE CAN'T MIX GRAVEYARDS AND FRESH MEAT AND COME UP WITH A MEANINGFUL POOLED VARIANCE!

POOLED STANDARD DEVIATION

Assumming we have a meaningful pooled variance, the pooled standard deviation is simply the square root of the pooled variance.

CHEBYSHEV'S RULE

Let a population/sample have any distribution imaginable including "chaos". Calculate the MEAN and SD. Pick any number you want bigger than one. Call this number K. The following is true :

$$\text{at least } 1 - (1/K)^2 \text{ PERCENT}$$

of the observations will huddle between the boundaries

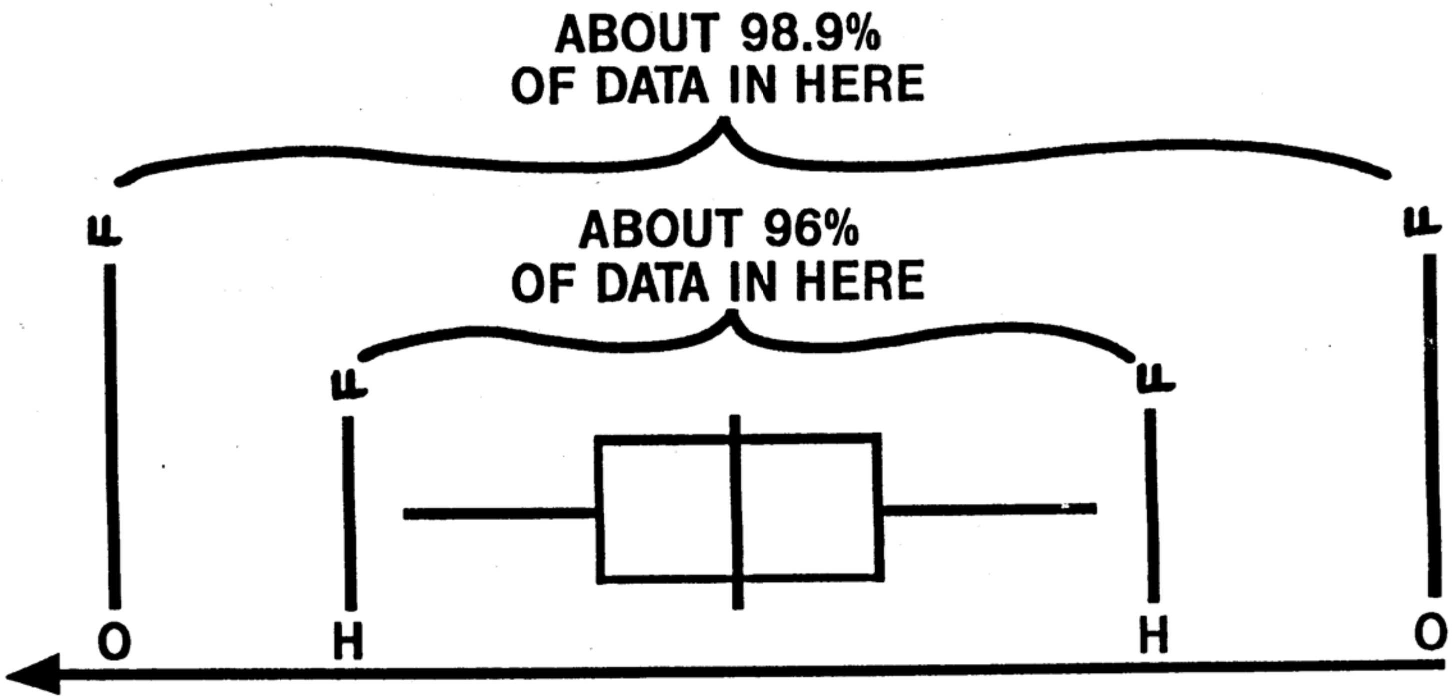
MEAN - K x SD and MEAN + K x SD.

NOTES ON CHEBYSHEV

1) K does not have to be a whole number!
2) Chebyshev sets a conservative LOWER LIMIT. Most distributions will " pack the critters in tighter", but we can not guarantee it.
3) Percentage limits for the INNER and OUTER FENCES of a BOX and WHISKER PLOT can be established by Chebyshev's Rule. Roughly 96 % of the observations huddle between the two INNER FENCES and roughly 98.9 % of the observations huddle between the two OUTER FENCES.
4) Chebyshev's Rule is DISTRIBUTION FREE meaning it will work for any distribution whatsoever!

PUTTING TEETH ON THE BOX AND WHISKER

PROOF: BY CHEBYSHEV



ESTIMATE IMPROVES AS SAMPLE SIZE INCREASES!

JENGA DEMONSTRATION

JENGA is a game made by Milton-Bradley. Each game set consists of 54 "identical" wooden blocks. The JENGA game is used to explore several statistical ideas. We are first going to build a data base. This data base will then be analyzed by all our investigative tools developed today. This session closes by taking a look at goal setting and quality improvement.Tomorrow, we will use the same data base to explore a great statistical idea which governs the formation of bell shaped distributions.

The demonstration itself is simple. We are going to make towers from our JENGA blocks and record the number of blocks successfully stacked before the tower tumbles and falls down on the table. Each table will have four participants and willbe responsible for eight trials. Blocks are to be stacked in the center of the table. Blocks are to be stacked with the flat side horizontal and with no outside support. The four participants will take rotating turns stacking the blocks, one block per turn. Turn rotation will be clockwise. The tower is not to be steadied before or after the block is placed. Once the tower falls, record the number of blocks successfully stacked on a 3X5 card and give the card to the instructor. Repeat foreight trials and take a short break.

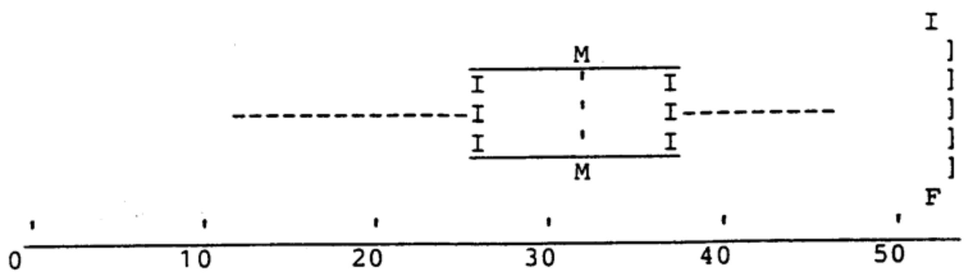
THIS	NOT THIS
-	-
-	-
-	--
-	--
-	---
-	---
-	---

QUALITY IMPROVEMENT PROCESS

1) ASSESS CURRENT CAPABILITY

10:2
10:7
20:0,2,2
20:5,5,5,6,7,9,9,9,9
30:0,1,1,2,2,2,3,3,4,4,4
30:5,7,8,8,9
40:0,1,2,2
40:7

2)GOAL SETTING



3) GOAL ATTAINMENT

20:5,6,6,7,7,9,9
30:0,0,1,1,3,3
30:6
40:
40:4,5,6,6,8,9
50:0,1,4

NOTES

1) The goal of 54 was established by the box and whisker plot.
2) Once established, the goal was achieved!
3) An overall quality improvement was a byproduct of the competition process.

SUGGESTED HOMEWORK PROBLEMS

1) 1982 was a very good year! The following gives the 1982 coin count for each of the 28 samples provided by my class in the Fall of 1990.

11	14	15	27	21	17	16
07	12	14	07	18	19	13
06	12	16	11	14	23	13
20	21	12	13	07	13	12

A) Construct a stem and leaf plot using 0 0 1 1 2 2 as the trunk.

B) Calculate the MEAN, MEDIAN, MODE and SD.

C) Construct a box and whisker plot.

D) Verify Chebyshev's Rule for the following values of K: 1.25, 1.5, 2.0 and 3.0.

E) Determine if 27 is a mild outlier. If it is, offer some rationale for this phenomena. HINT: GRAVEYARD

2) My wife is a bargain hunter. She came home with a bag of seven "big" apples purchased for $1.40 at a local market. A runt was found in the bottom of the bag. A clever arrangement! Sure enough, the weights turned out to be .67, .43, .17, .5, .55, .47 and .58 pounds. "Do something", Carolyn said! A quick "box and whisker" proved the point and Carolyn marched right back to the market and recovered our 20 cents. How did I establish this "runtness"? What are some possible holes in my argument? HINT: CONVICTED BY SEVEN DATA POINTS!

THE INTUITIVE
CENTRAL LIMIT THEOREM

Suppose a distribution is generated from observations where each individual observation is the additive effect (or output) of "several" equally contributing causes (or factors). The resulting distribution will "shape up" to the "Old Bell Curve" provided we have enough observations to "flesh the curve out". The center (mean) of the bell curve will be determined by the basic plan or physics governing the action.

COMMENTS

The three data sets shown below were generated in the Fall of 1990. Note that all three stem and leaf plots are trying to conform to "The Old Bell Curve". An indicator of this (other than visual) is the close match between all three measures of central tendency (MEAN, MEDIAN and MODE). Compare these three data sets with any of our penny samples. What do you notice about the MEAN, MEDIAN and MODE for pennies?

