H. G. Wells

"Statistical thinking will one day be as necessary for efficient citizenship as the ability to read and write."

Develop

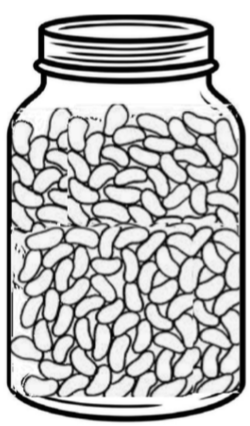
mth_150

Next

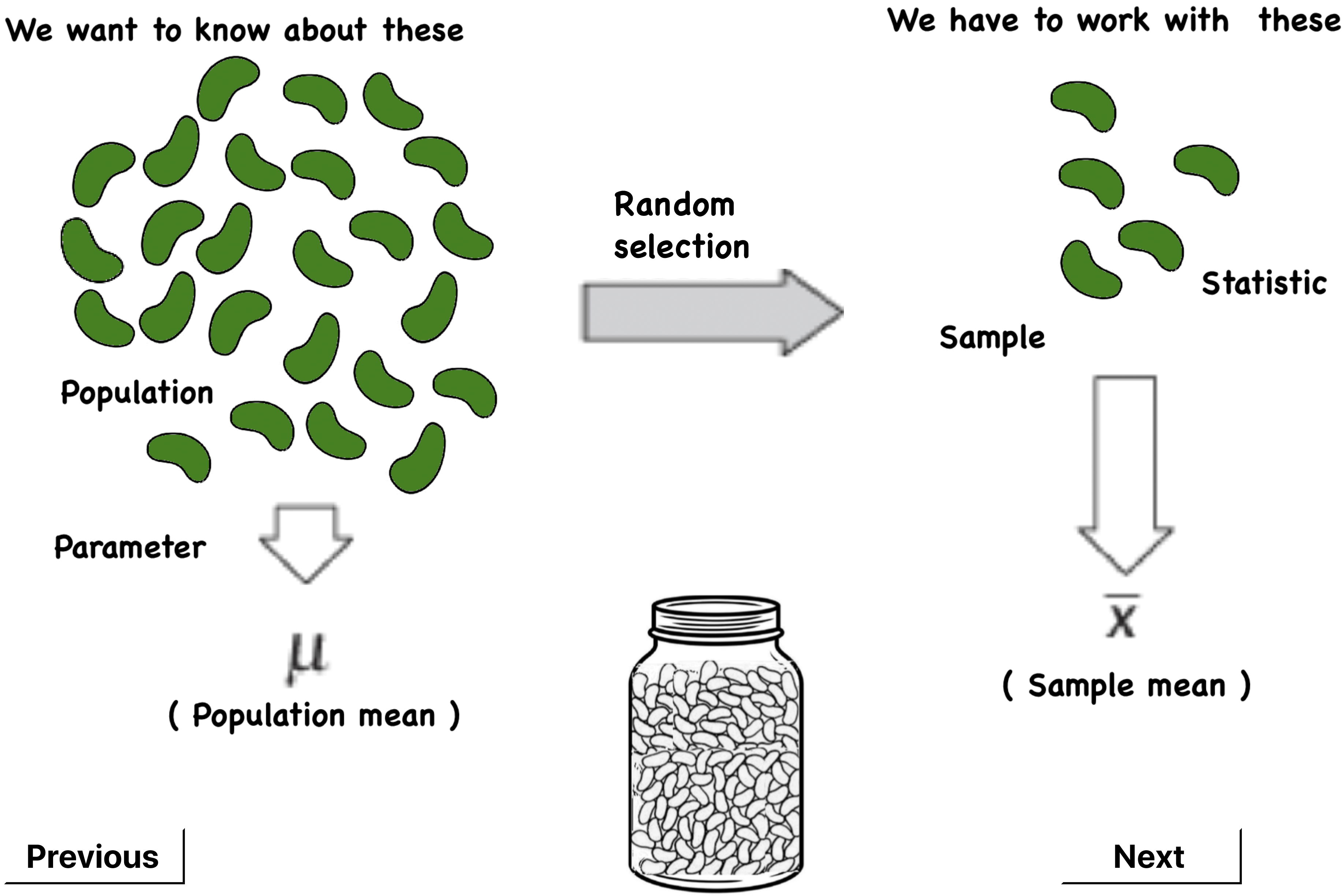
Example:
A virus has attacked the bean population and is expected to survive a mean duration of time equal to 38.3 months with a standard deviation of 6.5 months. Investigators are hopeful that a new therapy will affect survival.

Suppose that the new treatment is administered to 100 cultures. It is observed that the average survival time is 37 months.

Is survival statistically significantly changed (5%) with receipt of the new treatment ? **or** Can you be 95% confident that the new therapy does change the mean survival time?

