DOCUMENT RESUME

ED 450 156		TM 032 346

AUTHOR	Rollins, Dahl A.
TITLE	There Are Infinitely Many Univariate Normal Distributions: Distribution "Bells" Come in Many Appearances.
PUB DATE	2001-02-01
NOTE	22p.; Paper presented at the Annual Meeting of the Southwest Educational Research Association (New Orleans, LA, February 1-3, 2001).
PUB TYPE	Numerical/Quantitative Data (110) -- Reports - Descriptive (141) -- Speeches/Meeting Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Heuristics; *Scores; *Statistical Distributions
IDENTIFIERS	*Normal Distribution; Normality Tests; *Univariate Analysis

ABSTRACT

Many univariate statistical methods assume that dependent variable data have a univariate distribution. Some statistical methods assume that the error scores are normally distributed. It is clear that the concept of data normality is an important one in statistics. This paper explains that, notwithstanding common misconceptions to the contrary, there are infinitely many normal distributions, each differing in appearance. The classic bell-shaped curve only depicts one single case of the infinitely many univariate normal distributions. The bell-shaped curve is the general shape of all normal univariate distributions, but the bells can have an infinite assortment of shapes and still be bells. A small heuristic data set involving scores of 100 people of a variable labeled "x" is used to illustrate this principle. (SLD)

dahl.wp 1/22/01

Running head: NORMAL DISTRIBUTIONS

There Are Infinitely Many Univariate Normal Distributions: Distribution "Bells" Come in Many Appearances

Dahl A. Rollins

Texas A&M University 77843-4225

Paper presented at the annual meeting of the Southwest Educational Research Association, New Orleans, February 1, 2001.

Abstract

Many univariate statistical methods assume that dependent variable data have a univariate normal distribution. And some statistical methods assume that the error scores are normally distributed. Clearly, the concept of data normality is an important one in statistics. The present paper explains, notwithstanding common misconceptions to the contrary, that there are infinitely many normal distributions, each differing in appearance.

Many univariate statistical methods assume that dependent variable data have a univariate normal distribution (cf. Hinkle, Weirsma & Jurs, 1998). And some statistical methods assume that the error scores are normally distributed (cf. Thompson, 1992). Clearly, the concept of data normality is an important one in statistics.

The normal distribution is a model reflecting a particular mathematical equation (Hinkle, Wiersma & Jurs, 1998). The normal distribution is not a single distribution, but for every potential mean and standard deviation there is a different unique normal distribution (Hinkle, Wiersma & Jurs, 1998).

However, because all normal distributions are symmetrical, the mean always equals the median of a normal distribution (Vogt, 1993). Furthermore, because all normal distributions are also unimodal, the mode also equals both the median and the mean of a given normal distribution.

Abraham Demoivre, a French mathematician, developed the normal distribution general equation in the eighteenth century (Hinkle, Wiersma & Jurs, 1998). The normal distribution is also known as the "Gaussian" distribution, after the European mathematician who devoted so much of his life to studying this distribution.

The depiction of the normal distribution in textbooks almost always shows the very famous bell-shaped curve. Many assume this classic bell-shaped curve is the only shape for the normal distribution. However, that classic bell-shaped curve only depicts one single case of the *infinitely many univariate normal* *distributions.*

The frequently-depicted normal distribution actually is a special case of the infinitely many normal distributions. The depicted "bell-shaped curve" is invariably a picture specifically of the "standard normal distribution" (i.e., z scores that are normally distributed) (Hinkle, Wiersma & Jurs, 1998).

To determine if a curve is normal the raw scores must be converted to z-scores to standardize them. It is impossible to determine whether or not even a symmetrical and unimodal distribution is normal simply by looking at it (Bump, 1991). The bell-shape is the general shape of all normal univariate distributions, however, bells can have an infinite assortment of shapes and still all be bells (Burdenski, 2000).

In statistics there are four important statistical estimates called "moments about the mean." These involve: (a) the mean itself, (b) the standard deviation, (c) the coefficient of skewness, and (d) the coefficient of kurtosis.

Skewness indicates whether a distribution or curve is symmetrical (i.e., mean = median) or asymmetrical (i.e., skewed: mean ≠ median). Kurtosis refers to how tall or flat the curve is in relation to how wide or narrow the distribution is.

In symmetrical distributions, if the two halves of the curve were divided at the median, they would mirror each other. Or distributions can be skewed positively or negatively. To be positively skewed or skewed right a few scores (called the "tail") are to the right of the majority or the body of the scores. To be negatively skewed or skewed left a few scores (called the "tail") are to the left of the majority or the body of the scores.