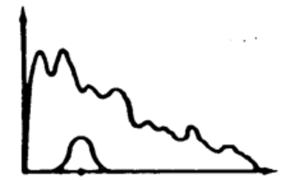
JENGA STEM AND LEAF	SUMMARY STATISTICS
1 : 3	SIZE = 26
1 : 7,7	MEAN = 25.9
2 : 0,0,1,2,3,3,4	MEDIAN = 25.5
*2 : 5,5,5,6,6,7,8,8,9,9	MODE = 25
3 : 2,2,3,4	SD = 6.1
3 : 6,8	

FSHOE STEM AND LEAF	SUMMARY STATISTICS
6 : 0,0	SIZE = 17
6 : 5,5,5	MEAN = 7.4
7 : 0	MEDIAN = 7.5
*7 : 5,5,5,5,5,5	MODE = 7.5
8 : 0,0	SD = .88
8 : 5,5	
9 : 0	

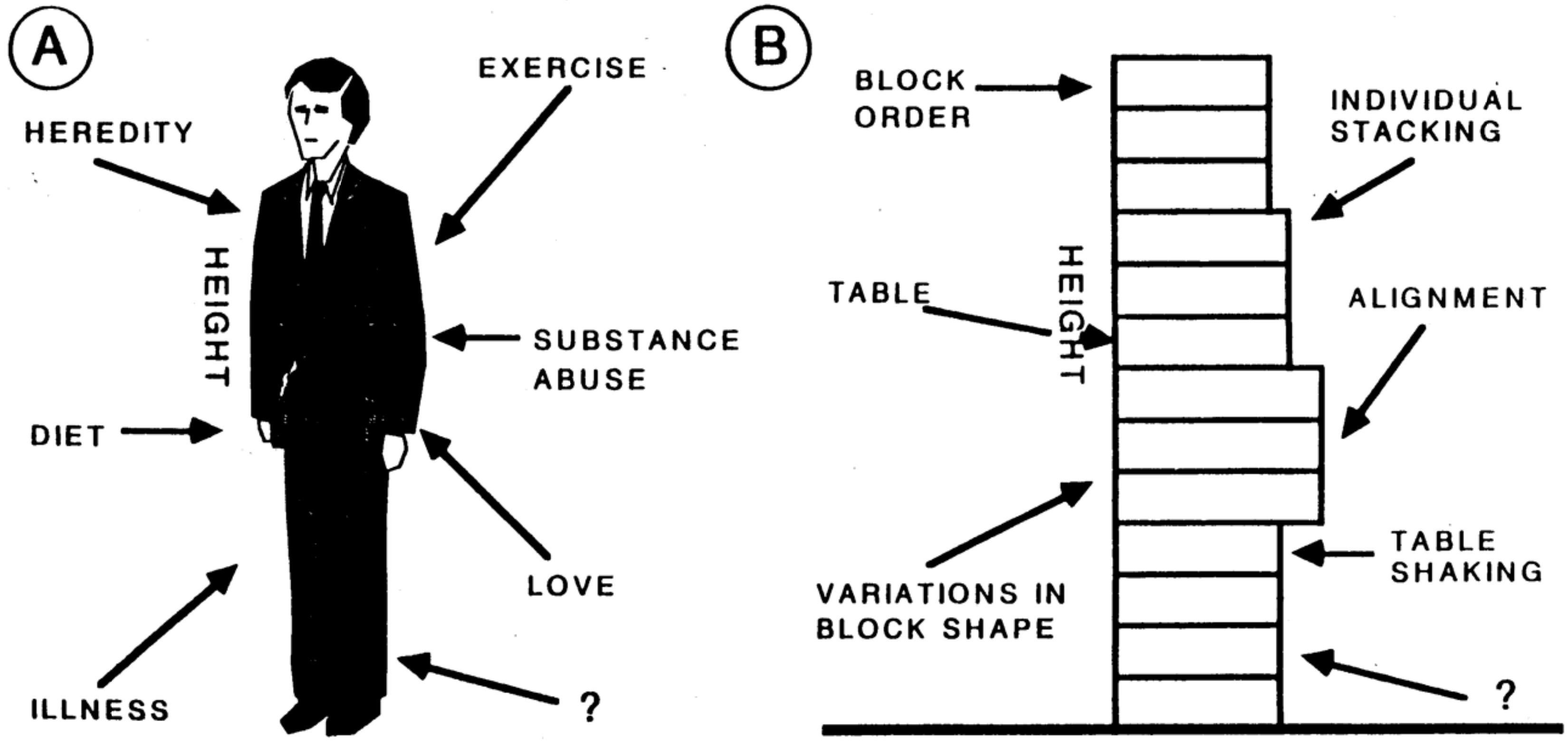
FHGHT STEM AND LEAF	SUMMARY STATISTICS
6 : 0,1	SIZE = 17
6 : 2,2,3,3	MEAN = 64.5
*6 : 4,4,4,5,5	MEDIAN = 64
6 : 6,6,7	MODE = 64
6 : 8,8,9	SD = 2.6

* What "sets the center"!

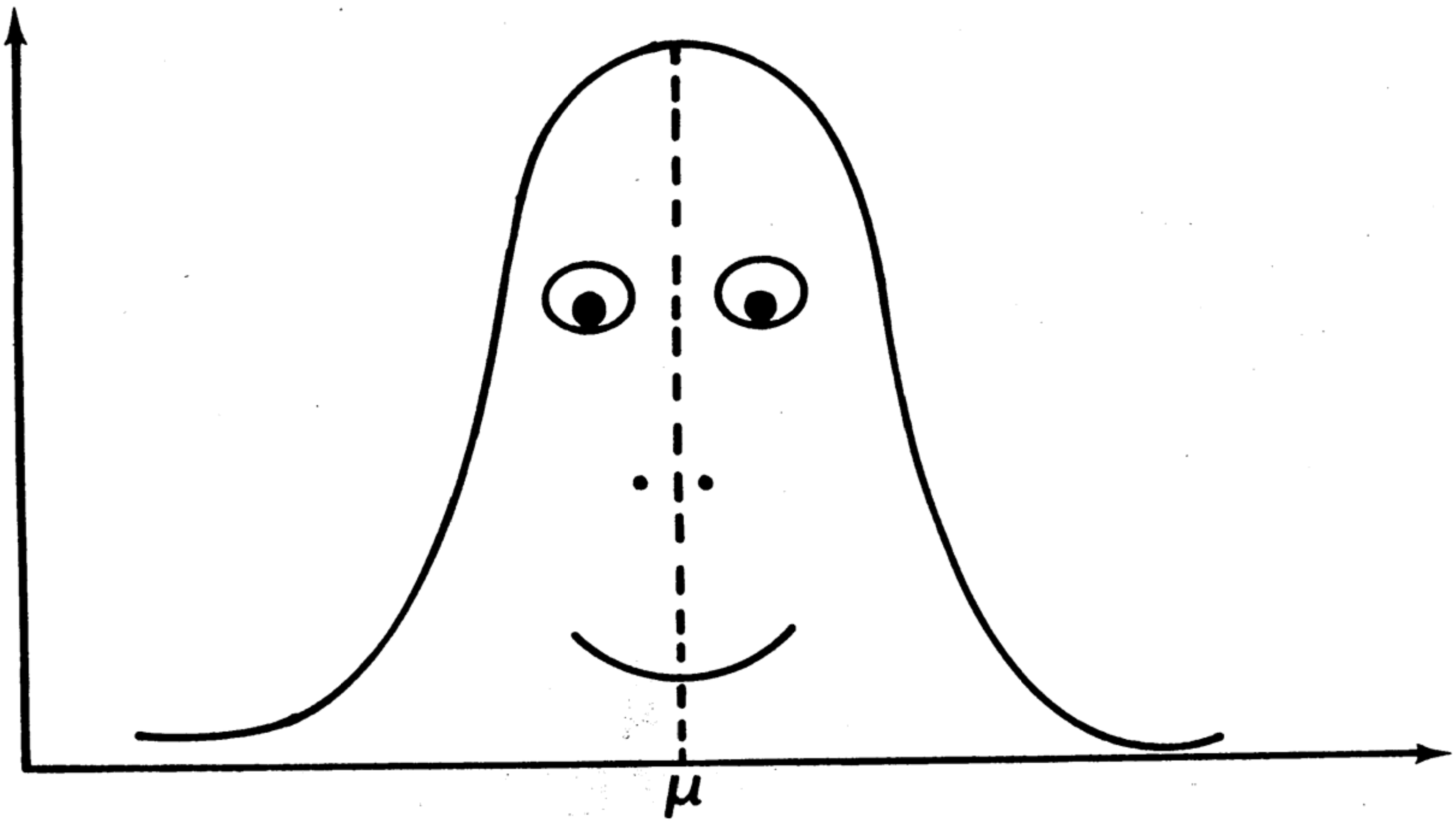
QUALITATIVE STATEMENT OF THE CENTRAL LIMIT THEOREM



LET A DISTRIBUTION BE MADEUP OF MANY OBSERVATIONS WHERE EACH OBSERVATION IS THE CUMULATIVE RESULT OF A NUMBER OF ROUGHLY EQUALLY CONTRIBUTING FACTORS, THEN, THE DISTRIBUTION FOLLOWS THE BELL CURVE !