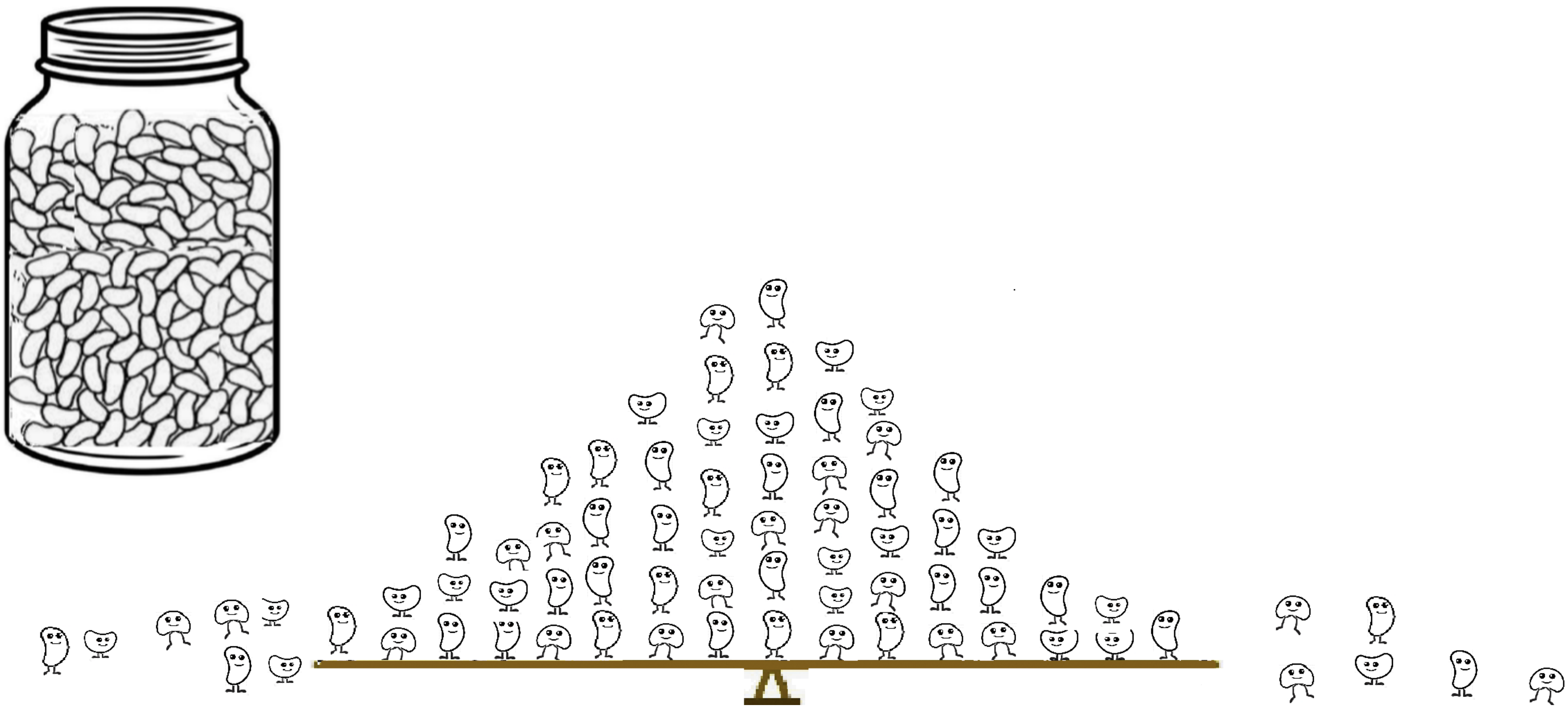


Inference



Previous Next

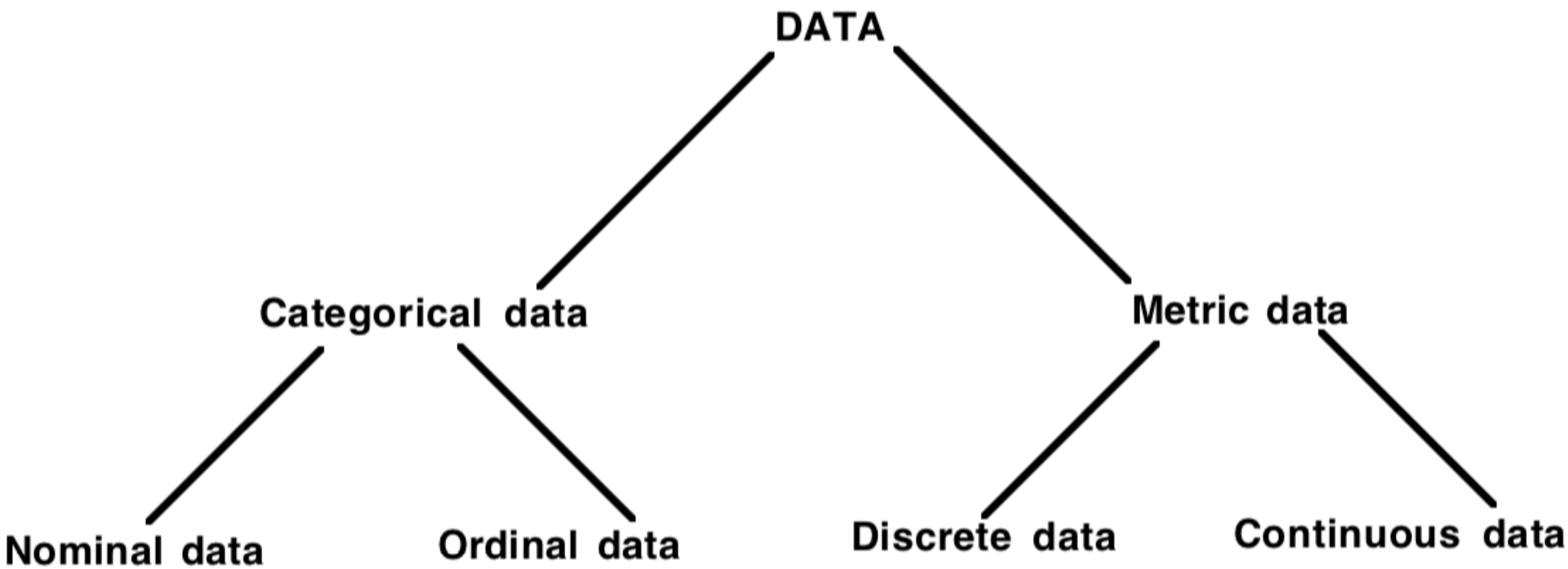
Descriptive Statistics



Previous

Next

Descriptive Statistics



Previous

Next

Descriptive Statistics

$$\text{mean} = \frac{5 + 7 + 8 + 10 + 11 + 12 + 13 + 14}{8} = \frac{80}{8} = 10$$

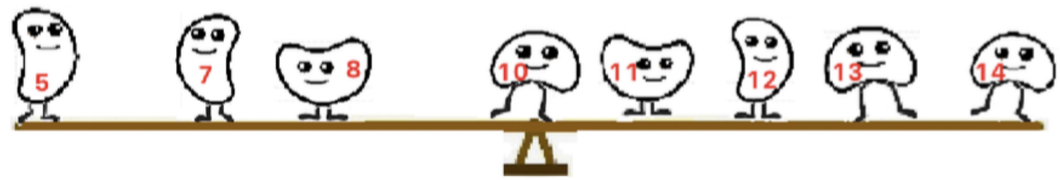
Deviations from the mean

(5 - 10)	=	-5
(7 - 10)	=	-3
(8 - 10)	=	-2
(10 - 10)	=	0
(11 - 10)	=	+1
(12 - 10)	=	+2
(13 - 10)	=	+3
(14 - 10)	=	+4
		0

Previous

Next

Descriptive Statistics



Average of the squares of the deviations from the mean

$$
\begin{array}{lcrcr}
(5 - 10)^2 & = & -5^2 & = & 25 \\
(7 - 10)^2 & = & -3^2 & = & 9 \\
(8 - 10)^2 & = & -2^2 & = & 4 \\
(10 - 10)^2 & = & 0^2 & = & 0 \\
(11 - 10)^2 & = & +1^2 & = & 1 \\
(12 - 10)^2 & = & +2^2 & = & 4 \\
(13 - 10)^2 & = & +3^2 & = & 9 \\
(14 - 10)^2 & = & +4^2 & = & 16 \\
\hline
 & & & & 68
\end{array}
$$

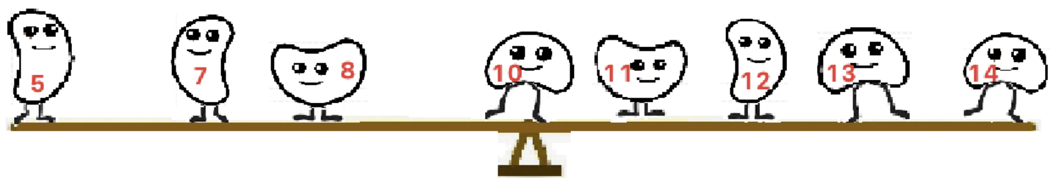
$$\frac{68}{7} = 9.714$$

(Variance)