Kurtosis, the last moment, is "the ratio of the average of the deviation scores raised to the fourth power to the standard deviation also raised to the fourth power" (Bump, 1991). The normal curve has a kurtosis value of 3.0 using this formula, which is why now the common practice is to subtract 3.0 from the kurtosis value so as to yield a value of zero for kurtosis of a normal curve.

To yield a curve with a positive coefficient of kurtosis, there has to be a higher concentration of scores around the mean, resulting in a narrow distribution. In a negative kurtosis curve, the distribution is broad and there is a low concentration of scores around the mean (Bump, 1991).

In order for normality to exist, there must be suitable distribution of height relative to width. A tall, thin distribution is called "leptokurtic." A wide, flat distribution is called "platykurtic." As Henson (1999) noted, "Symmetry is a necessary, but insufficient condition for normality."

The purpose of the present paper is to illustrate in concrete fashion that there are infinitely many normal distributions. This will be demonstrated with heuristic data.

Heuristic Examples

Table 1 presents a small heuristic data set involving scores of 100 people of a variable labelled "x". These data are very close to normally distributed, as can be seen from the descriptive statistics presented in Table 2.

INSERT TABLES 1 AND 2 ABOUT HERE.

Four Different Normal Distributions Each with Means of 0

Using the SPSS syntax presented in the Appendix, 4 variables were computed (i.e., "mean0_1," "mean0_2," "mean0_3," "mean0_4"), each with a mean equal to zero. However, as reported in Table 2, the standard deviations of the variables differ.

Even though the 4 distributions differ in their "spreadoutness," all these variables are approximately normally distributed, as reflected in the coefficients of skewness and kurtosis reported in Table 2 all being close to zero. This can also be seen in Figures 1 through 4, which present graphs of the data distributions with normal curves imposed on the actual data distributions.

INSERT FIGURES 1 THROUGH 4 ABOUT HERE.

Three Different Normal Distributions, None with Means of 0

Using the SPSS syntax presented in the Appendix, 3 additional variables were computed (i.e., "mean_1," "mean_2," and "mean_3"), each with a mean not equal to zero. As reported in Table 2, the means and the standard deviations of the variables differ.

Again, even though the 3 distributions differ in their "spreadoutness," all these variables are approximately normally distributed, as reflected in the coefficients of skewness and kurtosis reported in Table 2 all being close to zero. This can also be seen in Figures 5 through 7, which present graphs of the data distributions with normal curves imposed on the actual data distributions.

INSERT FIGURES 5 THROUGH 7 ABOUT HERE.

Summary

Although all the graphs look different and the second set of variables all had different means and standard deviations, the skewness and kurtosis values were exactly the same for all 7 variables. All these variables were representative of normal curves. In summary, it is possible that univariate normal curve distributions may look very different.

It is impossible to judge whether even symmetrical, unimodal data are normally distributed, unless coefficients of skewness and kurtosis are computed. Clearly, while the "standard normal distribution" has a classical bell shape, bells come in many widths and tallnesses, and so do normal distributions.

References

Bump, W. (1991, January). The normal curve takes many forms: A review of skewness and kurtosis. Paper presented at the annual meeting of the Southwest Educational Research Association, San Antonio. (ERIC Document Reproduction Service No. ED 342 790)

Burdenski, Jr., T.K. (2000, April). Evaluating univariate, bivariate and multivariate normality using graphical procedures. Paper presented at the annual meeting of the American Educational Research Association, New Orleans.

Henson, R.K. (1999). Multivariate normality: What is it and how is it assessed?. In B. Thompson (Ed.), Advances in social science methodology (Vol. 5, pp. 193-212). Stamford, CT: JAI Press.

Hinkle, D.E., Wiersma, W., & Jurs, S.G. (1998). Applied statistics for the behavioral sciences (4th ed.). Boston: Houghton Mifflin.

Thompson, B. (1992, April). Interpreting regression results: beta weights and structure coefficients are both important. Paper presented at the annual meeting of the American Educational Research Association, San Francisco. (ERIC Document Reproduction Service No. ED 344 897)

Vogt, P.W. (1993). Dictionary of statistics and methodology. Newbury Park, CA: Sage Publications.