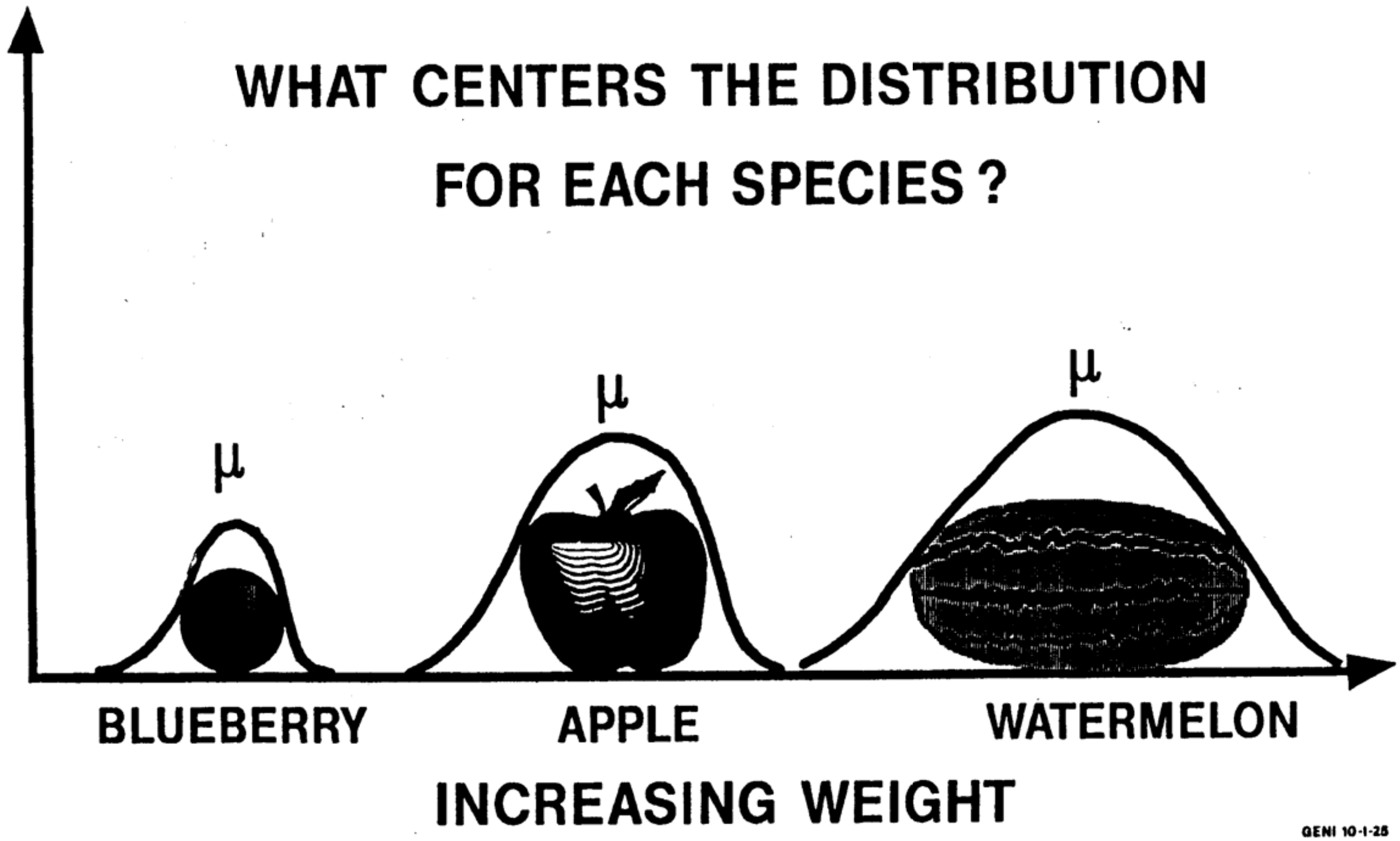


NORMAL



HELLO ! I'M THE BELL CURVE

LINING UP THE FRUIT



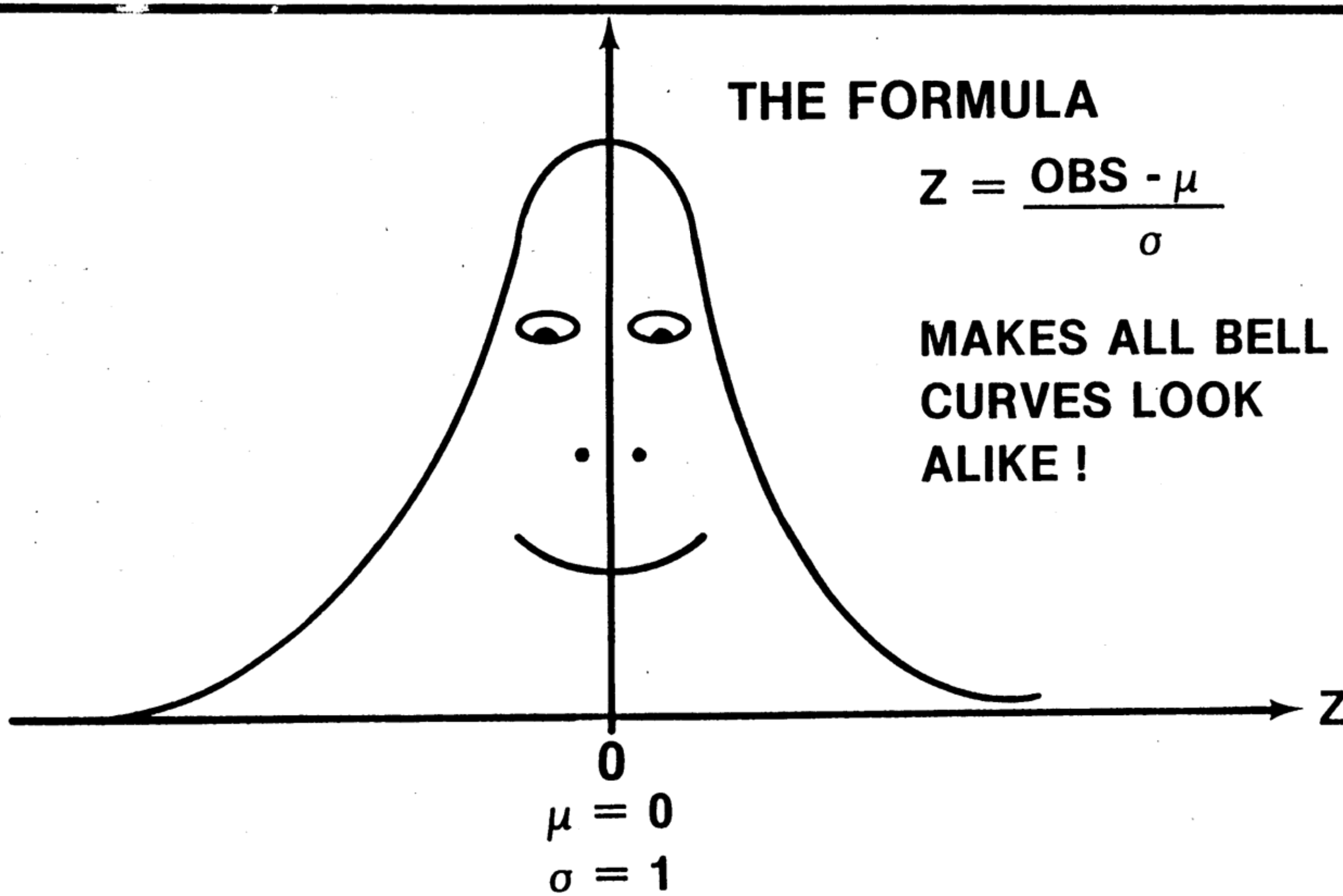
GENI 10-1-25

The STANDARD NORMAL distribution is simply a special edition of the BELL CURVE where the MEAN is zero and the SD is one. Any set of numbers that track the BELL CURVE can also be made to track the STANDARD NORMAL distribution. The tracking mechanism which converts our BELL CURVE numbers to STANDARD NORMAL numbers is called a Z-SCORE. Here is how it is done. Compute the MEAN and SD for any set of numbers that track the BELL CURVE. Let OBS be any individual observation in the original data set. The following is true. Z where

$$Z = \frac{OBS - MEAN}{SD}$$

tracks the STANDARD NORMAL distribution. The individual Z values or Z-SCORES measure the distance between OBS and MEAN in terms of SD multiples. Z-SCORES can be positive or negative. The larger the Z-SCORE is in absolute value, the rarer the associated observation becomes. Outliers usually have Z-SCORES starting at + 2 or less than - 2. More extreme outliers will have more extreme Z-SCORES! TABLE A-3 (one of the handouts) associates a Z-SCORE and the percentage of observations that huddle inside the Z-SCORE boundary. The same table also can be made to tell us the percentage of observations on the outside of the Z-SCORE boundary which is a good indication of rarity!

STANDARD NORMAL



THE FORMULA

$$Z = \frac{OBS - \mu}{\sigma}$$

MAKES ALL BELL CURVES LOOK ALIKE !