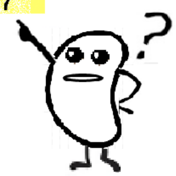
Previous

Next

Descriptive Statistics



$$\frac{68}{7} = 9.714$$



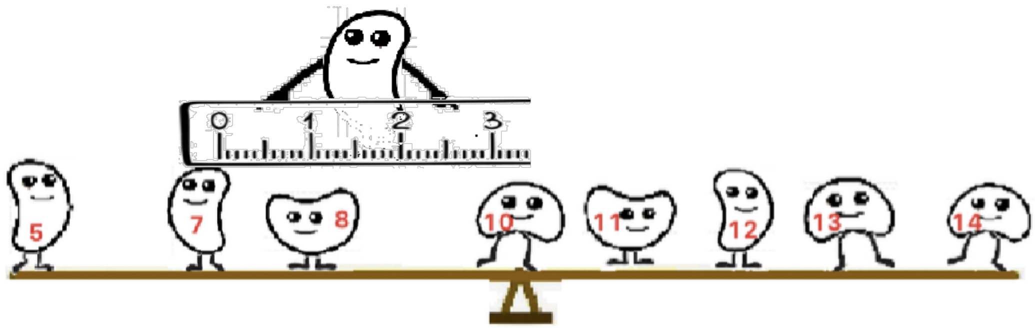
$$\text{mean} = \frac{5 + 7 + 8 + 10 + 11 + \quad + 13 + 14}{8} = 10$$

(Variance)

Previous

Next

Descriptive Statistics



Average of the squares of the deviations from the mean

$$
\begin{aligned}
(5 - 10)^2 &= -5^2 = 25 \\
(7 - 10)^2 &= -3^2 = 9 \\
(8 - 10)^2 &= -2^2 = 4 \\
(10 - 10)^2 &= 0^2 = 0 \\
(11 - 10)^2 &= +1^2 = 1 \\
(12 - 10)^2 &= +2^2 = 4 \\
(13 - 10)^2 &= +3^2 = 9 \\
(14 - 10)^2 &= +4^2 = 16 \\
&\quad\ \ 68
\end{aligned}
$$

$$\frac{68}{7} = 9.714$$

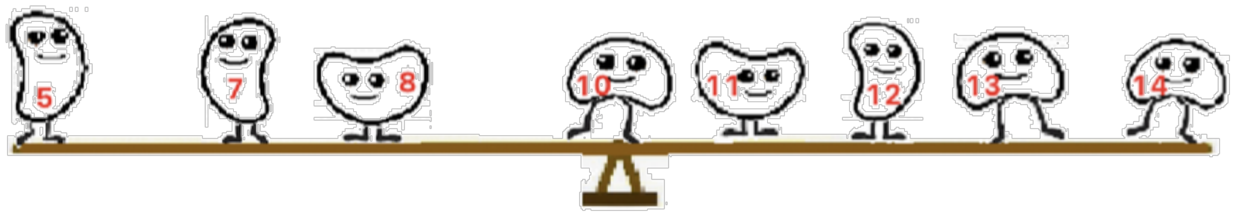
$$\sqrt{9.714} = 3.116$$

(Standard Deviation)

Previous

Next

Descriptive Statistics

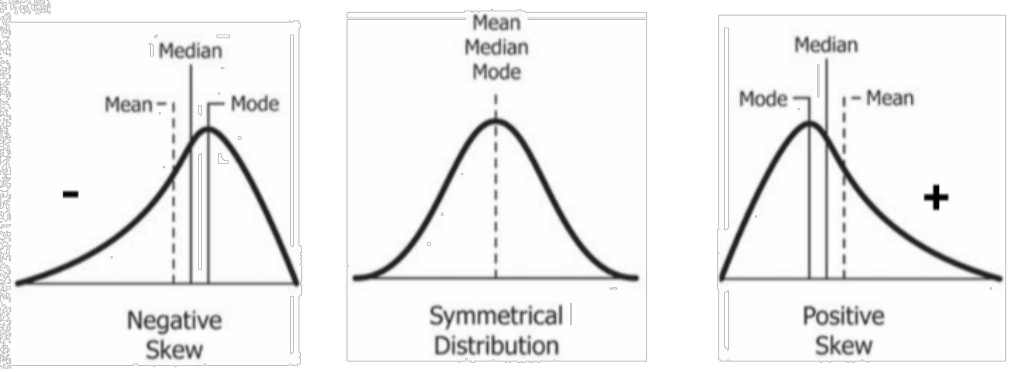


Average of the cubes of the deviations from the mea (Skewness)

$(5 - 10)^3 = -5^3 = -125$
$(7 - 10)^3 = -3^3 = -27$
$(8 - 10)^3 = -2^3 = -8$

$(10 - 10)^3 = 0^3 = 0$

$(11 - 10)^3 = +1^3 = +1$
$(12 - 10)^3 = +2^3 = +8$
$(13 - 10)^3 = +3^3 = +27$
$(14 - 10)^3 = +4^3 = +64$



$$s = \frac{1}{\sigma^3} \frac{\sum_{i=1}^{n} (x_i - \mu)^3}{n-1} = -0.377$$

Previous

Next

Descriptive Statistics

To summarize **generally** if the distribution of data is skewed to the left, the mean is less than the median, which is often less than the mode. If the distribution of data is skewed to the right, the mode is often less than the median, which is less than the mean.

One of the basic tenets of statistics that every student learns in about the second week of intro stats is that in a skewed distribution, the mean is closer to the tail in a skewed distribution.

So in a right skewed distribution (the tail points right on the number line), the mean is higher than the median.

It's a rule that makes sense, and I have to admit, I never questioned it.

But a great article in the Journal of Statistical Education titled "Mean, Median, and Skew: Correcting a Textbook Rule" shows that it really only holds in idealized, unimodal, continuous distributions:

http://jse.amstat.org/v13n2/vonhippel.html.

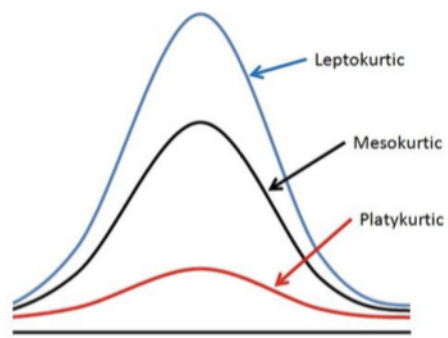
Previous

Next

Descriptive Statistics

Average of the fourth powers of the deviations from the mean (Kurtosis)

$$
\begin{array}{lllll}
(5 - 10)^4 & = & -5^4 & = & +625 \\
(7 - 10)^4 & = & -3^4 & = & +81 \\
(8 - 10)^4 & = & -2^4 & = & +16 \\
 & & & & \\
(10 - 10)^4 & = & 0^4 & = & 0 \\
 & & & & \\
(11 - 10)^4 & = & +1^4 & = & +1 \\
(12 - 10)^4 & = & +2^4 & = & +16 \\
(13 - 10)^4 & = & +3^4 & = & +81 \\
(14 - 10)^4 & = & +4^4 & = & +256
\end{array}
$$