Table 1
Heuristic Data (n = 100)
from Thompson (2001)

ID	x
1	2.4
2	2.8
3	3.0
4	3.2
5	3.3
6	3.4
7	3.5
8	3.6
9	3.6
10	3.7
11	3.7
12	3.8
13	3.8
14	3.9
15	3.9
16	4.0
17	4.0
18	4.1
19	4.1
20	4.1
21	4.2
22	4.2
23	4.2
24	4.3
25	4.3
26	4.3
27	4.4
28	4.4
29	4.4
30	4.5
31	4.5
32	4.5
33	4.6
34	4.6
35	4.6
36	4.6
37	4.7
38	4.7
39	4.7
40	4.7
41	4.8
42	4.8
43	4.8
44	4.8
45	4.9
46	4.9
47	4.9
48	4.9
49	5.0
50	5.0
51	5.0
52	5.0
53	5.1
54	5.1
55	5.1
56	5.1

57	5.2
58	5.2
59	5.2
60	5.2
61	5.3
62	5.3
63	5.3
64	5.3
65	5.4
66	5.4
67	5.4
68	5.4
69	5.5
70	5.5
71	5.5
72	5.6
73	5.6
74	5.6
75	5.7
76	5.7
77	5.7
78	5.8
79	5.8
80	5.8
81	5.9
82	5.9
83	5.9
84	6.0
85	6.0
86	6.1
87	6.1
88	6.2
89	6.2
90	6.3
91	6.3
92	6.4
93	6.4
94	6.5
95	6.6
96	6.7
97	6.8
98	7.0
99	7.2
100	7.6

Table 2
Descriptive Statistics

Variable	Mean	SD	Coef. of Skewness	Coef. of Kurtosis
	Statistics			
x	5.00	1.005	.000	-.090
mean0_1	0.00	1.005	.000	-.090[a]
mean0_2	0.00	**0.502**	.000	-.090[b]
mean0_3	0.00	**2.010**	.000	-.090[c]
mean0_4	0.00	**5.025**	.000	-.090[d]
mean_1	**2.50**	**0.502**	.000	-.090[e]
mean_2	**10.00**	**2.010**	.000	-.090[f]
mean_3	**25.00**	**5.025**	.000	-.090[g]

[a] "mean0_1" = "x" - 5.
[b] = "mean0_1" * .5
[c] = "mean0_1" * 2.
[d] = "mean0_1" * 5.

[e] = "x" * .5
[f] = "x" * 2.
[g] = "x" * 5.

Table 3
Scores (n = 100) for Three Variables
Each with **Means = .0**

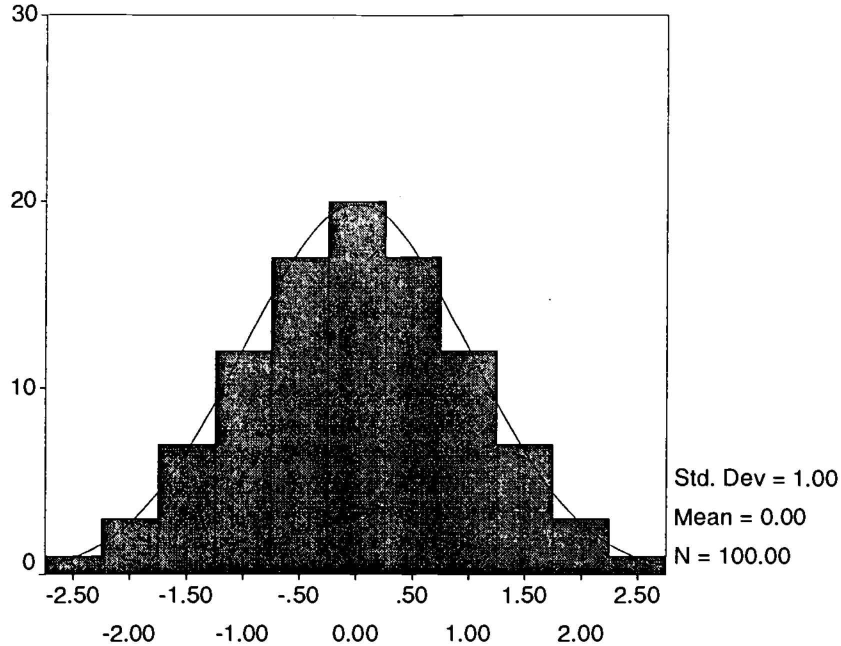
ID	mean0_1	mean0_2	mean0_3	mean0_4
1	-2.60	-1.30	-5.20	-13.00
2	-2.20	-1.10	-4.40	-11.00
3	-2.00	-1.00	-4.00	-10.00
4	-1.80	-.90	-3.60	-9.00
5	-1.70	-.85	-3.40	-8.50
6	-1.60	-.80	-3.20	-8.00
7	-1.50	-.75	-3.00	-7.50
8	-1.40	-.70	-2.80	-7.00
9	-1.40	-.70	-2.80	-7.00
10	-1.30	-.65	-2.60	-6.50
11	-1.30	-.65	-2.60	-6.50
12	-1.20	-.60	-2.40	-6.00
13	-1.20	-.60	-2.40	-6.00
14	-1.10	-.55	-2.20	-5.50
15	-1.10	-.55	-2.20	-5.50
16	-1.00	-.50	-2.00	-5.00
17	-1.00	-.50	-2.00	-5.00
18	-.90	-.45	-1.80	-4.50
19	-.90	-.45	-1.80	-4.50
20	-.90	-.45	-1.80	-4.50
21	-.80	-.40	-1.60	-4.00
22	-.80	-.40	-1.60	-4.00
23	-.80	-.40	-1.60	-4.00
24	-.70	-.35	-1.40	-3.50
25	-.70	-.35	-1.40	-3.50
26	-.70	-.35	-1.40	-3.50
27	-.60	-.30	-1.20	-3.00
28	-.60	-.30	-1.20	-3.00
29	-.60	-.30	-1.20	-3.00
30	-.50	-.25	-1.00	-2.50
31	-.50	-.25	-1.00	-2.50
32	-.50	-.25	-1.00	-2.50
33	-.40	-.20	-.80	-2.00
34	-.40	-.20	-.80	-2.00
35	-.40	-.20	-.80	-2.00
36	-.40	-.20	-.80	-2.00
37	-.30	-.15	-.60	-1.50
38	-.30	-.15	-.60	-1.50
39	-.30	-.15	-.60	-1.50
40	-.30	-.15	-.60	-1.50
41	-.20	-.10	-.40	-1.00
42	-.20	-.10	-.40	-1.00
43	-.20	-.10	-.40	-1.00
44	-.20	-.10	-.40	-1.00
45	-.10	-.05	-.20	-.50
46	-.10	-.05	-.20	-.50
47	-.10	-.05	-.20	-.50
48	-.10	-.05	-.20	-.50
49	.00	.00	.00	.00
50	.00	.00	.00	.00
51	.00	.00	.00	.00
52	.00	.00	.00	.00
53	.10	.05	.20	.50
54	.10	.05	.20	.50