EMPIRICAL RULE

Let a population/sample have a bell or mound shaped distribution. Calculate the MEAN and SD. The following statements are true:

1) at least 68 % of the observations huddle between the boundaries

MEAN - SD and MEAN + SD

2) at least 95 % of the observations huddle between the boundaries

MEAN - 2 x SD and MEAN + 2 x SD

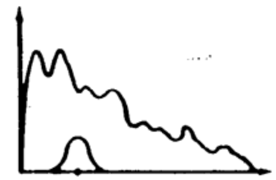
3) at least 99 % of the obsevations huddle between the boundaries

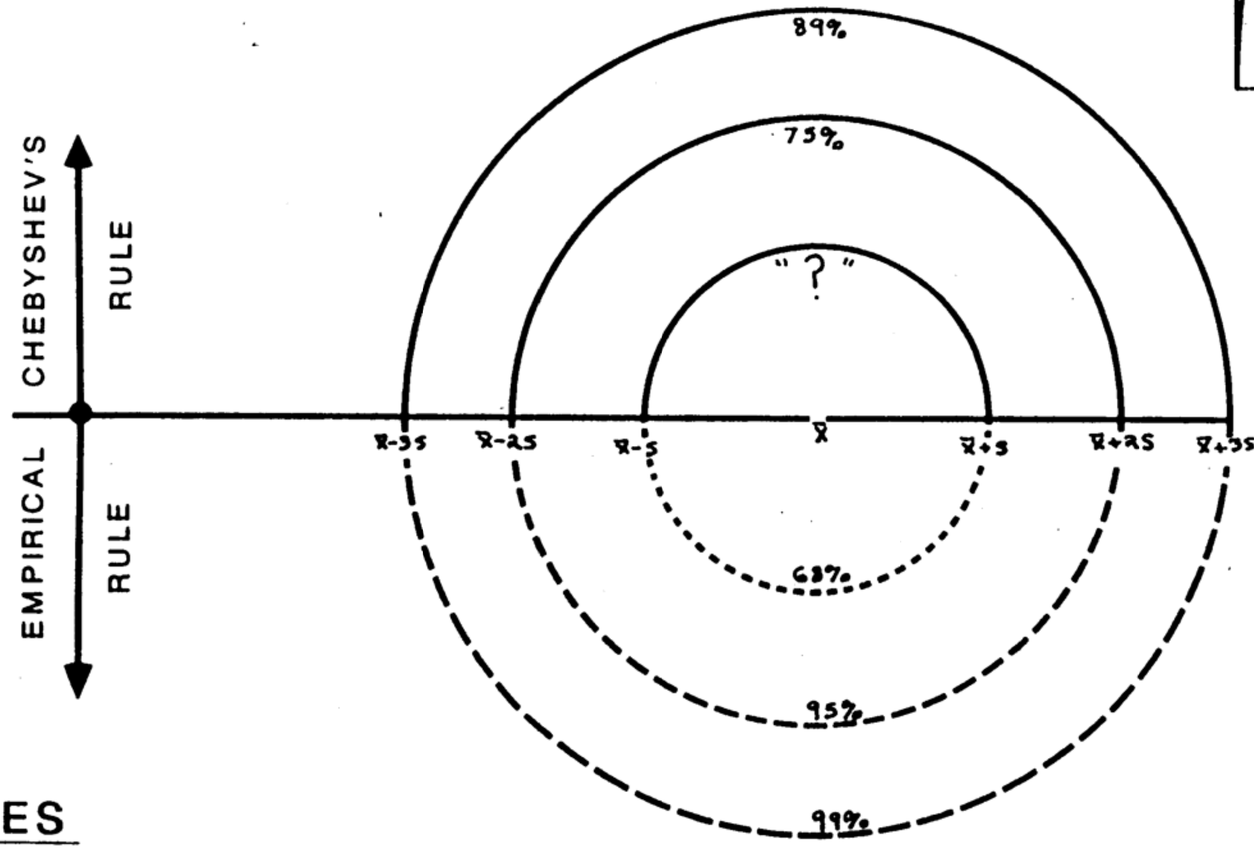
MEAN - 3 x SD and MEAN + 3 x SD.

NOTES ON EMPIRICAL

1) The distribution MUST BE MOUND OR BELL SHAPED for the Empirical Rule to hold.
2) The Empirical Rule will give tighter estimates of distribution "packing" than the Chebyshev Rule provided we can use it!
3) We can use the boundaries erected by the Empirical Rule to establish OUTLIER ZONES for BELL SHAPED CURVES. Mild outliers are on the outside of the 95% boundary and extreme outliers are on the outside of the 99% boundary
4) The Empirical Rule IS NOT DISTRIBUTION FREE! WE MUST HAVE A BELL CURVE FOR IT TO WORK!
5) The Empirical Rule can be used as an investigative tool to see if a distribution is following a BELL SHAPED PATTERN.

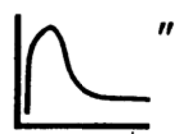
COMPARISON OF CHEBYSHEV TO EMPIRICAL

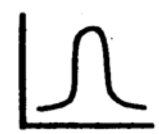




NOTES

1) CHEBYSHEV'S RULE IS GOOD FOR ANY DISTRIBUTION SHAPE

THIS "  " OR "  " OR "  " ETC.

2) EMPIRICAL RULE IS ONLY GOOD FOR MOUND OR BELL SHAPED DISTRIBUTIONS ! "  "

THE CENTRAL LIMIT THEOREM
or
"THE HEART OF IT ALL"

Let a population have any wierd distribution whatsoever like the penny samples we brought to this workshop. Suppose we draw lots of samples (with replacement) from our population where all sample sizes are identical. Calculate the MEAN for each sample. THE SAMPLING DISTRIBUTION OF THESE MEANS WILL LINE UP WITH THE "OLD BELL CURVE" IF THE COMMON SAMPLE SIZE IS THIRTY SOMETHING! The MEAN of the SAMPLING DISTRIBUTION will be identical to the Population MEAN. The Standard Deviation of the SAMPLING DISTRIBUTION is simply the Population SD divided by the square root of the common sample size.

AN ACTUAL CLT DEMONSTRATION

The following stem and leaf display is actual data generated via an in-class demonstration of the Central Limit Theorem. Means were computed for 68 "random" samples of 30 "thirty something" coins each . Sampling was with replacement within the confines of the 5% rule. The underlying population consisted of about 2500 pennies collected over a 10 year period.

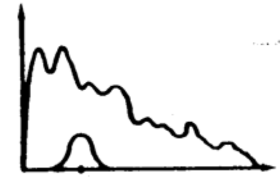
```
1973. : 50
1974. :
1975. : 40,46
1976. : 40,50,73,80
1977. : 23,36,60,60,60,77
1978. : 00,16,23,30,43,50,50,60,70,73,83,90,90,97     LET THE
1979. : 13,13,30,36,36,40,50,57,60,63,67,70,73,73   --MODAL CLASS
1980. : 00,06,23,36,37,38,40,50,60,63,70,80,86,97     BE THE
1981. : 10,37,56,60,60,60,60,76,83                    LONG STEM
1982. : 20,40,83,97
1983. :
1984. :
```

SAMPLING DISTRIBUTION STATISTICS

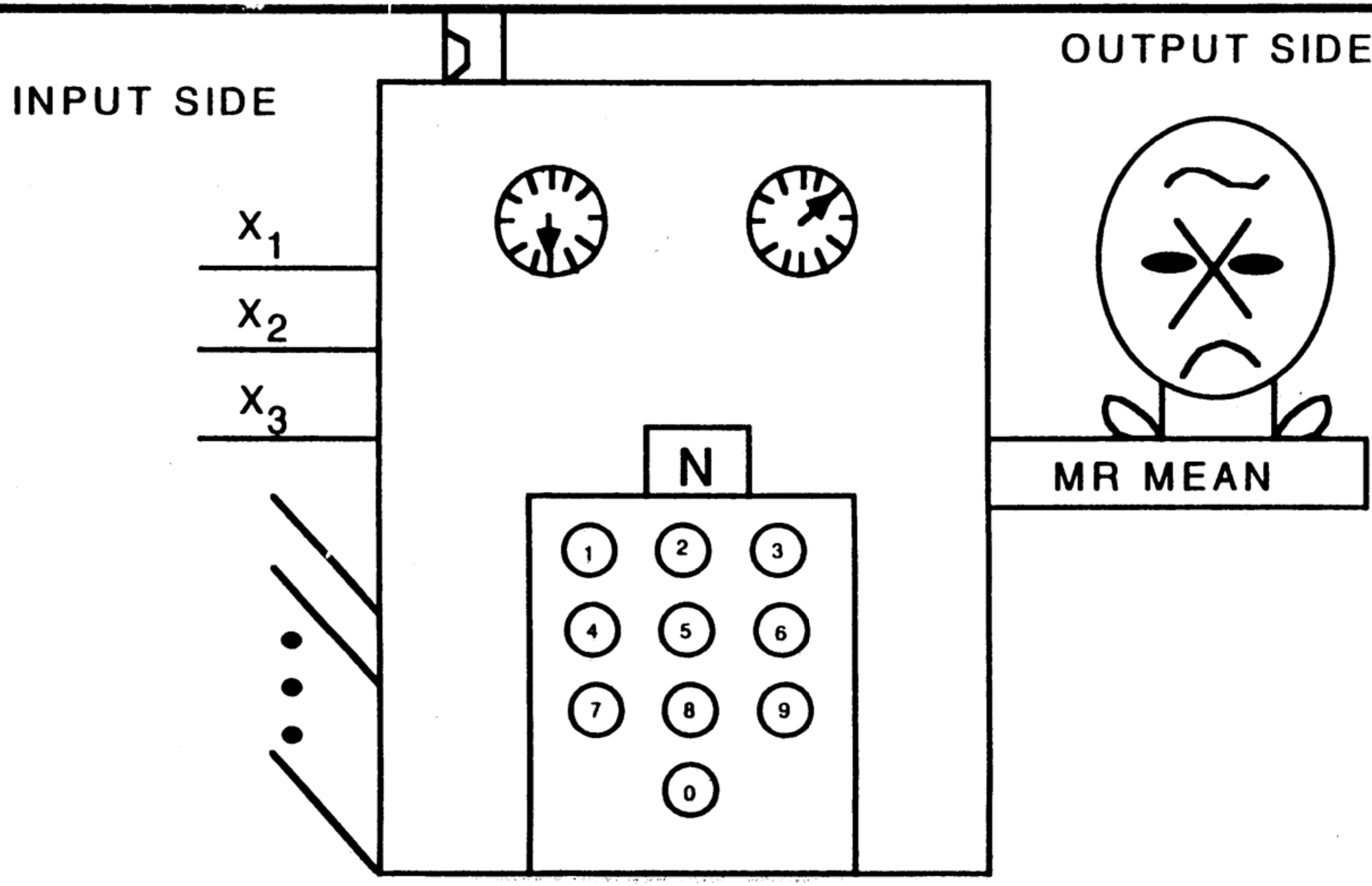
```
MEAN   = 1979.408
MEDIAN = 1979.535 (Use Underlined Values)
MODE   = 1979 STEM
SD     = 1.8613
```

NOTES

A good estimate of the POPULATION MEAN is 1979.408. WHY! Since SD/SQR(30) = 1.8613, a good estimate of the POPULATION STANDARD DEVIATION is 10.1946. WHY!