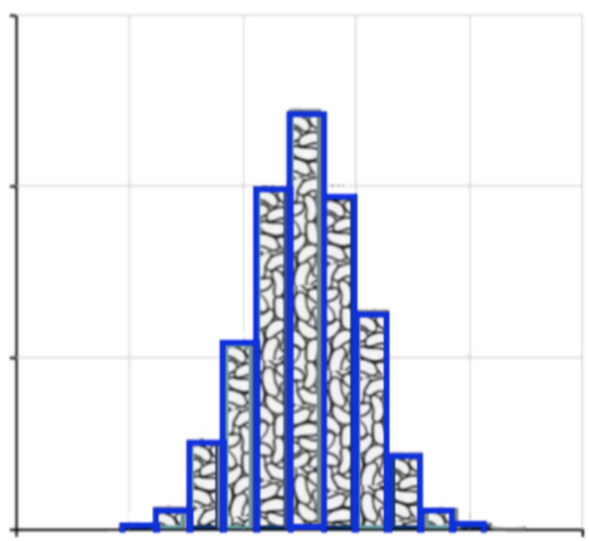
$$\text{Pearson's } K = \frac{1}{\sigma^4}\frac{\sum_{i=1}^{n}(x_i-\mu)^4}{n-1} = 2.009$$

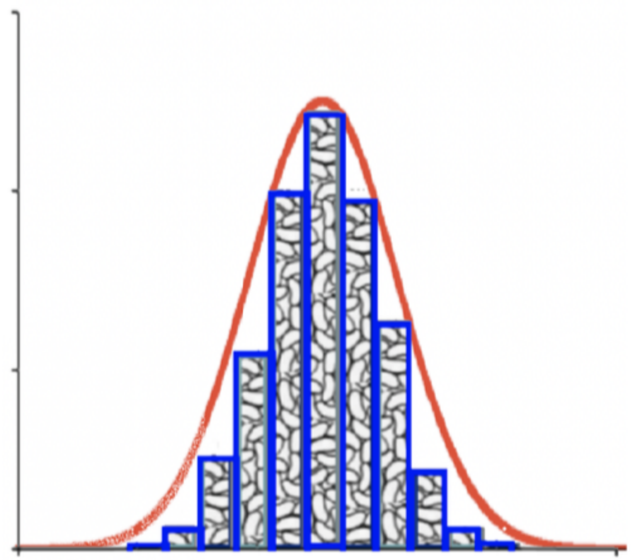
✓ excess kurtosis, $\kappa = K - 3 = -0.990$

Previous

Next

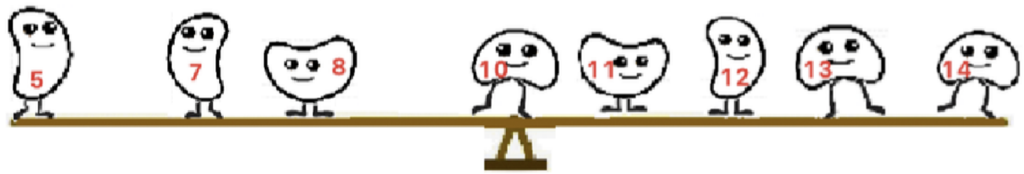
Descriptive Statistics





Descriptive Statistics

- Sensitivities -



$$\frac{1}{n}\sum_{k=1}^{n}\left(x_k \pm c\right) = \bar{x} \pm c \qquad \frac{1}{n}\sum_{k=1}^{n}\left(cx_k\right) = c\bar{x}$$

- Adding or subtracting from each item results shift of the mean by that amount.
- Multiplying each item by 3, enlarges the mean by a factor of 3.

$\forall_k\left(x_k \pm c\right)$ leaves σ^2 and σ unchanged

$\forall_k\left(cx_k\right)$ results in $c^2\sigma^2$ and $|c|\sigma$

- Adding or subtracting from each item results in no change to the variation.
- Multiplying each item by 3, enlarges the variance by a factor of 9 and the standard deviation by a factor of 3.

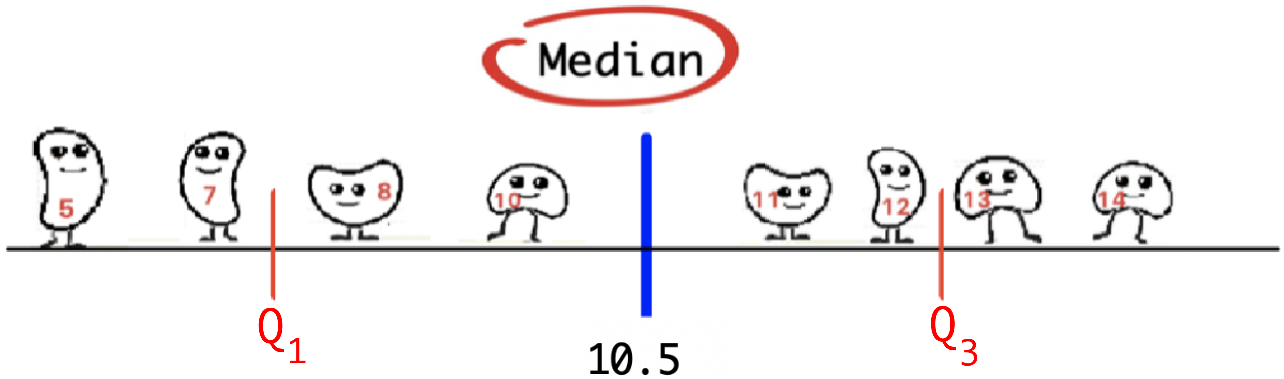
$$E\left(\bar{x}\right) = \mu \qquad E\left(s^2\right) = \sigma^2 \qquad E\left(s\right) \to \sigma$$

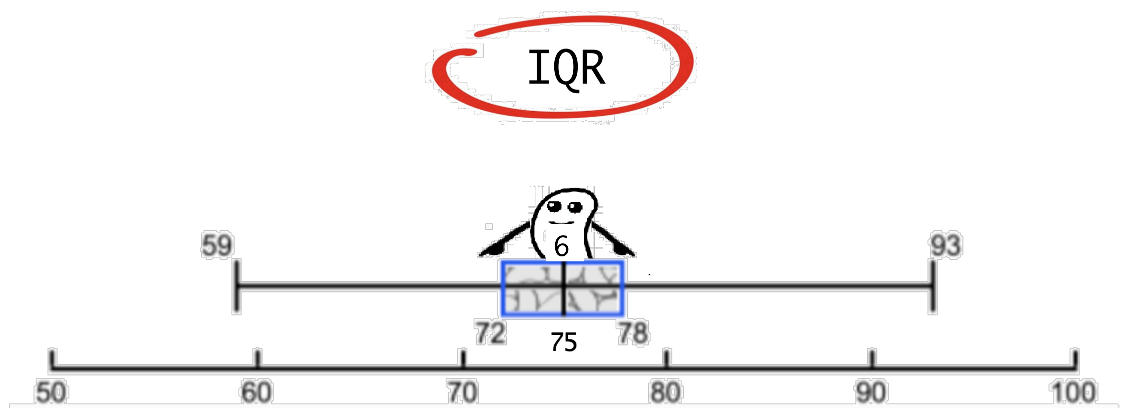
- The sample means, sample variances, and sample standard deviations are good predictors of the population means, variances, and standard deviations.