55	.10	.05	.20	.50
56	.10	.05	.20	.50
57	.20	.10	.40	1.00
58	.20	.10	.40	1.00
59	.20	.10	.40	1.00
60	.20	.10	.40	1.00
61	.30	.15	.60	1.50
62	.30	.15	.60	1.50
63	.30	.15	.60	1.50
64	.30	.15	.60	1.50
65	.40	.20	.80	2.00
66	.40	.20	.80	2.00
67	.40	.20	.80	2.00
68	.40	.20	.80	2.00
69	.50	.25	1.00	2.50
70	.50	.25	1.00	2.50
71	.50	.25	1.00	2.50
72	.60	.30	1.20	3.00
73	.60	.30	1.20	3.00
74	.60	.30	1.20	3.00
75	.70	.35	1.40	3.50
76	.70	.35	1.40	3.50
77	.70	.35	1.40	3.50
78	.80	.40	1.60	4.00
79	.80	.40	1.60	4.00
80	.80	.40	1.60	4.00
81	.90	.45	1.80	4.50
82	.90	.45	1.80	4.50
83	.90	.45	1.80	4.50
84	1.00	.50	2.00	5.00
85	1.00	.50	2.00	5.00
86	1.10	.55	2.20	5.50
87	1.10	.55	2.20	5.50
88	1.20	.60	2.40	6.00
89	1.20	.60	2.40	6.00
90	1.30	.65	2.60	6.50
91	1.30	.65	2.60	6.50
92	1.40	.70	2.80	7.00
93	1.40	.70	2.80	7.00
94	1.50	.75	3.00	7.50
95	1.60	.80	3.20	8.00
96	1.70	.85	3.40	8.50
97	1.80	.90	3.60	9.00
98	2.00	1.00	4.00	10.00
99	2.20	1.10	4.40	11.00
100	2.60	1.30	5.20	13.00

Table 4
Scores ($\underline{n}$ = 100) for Three Variables
Each with **Means ≠ .0**