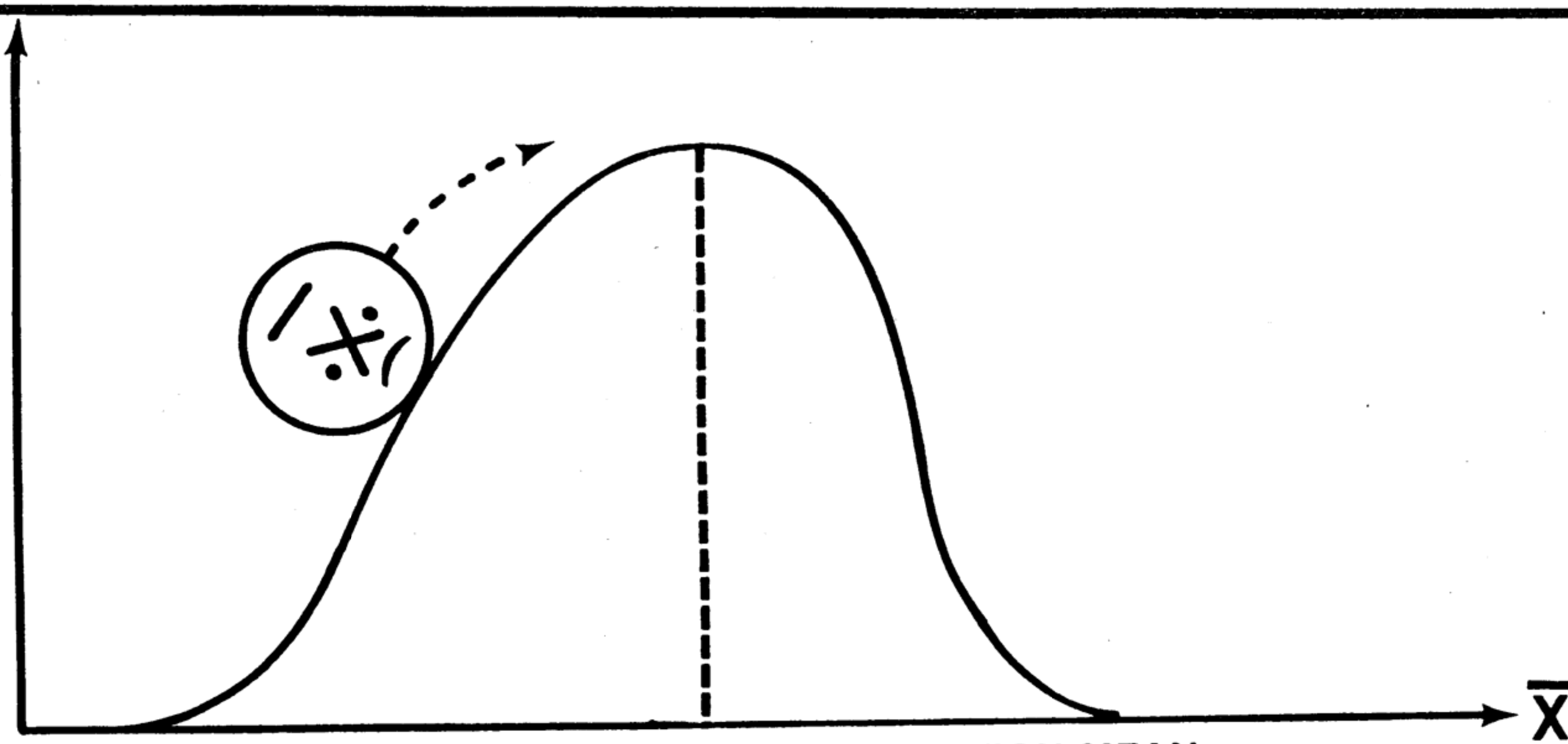


THE MEAN MACHINE



MAKING MEANS WHEN N IS 3

MR MEAN RIDES THE BELL CURVE



μ = POPULATION MEAN

σ = POPULATION STANDARD DEVIATION DIVIDED BY THE SQUARE ROOT OF THE SAMPLE SIZE N

N = THIRTY SOMETHING

TWO MOTHERLODE EXPRESSIONS

The CENTRAL LIMIT THEOREM is to statistics as the PYTHAGOREAN THEOREM is to geometry. Since our High School days, we have been bombarded with the expression

$$A^2 = B^2 + C^2.$$

This expression has been a MOTHERLODE that has brought with it a thousand and one applications in all areas of science and technology. It is part of our cultural heritage in the West!

The CENTRAL LIMIT THEOREM guarantees that $MEAN_S$ (SAMPLE MEAN) tracks the BELL CURVE as long as the sample size N is THIRTY SOMETHING.

Recall SD_{MEAN} is simply $SD_P/SQR(N)$.

Since we usually don't know SD_P, we substitute SD_S with "little loss of accuracy". Putting the pieces together, the following is true. Z where

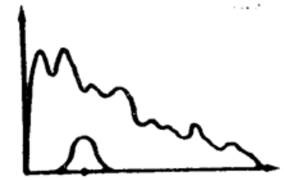
$$Z = \frac{MEAN_S - MEAN_P}{SD_S/SQR(N)}$$

tracks the STANDARD NORMAL distribution.

The above MOTHERLODE expression and expressions like it are cornerstones for the inference portion of the statistical triad. WITHOUT THESE EXPRESSIONS, STATISTICAL PREDICTION would be next to impossible! You must now make these expressions part of your cultural heritage.

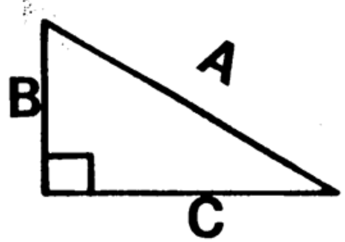
We will be using the above expression in this workshop to build CONFIDENCE INTERVALS and to do STATISTICAL TESTING. By working with this expression, you will begin to appreciate it's power!