Previous

Next

Descriptive Statistics

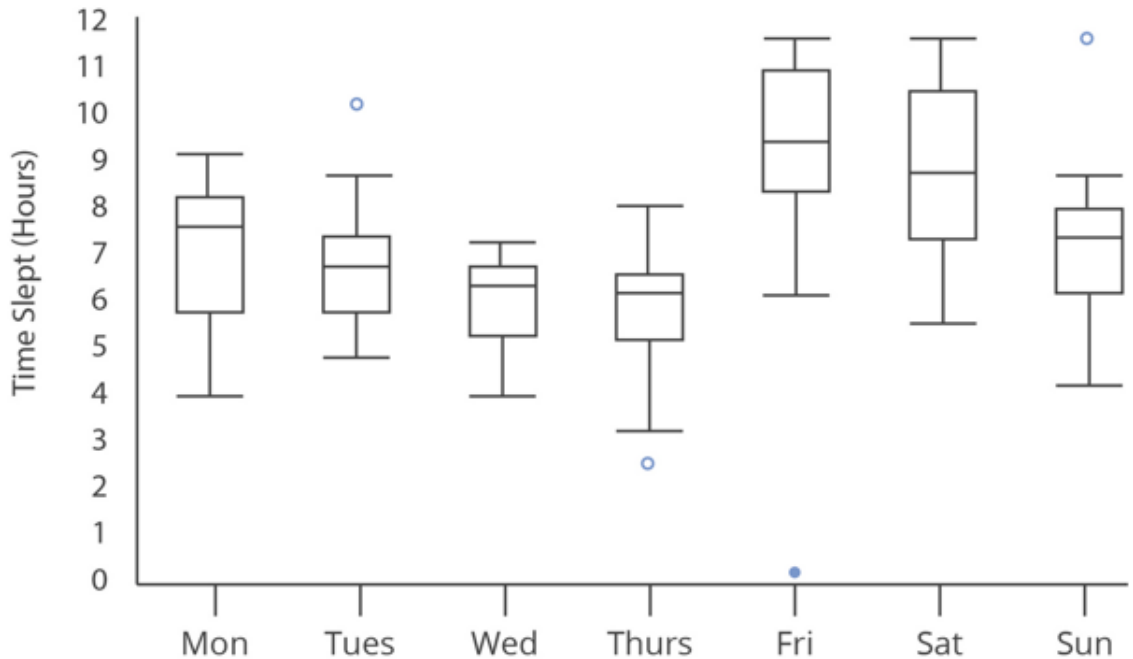




Previous

Next

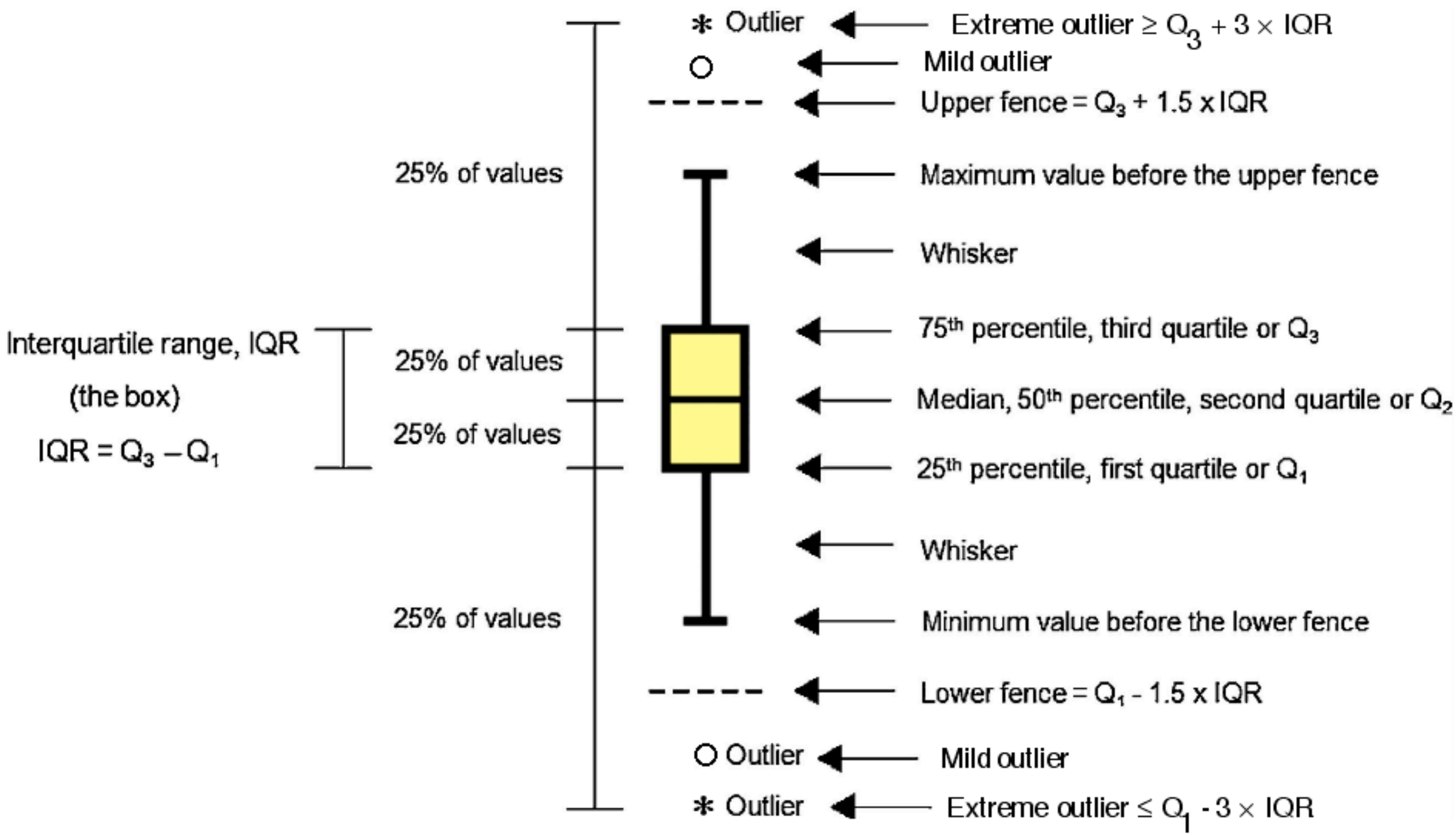
Student's Time Sleeping



Previous

Next

Descriptive Statistics



Previous

Next

COEFFICIENT OF VARIATION

	PRICE PER GALLON OF GASOLINE			
	USA (dollars)			Vietnam (dong)
	$	2.70		11,612
	$	3.06		12,138
	$	2.87		12,980
	$	2.69		13,110
	$	2.71		12,084
MEAN	$	2.81	MEAN	12,385
ST. DEV.	$	0.16	ST. DEV.	638.12
CV =		5.71%	CV =	5.15%

$$CV = \frac{\text{standard deviation}}{\text{mean}} = \frac{\sigma}{\mu}$$

Previous

Next

Go

Statistics

Percentile

Percentile Rank

Mean	
Std. Dev.	
Std. Error	1.720237
Count	50
Minimum	2
Maximum	45
Variance	
Coef. Var.	52.250494%
Range	43
Sum	1164
R.M.S	0
Geom. Mean	
Harm. Mean	
Skewness	
Kurtosis	
Median	
IQR	19.5
10% Tr. Mean	23.375
MAD	
Biweight	23.434159
Mode(s)	