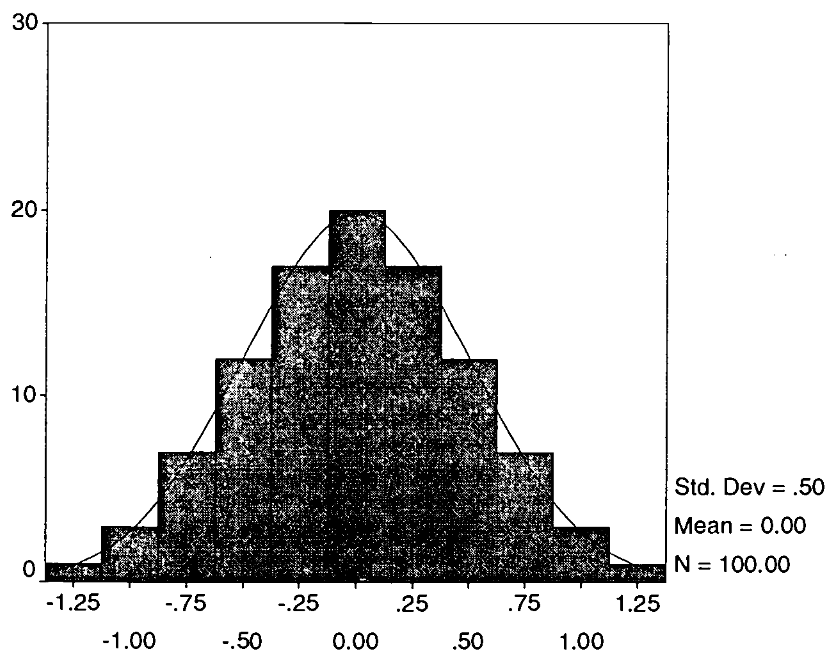
ID	mean_1	mean_2	mean_3
1	1.20	4.80	12.00
2	1.40	5.60	14.00
3	1.50	6.00	15.00
4	1.60	6.40	16.00
5	1.65	6.60	16.50
6	1.70	6.80	17.00
7	1.75	7.00	17.50
8	1.80	7.20	18.00
9	1.80	7.20	18.00
10	1.85	7.40	18.50
11	1.85	7.40	18.50
12	1.90	7.60	19.00
13	1.90	7.60	19.00
14	1.95	7.80	19.50
15	1.95	7.80	19.50
16	2.00	8.00	20.00
17	2.00	8.00	20.00
18	2.05	8.20	20.50
19	2.05	8.20	20.50
20	2.05	8.20	20.50
21	2.10	8.40	21.00
22	2.10	8.40	21.00
23	2.10	8.40	21.00
24	2.15	8.60	21.50
25	2.15	8.60	21.50
26	2.15	8.60	21.50
27	2.20	8.80	22.00
28	2.20	8.80	22.00
29	2.20	8.80	22.00
30	2.25	9.00	22.50
31	2.25	9.00	22.50
32	2.25	9.00	22.50
33	2.30	9.20	23.00
34	2.30	9.20	23.00
35	2.30	9.20	23.00
36	2.30	9.20	23.00
37	2.35	9.40	23.50
38	2.35	9.40	23.50
39	2.35	9.40	23.50
40	2.35	9.40	23.50
41	2.40	9.60	24.00
42	2.40	9.60	24.00
43	2.40	9.60	24.00
44	2.40	9.60	24.00
45	2.45	9.80	24.50
46	2.45	9.80	24.50
47	2.45	9.80	24.50
48	2.45	9.80	24.50
49	2.50	10.00	25.00
50	2.50	10.00	25.00
51	2.50	10.00	25.00
52	2.50	10.00	25.00
53	2.55	10.20	25.50
54	2.55	10.20	25.50

55	2.55	10.20	25.50
56	2.55	10.20	25.50
57	2.60	10.40	26.00
58	2.60	10.40	26.00
59	2.60	10.40	26.00
60	2.60	10.40	26.00
61	2.65	10.60	26.50
62	2.65	10.60	26.50
63	2.65	10.60	26.50
64	2.65	10.60	26.50
65	2.70	10.80	27.00
66	2.70	10.80	27.00
67	2.70	10.80	27.00
68	2.70	10.80	27.00
69	2.75	11.00	27.50
70	2.75	11.00	27.50
71	2.75	11.00	27.50
72	2.80	11.20	28.00
73	2.80	11.20	28.00
74	2.80	11.20	28.00
75	2.85	11.40	28.50
76	2.85	11.40	28.50
77	2.85	11.40	28.50
78	2.90	11.60	29.00
79	2.90	11.60	29.00
80	2.90	11.60	29.00
81	2.95	11.80	29.50
82	2.95	11.80	29.50
83	2.95	11.80	29.50
84	3.00	12.00	30.00
85	3.00	12.00	30.00
86	3.05	12.20	30.50
87	3.05	12.20	30.50
88	3.10	12.40	31.00
89	3.10	12.40	31.00
90	3.15	12.60	31.50
91	3.15	12.60	31.50
92	3.20	12.80	32.00
93	3.20	12.80	32.00
94	3.25	13.00	32.50
95	3.30	13.20	33.00
96	3.35	13.40	33.50
97	3.40	13.60	34.00
98	3.50	14.00	35.00
99	3.60	14.40	36.00
100	3.80	15.20	38.00