TWO MOTHERLODES



PLANE GEOMETRY

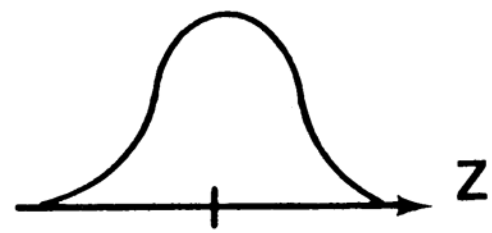
$$A^2 = B^2 + C^2$$



RIGHT TRIANGLE

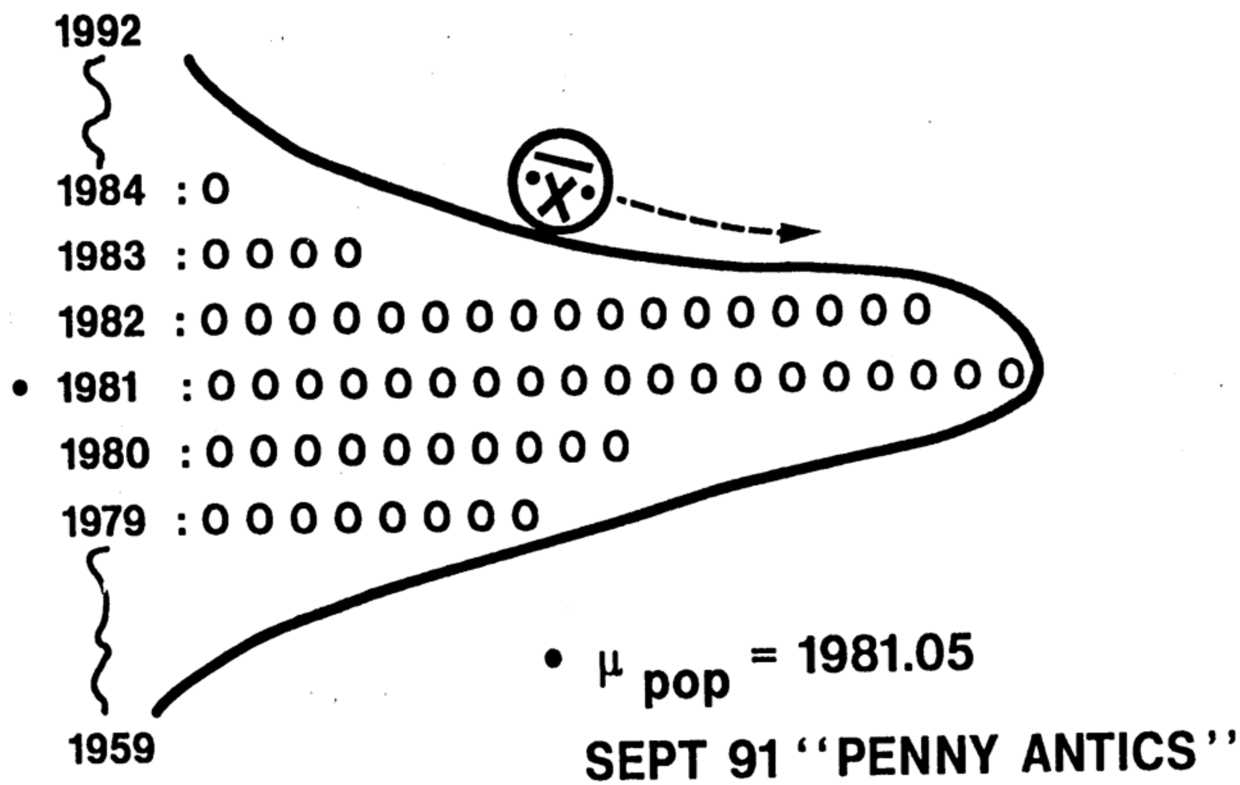
STATISTICAL INFERENCE

$$Z = \frac{\overline{X} - \mu}{\frac{\sigma}{\sqrt{N}}}$$



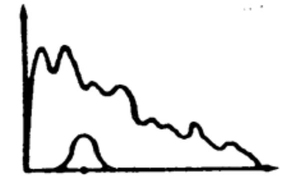
STANDARD NORMAL

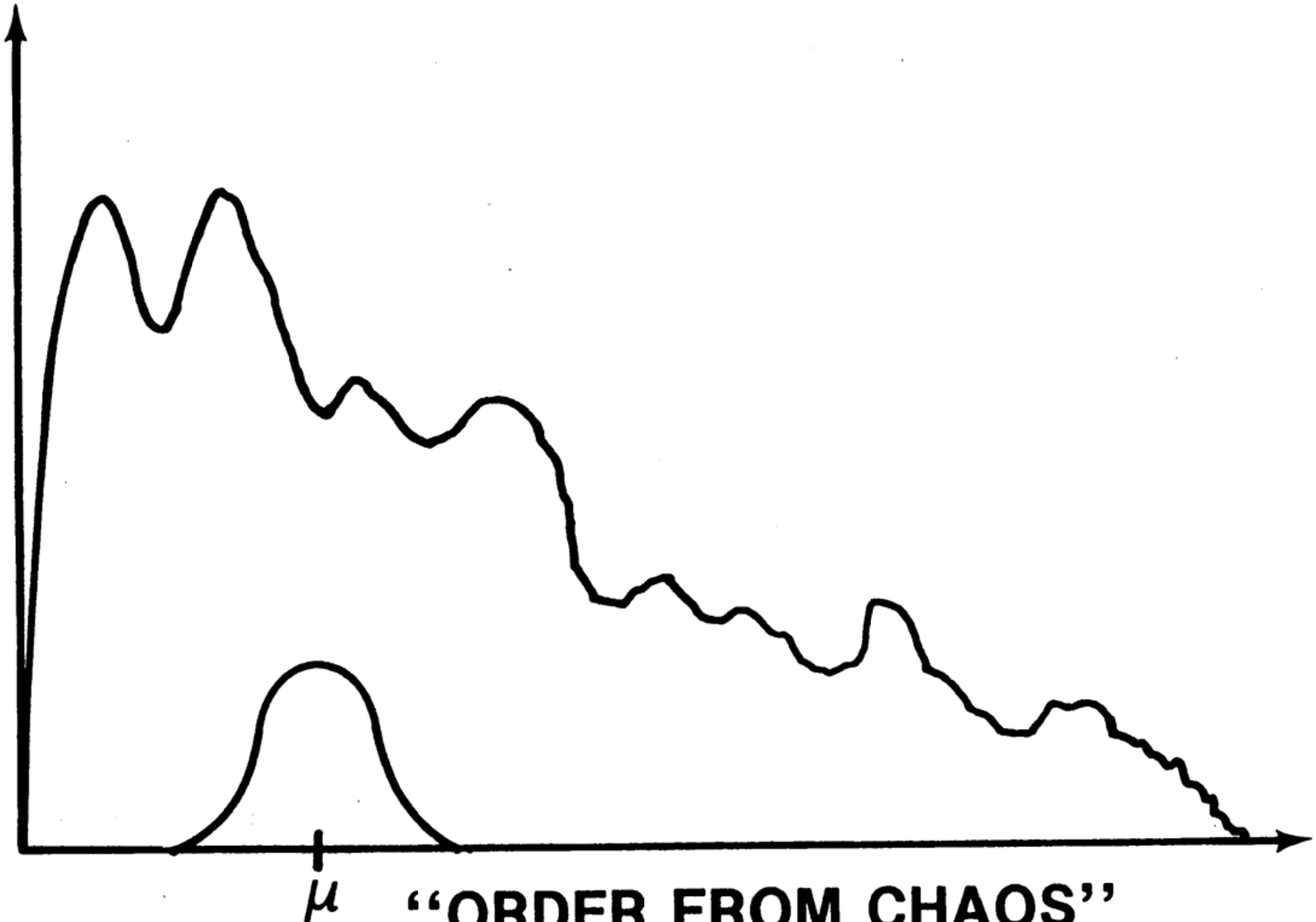
MR MEAN STILL RIDES THE BELL CURVE



GENI 10-1-20

CENTRAL LIMIT THEOREM

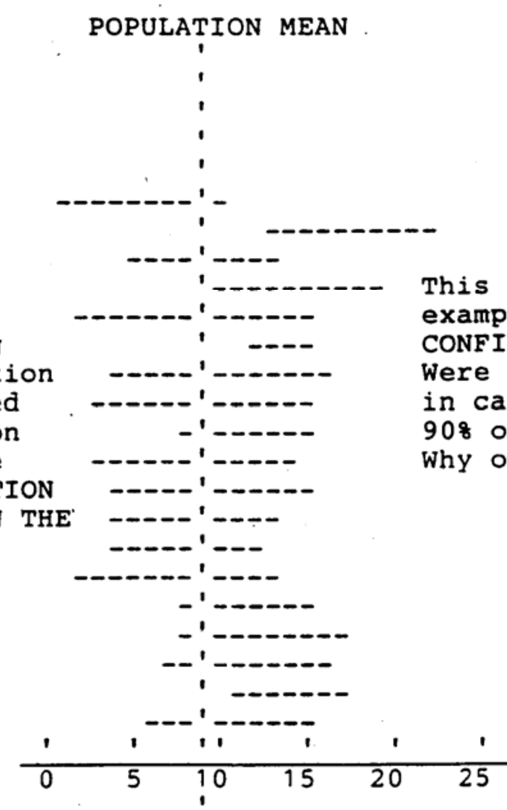




"ORDER FROM CHAOS"

FISHING FOR THE MEAN

Calculate the MEAN and SD for a SAMPLE and note the sample size N. We are going to use our MOTHERLODE expression to construct a "mathematical fish net" known as a CONFIDENCE INTERVAL. This fish net is used to trap the population mean. The CONFIDENCE that we have in a confidence interval is the confidence that we have in the PROCEDURE OF CONSTRUCTING THE NET AND USING THE NET TO CATCH THE POPULATION MEAN. Different samples will produce different nets since nets are constructed using the sample MEAN, SD and size N. The details of this construction process will be given during the workshop. When we talk about a 95% CONFIDENCE LEVEL, we are saying that the procedure used to make nets and catch means will lead to successful fishing about 95% of the time for a given sample size. Similiar statements can be made for 80%, 90% or 99% confidence levels. The stylized diagram below illustrates the process of catching a population mean using a confidence interval. 20 different nets were constructed using 20 different samples. All the samples had identical size. How many times did we catch the population mean?



The CONFIDENCE INTERVAL is an actual line having a length and position which is determined by the construction procedure. We have caught the POPULATION MEAN if it sits ON THE LINE.

This particular example is for a 90% CONFIDENCE LEVEL. Were we successful in catching fish 90% of the time? Why or why not!

CONSTRUCTING A NET

α PERCENT OF

Z'S IN RED

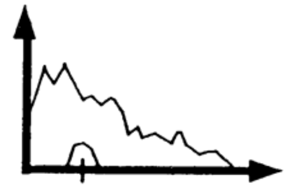
$-Z_{\alpha/2}$ $\quad Z_{\alpha/2}$

- $1-\alpha$ PERCENT OF $\left|\dfrac{\mu-\bar{X}}{\frac{S}{\sqrt{N}}}\right|$ IN PURPLE

- $\left|\dfrac{\mu-\bar{X}}{\frac{S}{\sqrt{N}}}\right| < Z_{\alpha/2}$ IS A $1-\alpha$ PERCENT HAPPENING !

- $\bar{X} - Z_{\alpha/2} \cdot \dfrac{S}{\sqrt{N}} < \mu < \bar{X} + Z_{\alpha/2} \cdot \dfrac{S}{\sqrt{N}}$ DITTO !