MidRange	
SemiQuartile	
MidQuartile	

$$\sigma \approx \frac{\text{range}}{4}$$

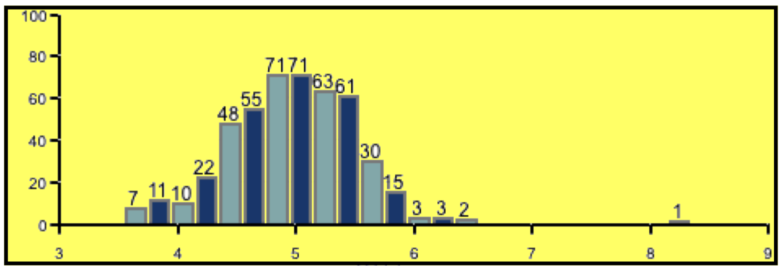
10.75

Previous

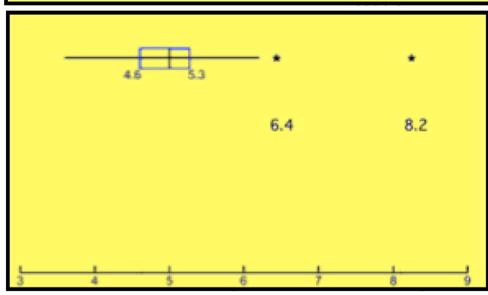
Next

Ungrouped Data

StatDIsk		Excel			
Sample Size, n:	473				
Mean:	4.976744	4.979	=AVERAGE(E2:E474)		
Median:	5	4.979	=MEDIAN(E2:E474)		
Midrange:	5.9				
RMS:	5.005008				
Variance, s^2:	0.2827207	0.281	=VAR(E2:E474)	0.2808	=VARP(E2:E474)
Standard Deviation, s:	0.5317149	0.530	=STDEV(E2:E474)	0.5299	=STDEVP(E2:E474)
Mean Absolute Deviatio	0.4130685				
Range:	4.6				
Coefficient of Variance:	10.68%				
Minimum:	3.6	3.571	=MIN(E2:E474)		
1st Quartile:	4.6	4.639	=QUARTILE(E2:E474,1)		
2nd Quartile:	5	4.979	=QUARTILE(E2:E474,2)		
3rd Quartile:	5.3	5.336	=QUARTILE(E2:E474,3)		
Sum:	2354	2355.007	=SUM(E2:E474)		
Sum of Squares:	11848.7	11858.106	=SUMSQ(E2:E474)		



	skewness->	0.27367	=SKEW(E2:E474)	
	kurtosis ->	2.460723	=KURT(E2:E474)	
	mode ->	#N/A	=MODE(E2:E474)	
	geometric mean ->	#NUM!	=GEOMEAN(E2:E474)	4.948
	harmonic mean ->	4.92186	=HARMEAN(E2:E474)	4.919
10%	trimmed mean ->	4.981582	=TRIMMEAN(E2:E474,H29)	



midrange ->	5.8854465	RANGE / 2	
range ->	4.629107	MAX - MIN	
rms ->	5.006994364	SQRT(SUMSQ / n)	
MAD ->	0.410851401	RANGE / 2	
coefficient of variation	10.65480494 %	STDEV/MEAN %	
midquartile ->	4.987591	(Q1+Q3) / 2	
semiquartile ->	0.348219	(Q3-Q1) / 2	
IQR ->	0.7	Q3-Q1	
biweight ->			4.979

	PERCENTILES		
95%	5.749	=PERCENTILE(E2:E474,G56)	
	RANK		
5.3	73%	=PERCENTILERANK(E2:E474,G59)	

Previous

Next

Ungrouped Data

Statistics

Mean	114.5
Std. Dev.	14.584064
Std. Error	1.882795
Count	60
Minimum	82
Maximum	139
Variance	212.694915
Coef. Var.	12.737174%
Range	57
Sum	6870
R.M.S	115.409705
Geom. Mean	113.549421
Harm. Mean	112.563352
Skewness	-0.311019
Kurtosis	-0.852841
Median	115
IQR	19.5
10% Tr. Mean	114.9375
MAD	12.1
Biweight	114.816251
Mode(s)	124

MidRange	28.5
SemiQuartile	9.75
MidQuartile	114.25
$\sigma \approx \frac{\text{range}}{4}$	14.25

Lower Bound	Upper Bound	Frequency
138	140	1
135	137	5
132	134	1
129	131	3
126	128	4
123	125	9
120	122	3
117	119	3
114	116	5
111	113	2
108	110	7
105	107	2
102	104	1
99	101	2
96	98	4
93	95	3
90	92	2
87	89	2
84	86	0
81	83	1

82
88
88
91
91
94
94
94
97
97
97
97
100
100
103
106
106
109
109
109
109
109
109
109
112
112
115