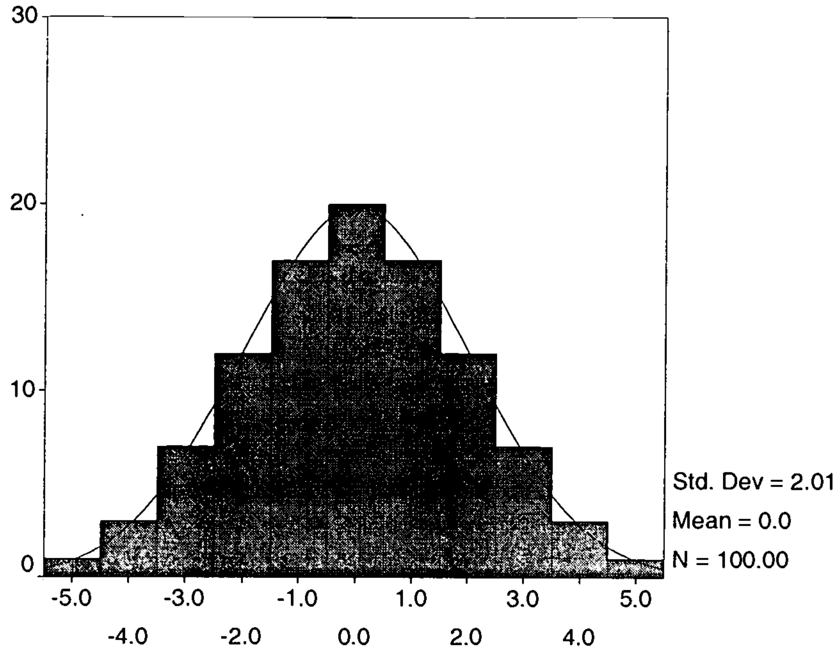
MEAN0_1

Graph

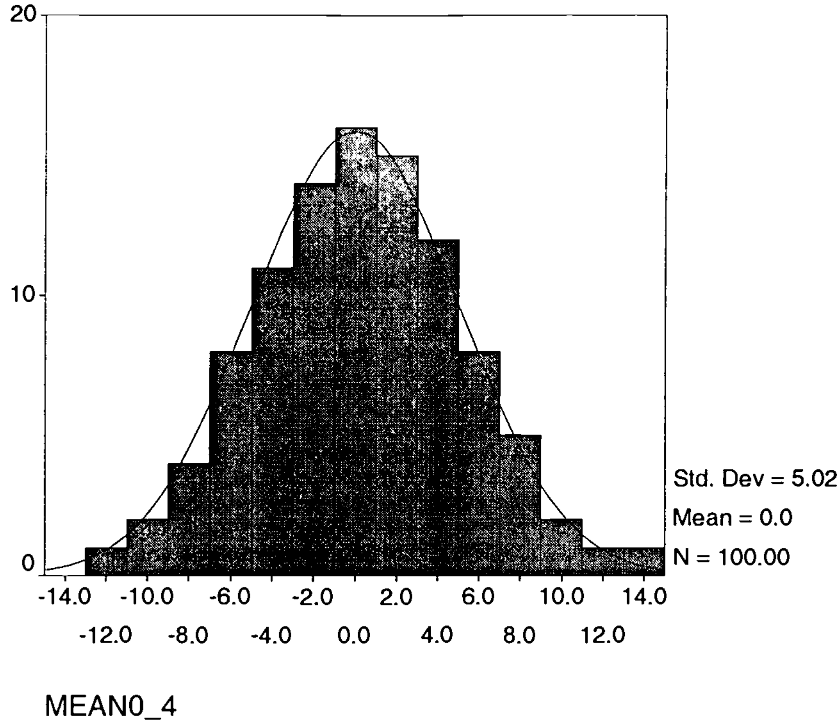


MEAN0_2

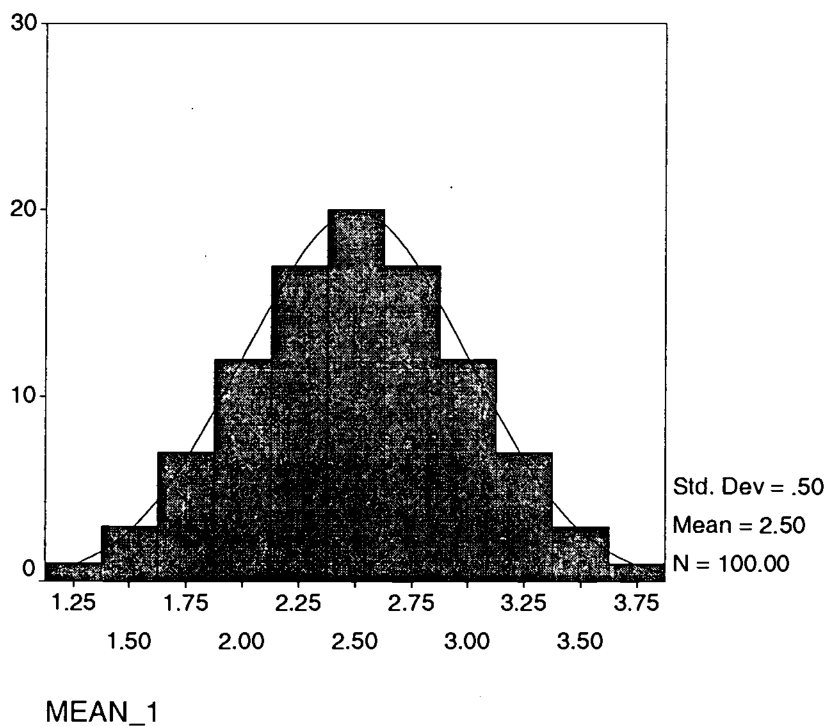
Graph



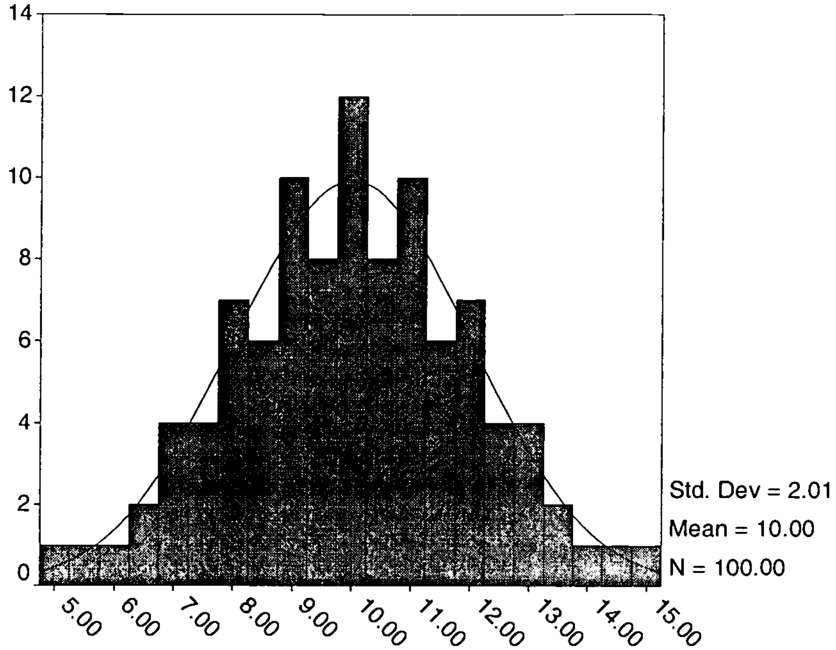
MEAN0_3

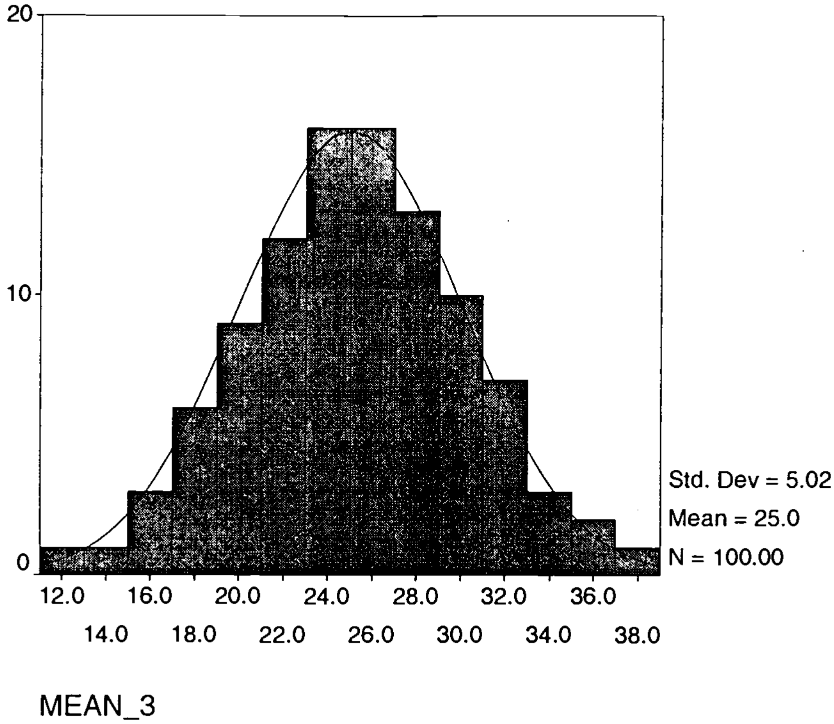

20
10
0
Std. Dev = 5.02
Mean = 0.0
N = 100.00
-14.0
-10.0
-6.0
-2.0
2.0
6.0
10.0
14.0
-12.0
-8.0
-4.0
0.0
4.0
8.0
12.0
MEAN0_4

Graph


30
20
10
0
Std. Dev = .50
Mean = 2.50
N = 100.00
1.25
1.75
2.25
2.75
3.25
3.75
1.50
2.00
2.50
3.00
3.50
MEAN_1

Graph


14
12
10
8
6
4
2
0
Std. Dev = 2.01
Mean = 10.00
N = 100.00
5.00
6.00
7.00
8.00
9.00
10.00
11.00
12.00
13.00
14.00
15.00


20
10
0
Std. Dev = 5.02
Mean = 25.0
N = 100.00
12.0
16.0
20.0
24.0
28.0
32.0
36.0
14.0
18.0
22.0
26.0
30.0
34.0
38.0
MEAN_3

APPENDIX
SPSS Syntax to Implement the Reported Analyses

```
SET BLANKS=SYSMIS UNDEFINED=WARN printback=listing.
TITLE 'Bruce Thompson^s normal data ***************' .
DATA LIST
  FILE='a:normal.dta' FIXED RECORDS=1 TABLE
  /1 id 1-3 x 4-8 .
list variables=all/cases=999 .

subtitle '1  show several normal with same Mean ***' .
execute .
compute mean0_1 = x - 5. .
compute mean0_2 = mean0_1 * .5 .
compute mean0_3 = mean0_1 * 2. .
compute mean0_4 = mean0_1 * 5. .