THE GAME OF INFERENCE FOR MEANS



SMALL SAMPLE

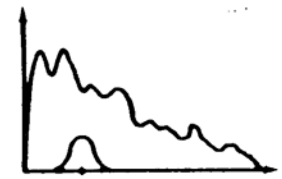
- $N < 30$
- PARENT POPULATION MUST BE BELL SHAPED
- INTUITIVE CLT MAKES THE GAME HAPPEN
- SCORING BY T

LARGE SAMPLE

- $N \geq 30$
- NO RESTRICTIONS ON PARENT POPULATION
- FORMAL CLT MAKES THE GAME HAPPEN
- SCORING BY Z

GENI 10-1-19

STATISTICAL TESTING



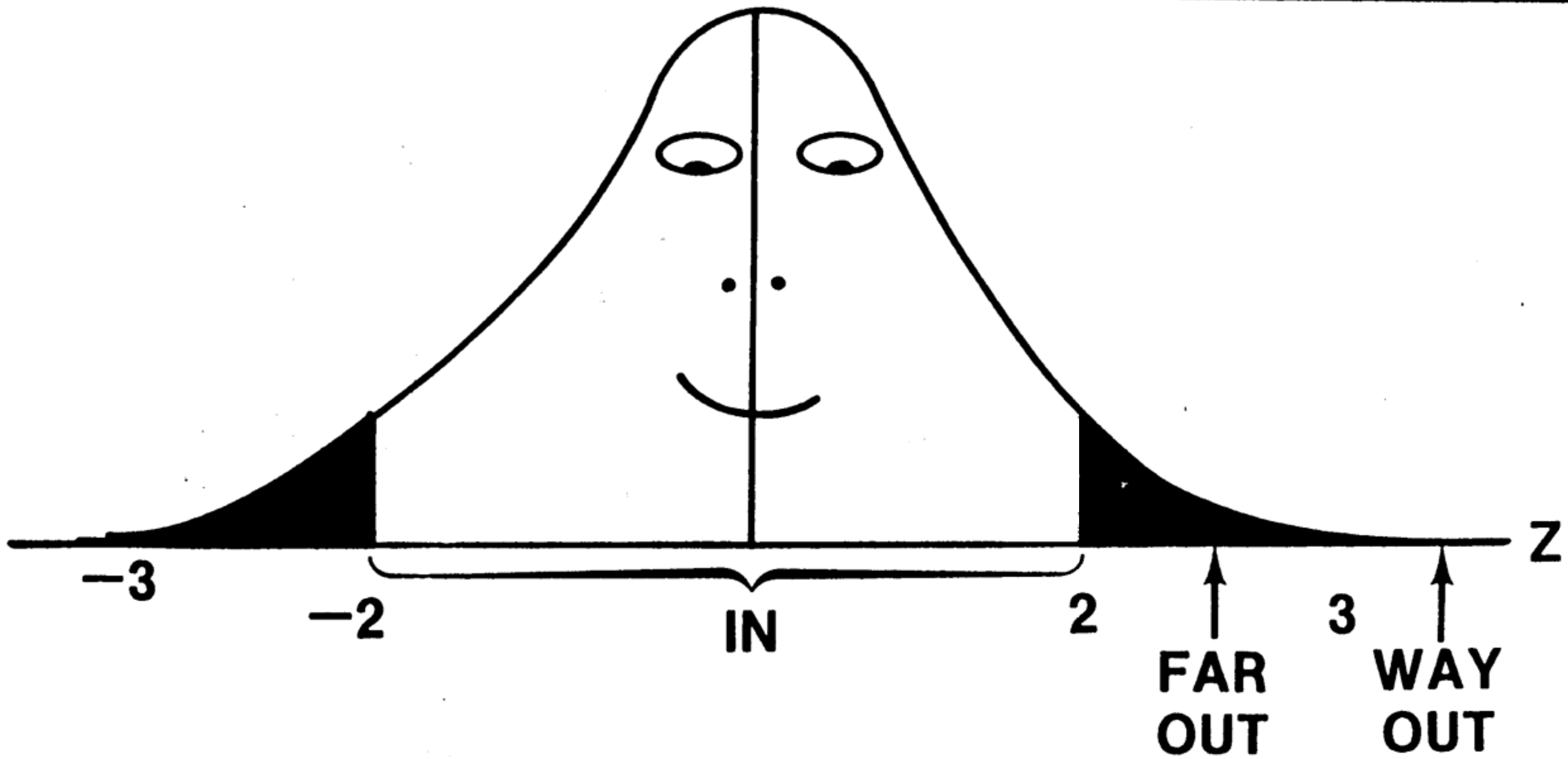
BASIS

- ASSUME THE STATUS QUO.

- THE POPULATION IS ASSUMED LEGALLY INNOCENT UNTIL PROVEN GUILTY.

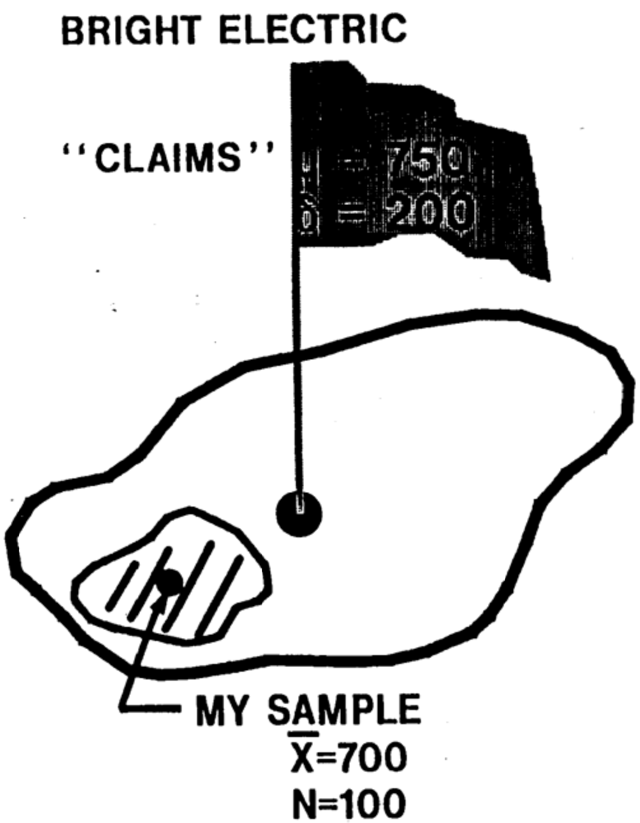
- GUILT IS ESTABLISHED USING SAMPLES.

STATISTICAL TESTING



YOU ARE EITHER IN OR OUT ! THE Z-SCORE TELLS THE STORY !

LIGHT BULBS
LIFE IN HOURS



THE Z SCORE SAYS

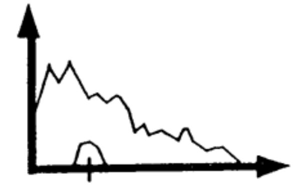
$$Z = \frac{700 - 750}{\frac{200}{\sqrt{100}}}$$

$$= -2.5$$

WHAT IS GOING ON HERE !

GENI 10-1-80

THE GAME OF INFERENCE FOR MEANS



SMALL SAMPLE	LARGE SAMPLE
• $N < 30$	• $N \geq 30$
• PARENT POPULATION MUST BE BELL SHAPED	• NO RESTRICTIONS ON PARENT POPULATION
• INTUITIVE CLT MAKES THE GAME HAPPEN	• FORMAL CLT MAKES THE GAME HAPPEN
• SCORING BY T	• SCORING BY Z

GENI 10-1-19

A TOUGH LESSON ON INFERENCE

This demonstration is designed to introduce you to several aspects of statistical testing. Statistical testing allows us to make statements about populations based on information found in samples. Statistical testing procedures are not perfect and always come with a measure of error. Wise statisticians try to control or bound the error in order to make the ICE SAFE! We are going to experience the "Life and Times of Pennyville" as the resident population undergoes several changes. Samples taken from our population may or may not reflect these changes which is part of the peril associated with statistical testing. Worse yet, a sample could cry "guilty" when our population is really innocent!

In this demonstration, we are going to check the health of our population mean utilizing sample means. The Central Limit Theorem and Z-SCORES will be the tools of the trade. Detailed use will be explained right before the demonstration.

THE FUNDAMENTAL ASSUMPTION

MADE

IS THAT

OUR POPULATION IS INNOCENT UNTIL PROVEN QUILTY!

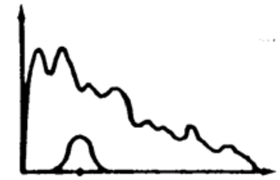
Please note that our population is deemed innocent in the LEGAL SENSE which means some criminals go free! The following table summarizes this discussion using the familiar OK, NOT OK terminology.

		SAMPLE SAYS OK	SAMPLE SAYS NOT OK
POPULATION IS	OK	GOOD TEST SLEEP WELL TONIGHT!	TYPE I ERROR BUM RAP!
POPULATION IS	NOT OK	TYPE II ERROR HANKY PANKY!	GOOD TEST JUSTICE IS SERVED!

NOTES

1) Where is the THIN ICE in the above table?
2) Try to identify which table region applies to each portion of the "Life and Times of Pennyville" demonstration as it is occurring!

LIFE AND TIMES OF PENNYVILLE



SAMPLE SAYS

POPULATION IS	OK	NOT OK
OK	GOOD TEST ! ● SLEEP WELL TONIGHT	TYPE I ERROR ● BUM RAP – SORRY
NOT OK	TYPE II ERROR ● UNDETECTED HANKY PANKY	GOOD TEST ! ● PROSECUTE THE VILLIAN

SCORECARD

The table below allows you to track the Z-SCORES as we move through the "Life and Times of Pennyville" demonstration. The six stages of this demonstration are as follows:

1) PEACEFULL BEGINNINGS,
2) TROUBLE IS BREWING,
3) JUSTICE IS SERVED,
4) RESTORATION,
5) A BUM RAP AND LAWSUIT,
6) PEACE RETURNS.