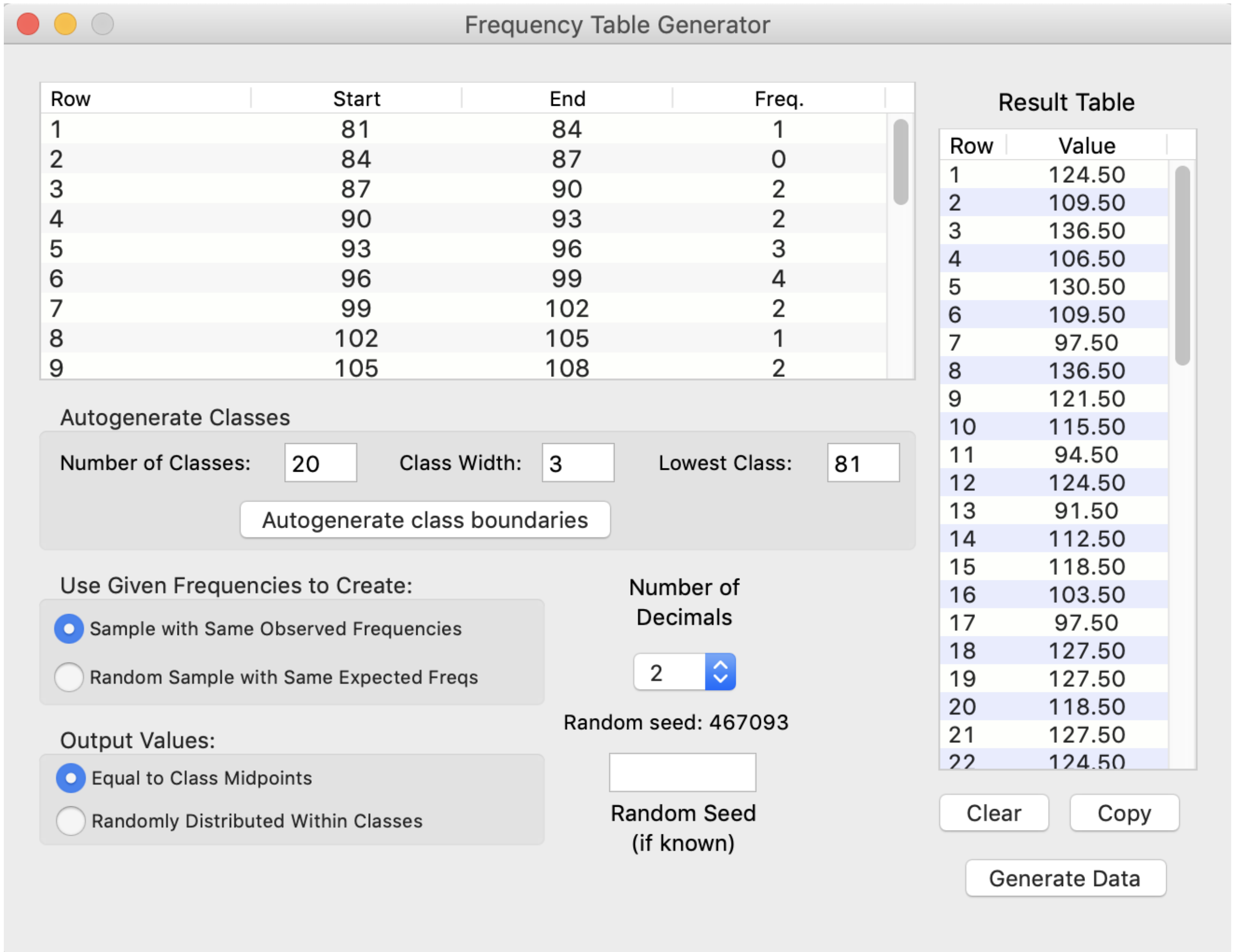
Percentile

Percentile Rank

Previous

Next

Grouped Data

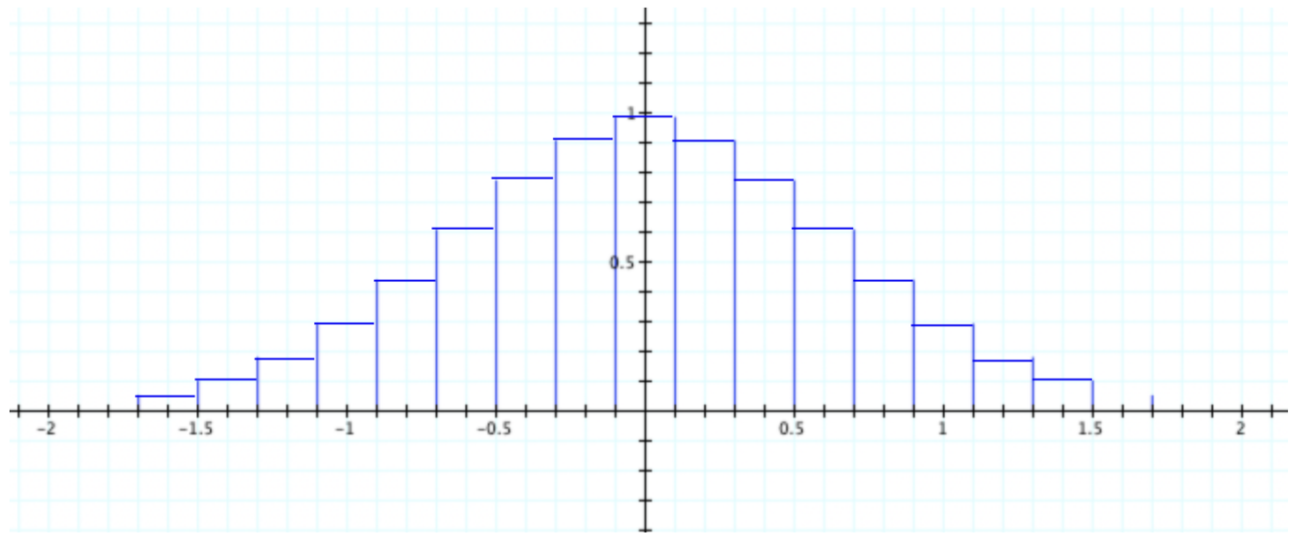


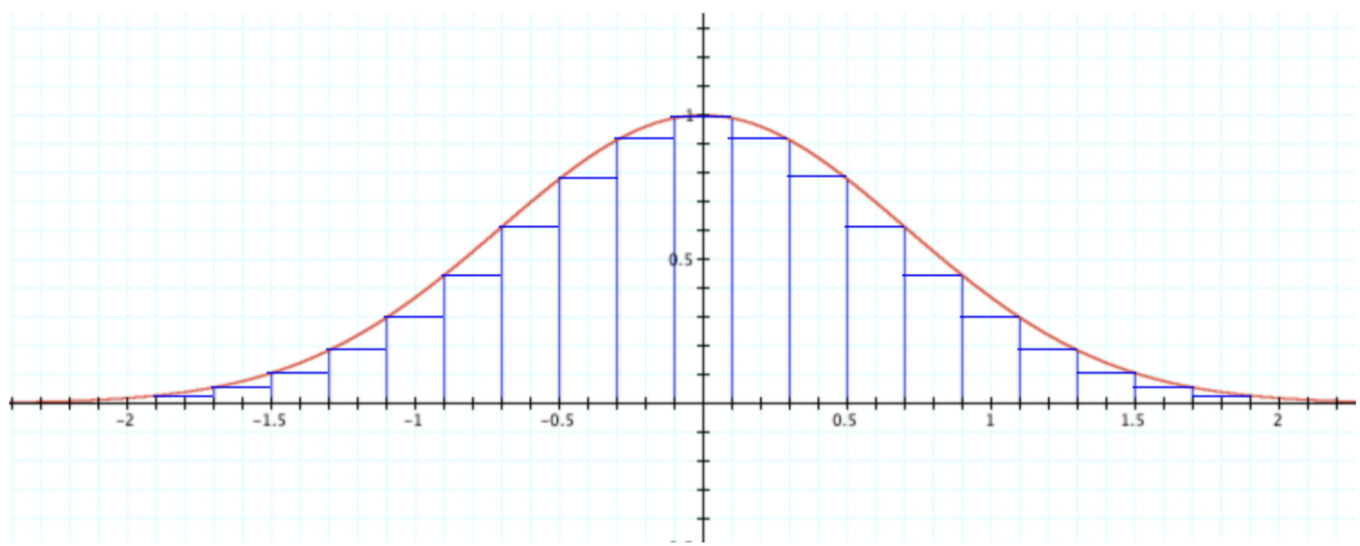
Grouped Data

Previous

Next

Distribution "Envelope"