list variables=id mean0_1 to mean0_4/cases=999 .
descriptives variables=x mean0_1 to mean0_4/
  statistics=all .

subtitle '2  show several normal with *dif* Mean ***' .
execute .
compute mean_1 = x * .5 .
compute mean_2 = x * 2. .
compute mean_3 = x * 5. .
list variables=id mean_1 to mean_3/cases=999 .
descriptives variables=x mean_1 to mean_3/
  statistics=all .
GRAPH  /HISTOGRAM(NORMAL)=mean0_1 .
GRAPH  /HISTOGRAM(NORMAL)=mean0_2 .
GRAPH  /HISTOGRAM(NORMAL)=mean0_3 .
GRAPH  /HISTOGRAM(NORMAL)=mean0_4 .
GRAPH  /HISTOGRAM(NORMAL)=mean_1 .
GRAPH  /HISTOGRAM(NORMAL)=mean_2 .
GRAPH  /HISTOGRAM(NORMAL)=mean_3 .
```

TM032346

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement (OERI)
Educational Resources Information Center (ERIC)

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: THERE ARE INFINITELY MANY UNIVARIATE NORMAL DISTRIBUTIONS: DISTRIBUTION "BELLS" COME IN MANY APPEARANCES	
Author(s): DAHL A. ROLLINS	
Corporate Source:	Publication Date: 2/1/01

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic/optical media, and sold through the ERIC Document Reproduction Service (EDRS) or other ERIC vendors. Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce the identified document, please CHECK ONE of the following options and sign the release below.

[X] Sample sticker to be affixed to document

Check here
Permitting microfiche (4"x 6" film), paper copy, electronic, and optical media reproduction

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
DAHL A. ROLLINS
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Level 1

Sample sticker to be affixed to document [ ]

or here
Permitting reproduction in other than paper copy.

"PERMISSION TO REPRODUCE THIS MATERIAL IN OTHER THAN PAPER COPY HAS BEEN GRANTED BY
Sample
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Level 2

Sign Here, Please

Documents will be processed as indicated provided reproduction quality permits. If permission to reproduce is granted, but neither box is checked, documents will be processed at Level 1.

"I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce this document as indicated above. Reproduction from the ERIC microfiche or electronic/optical media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries."

Signature: X Dahl Rollins	Position: RES ASSOCIATE
Printed Name: DAHL A. ROLLINS	Organization: TEXAS A&M UNIVERSITY
Address: TAMU DEPT EDUC PSYC COLLEGE STATION, TX 77843-4225	Telephone Number: 979/845-1335
	Date: 1/22/01