Z-SCORES are going to be calculated and displayed by designated workshop participants during all six stages of the demonstration. You are invited to use the table below to record the following information as you observe and participate in the process.

STAGE - Record NUMBER
POPULATION - OK or NOT OK
SAMPLE - OK or NOT OK
Z-SCORE - Record VALUE
ERROR - NONE, TYPE I or TYPE II

SCOREKEEPING TABLE

STAGE	POPULATION	SAMPLE	Z-SCORE	ERROR

ASD EXPERIENCES PENNY ANTICS
5-6 AUGUST 91

3200 RESIDENTS IN PENNYVILLE
MEAN = 1982.75

MR MEAN STILL RIDES THE BELL CURVE

```
1985: 0
1984: 00000 00
1983: 00000 00000 0
1982: 00000 000
1981: 00000 00000 00
1980: 000
1979: 00000 0
1978: 0
```

A "SLICE OF 89"

```
 0: 8,9,9
*1: 0,3,3,4,6,6,6,7
 2: 2,4,7
 3: 7
 4: 1
```

GO GO EIGHTIES

```
 08-09: 0
 10-11: 00000 0
 12-13: 00000 00
 14-15: 00000 000
*16-17: 00000 00000 00000 0
 18-19: 00000 00000 00
 20-21: 000
```

"FISHIN" IN PENNYVILLE

CONFIDENCE	90%	95%
ACTUALS	41/50	46/50

MALE HEIGHT

```
 6:0,1
 6:
 6:5
 6:6,6,6,7,7,7
*6:8,8,9,9,9,9,9,9
 7:0,0,0,1,1,1,1
 7:2,2,2,2,2,3,3
 7:5,5
 7:6,6,6
```

COMBINED JENGA

```
 1:999
 2:01
 2:5666666777889
 3:000122233333444
*3:56777788889
 4:000000023334444
 4:5566689
 5:234
```

* What determines the "long pole" in each of these situations?

WL EXPERIENCES PENNY ANTICS
26-27 SEPTEMBER 1991

4400 RESIDENTS IN PENNYVILLE
MEAN = 1981.05

MR MEAN STILL RIDES THE BELL CURVE

```
 1985:
 1984: 0
 1983: 00000
 1982: 00000 00000 00000 00
*1981: 00000 00000 00000 0000
 1980: 00000 00000
 1979: 00000 000
 1978:
```

MILEAGE DATA

```
 0: 1,3,3,5,5,5,6,6,8
*1: 0,0,0,1,2,3,3,3,4,5,6,7
 2: 3
 3: 0,6
 4: 0
 5:
 6: 5
 7:
```

GO GO EIGHTIES

```
 08: 0
 10: 00
 12: 0000000
 14: 0000000000
 16: 00000000000
*18: 00000000000000000
 20: 000000000
 22: 0000
```

"FISHIN" IN PENNYVILLE

CONFIDENCE	96%	99%
ACTUALS	57/60	58/60

"FISHIN" IN THE USA : 1981.14< MEAN <1983.35

MALE

```
 6:
 6:
 6:55
 6:677
 6:88
*7:00000011111
 7:233
 7:555
 7:7
```

FEMALE

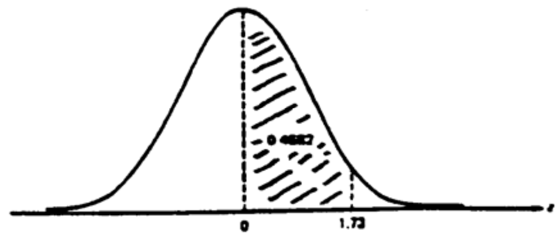
```
 5:9
 6:1
*6:222222333333
 6:44445
 6:6677
 6:9
 7:1
 7:
 7:
```

JENGA

```
 1:
 2:23
 2:58888899
 3:011224
*3:56888899
 4:133333
 4:5556
 5:0144
 5:
```

* What determines the "long pole" in each of these situations?

TABLE
The standard normal distribution



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4692	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990

FOUR GOOD REFERENCES

1) Brightnan, H.J. Statistics In Plain English. Cincinnati: South-Western Publishing Company, 1985.

2) Khazanie, R. Elementary Statistics In a World of Applications. Glenview, Illinois: Scott, Foresman and Company, 1979.

3) Miller, I.R., Freund, J.E.; and Johnson, R. Probability and Statistics for Engineers. 4th ed. Englewood Cliffs, New Jersey: Prentice Hall, 1990.

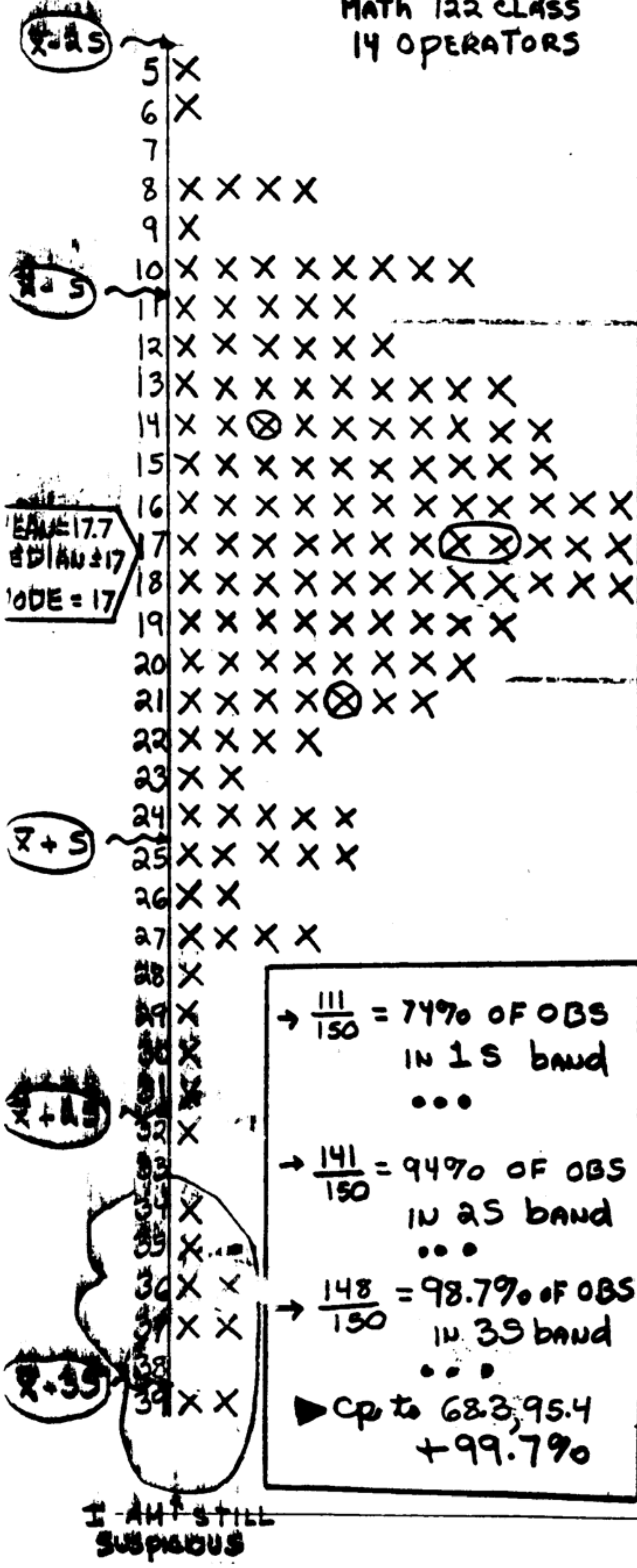
4) Mendenhall, W., Scheaffer, R.L.; and Wackerly, D.D. Mathematical Statistics with Applications. 3rd ed. Boston, Massachusetts: Duxbury Press, 1985.

NOTES

References are given in increasing order of difficulty. 1 is CAKE level, 2 is BASIC level, 3 is ENGINEERING level and 4 is MATHEMATICAL level. 1,3 and 4 contain material on the Design of Experiments (DOX).

K-Mart Paper Clip Data : 25 Feb 92

MATH 122 CLASS
14 OPERATORS



$\bar{X}$	SS
1×5 = 5	161.29
1×6 = 6	138.89
0×7 = 0	0.0
4×8 = 32	376.36
1×9 = 9	75.69
8×10 = 80	474.32
5×11 = 55	224.45
6×12 = 72	194.94
9×13 = 117	198.81
10×14 = 140	136.90
10×15 = 150	72.90
12×16 = 192	34.68
12×17 = 204	5.88
12×18 = 216	1.08
9×19 = 171	15.21
8×20 = 160	42.32
7×21 = 147	76.23
4×22 = 88	73.96
2×23 = 46	56.18
5×24 = 120	198.45
5×25 = 125	266.45
2×26 = 52	137.78
4×27 = 108	345.96
1×28 = 28	106.09
1×29 = 29	127.69
1×30 = 30	151.29
1×31 = 31	176.89
1×32 = 32	204.49
0×33 = 0	0.0
1×34 = 34	265.69
1×35 = 35	299.29
2×36 = 72	669.78
2×37 = 74	744.98
0×38 = 0	0.0
2×39 = 78	907.38
150 } 2666	6962.30
$\bar{X}$ = 17.77	S^2 = 46.415
	S = 6.813