Previous

Next

Go

Stem and Leaf Plot

10
9
8
7
6
5
4
3
2
1
0

10 |
9 |
8 |
7 |
6 | 145
5 | 2
4 | 6
3 |
2 | 045589
1 |
0 | 0

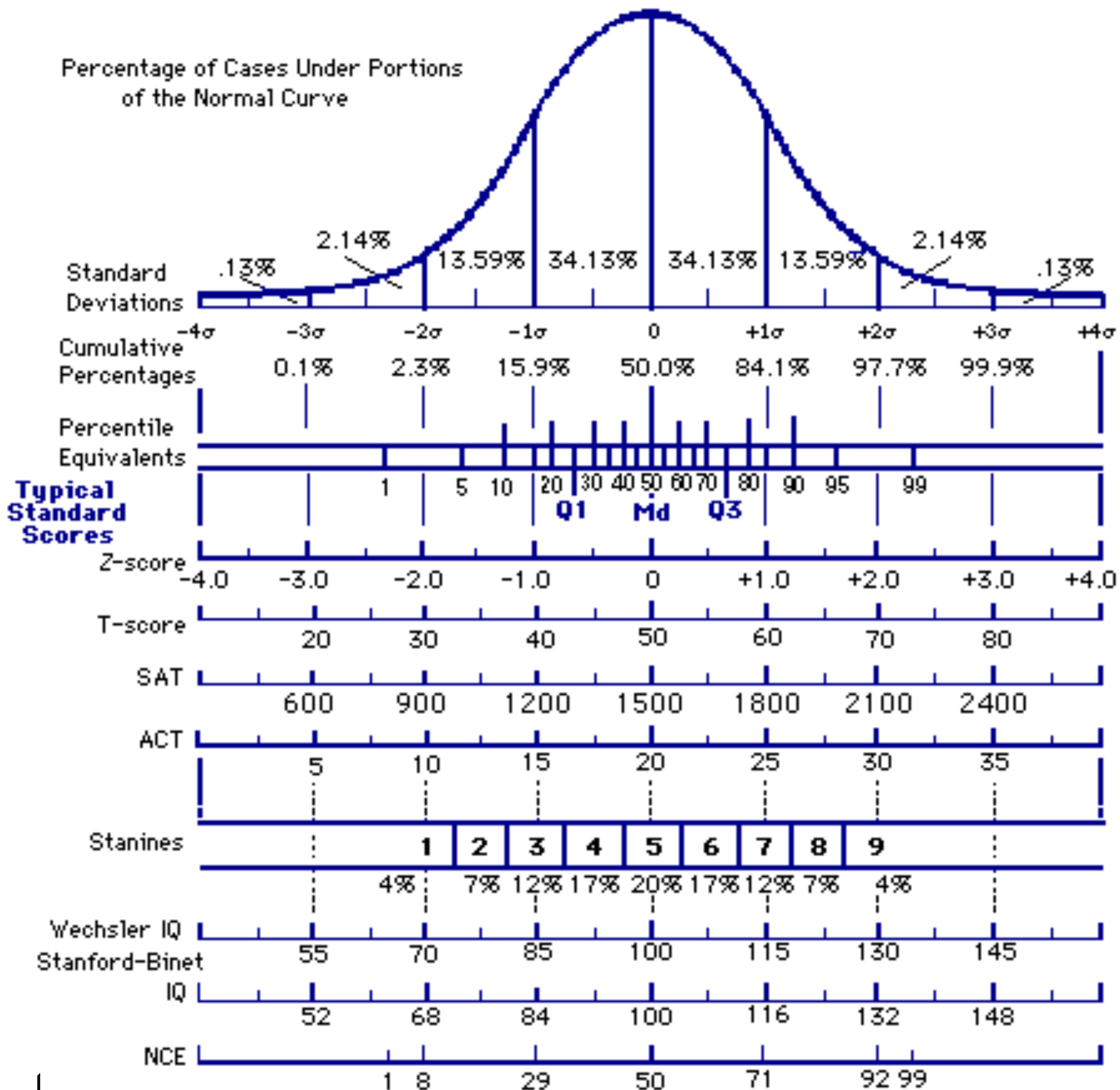
0
20
24
25
25
28
29
46
52
61
64
65

Girls		Boys
5	14	
7, 5, 5, 5, 4	15	3, 8, 9
8, 4, 2, 1, 0	16	2, 5, 7, 7, 7, 8, 8, 9
9, 8, 7, 6, 6, 4, 2, 1, 1, 0, 0	17	0, 2, 3, 6, 6, 7, 7
	18	0, 1, 4, 5

Previous

Next

The Normal Curve, Percentiles, and Standard Scores



Previous

Next

score -> 45

μ -> 65.5

σ -> 8.2

Get z

$$z = \frac{x - \mu}{\sigma}$$

z- score :

-2.5

standardize

130 <- standardized mean

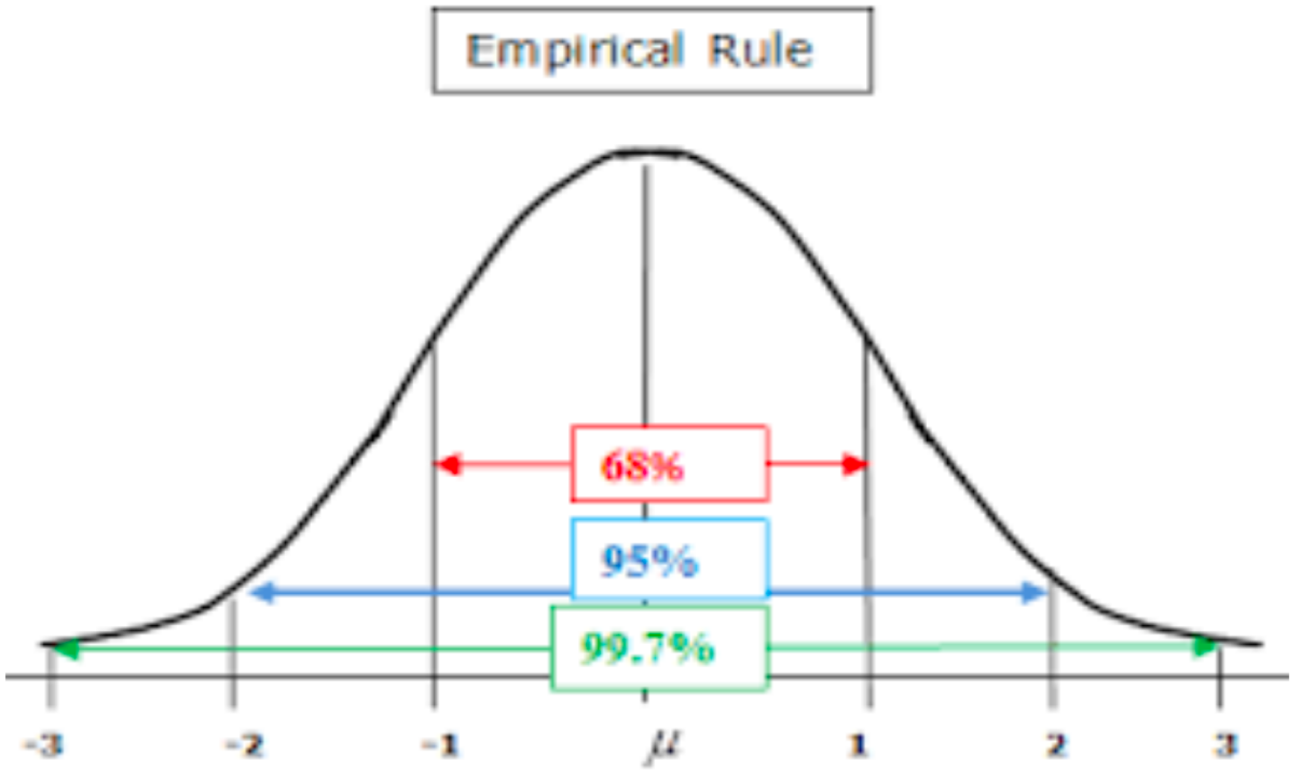
15 <- standardized st.dev.

$$x = \mu + z\sigma$$

92.5

Previous

Next



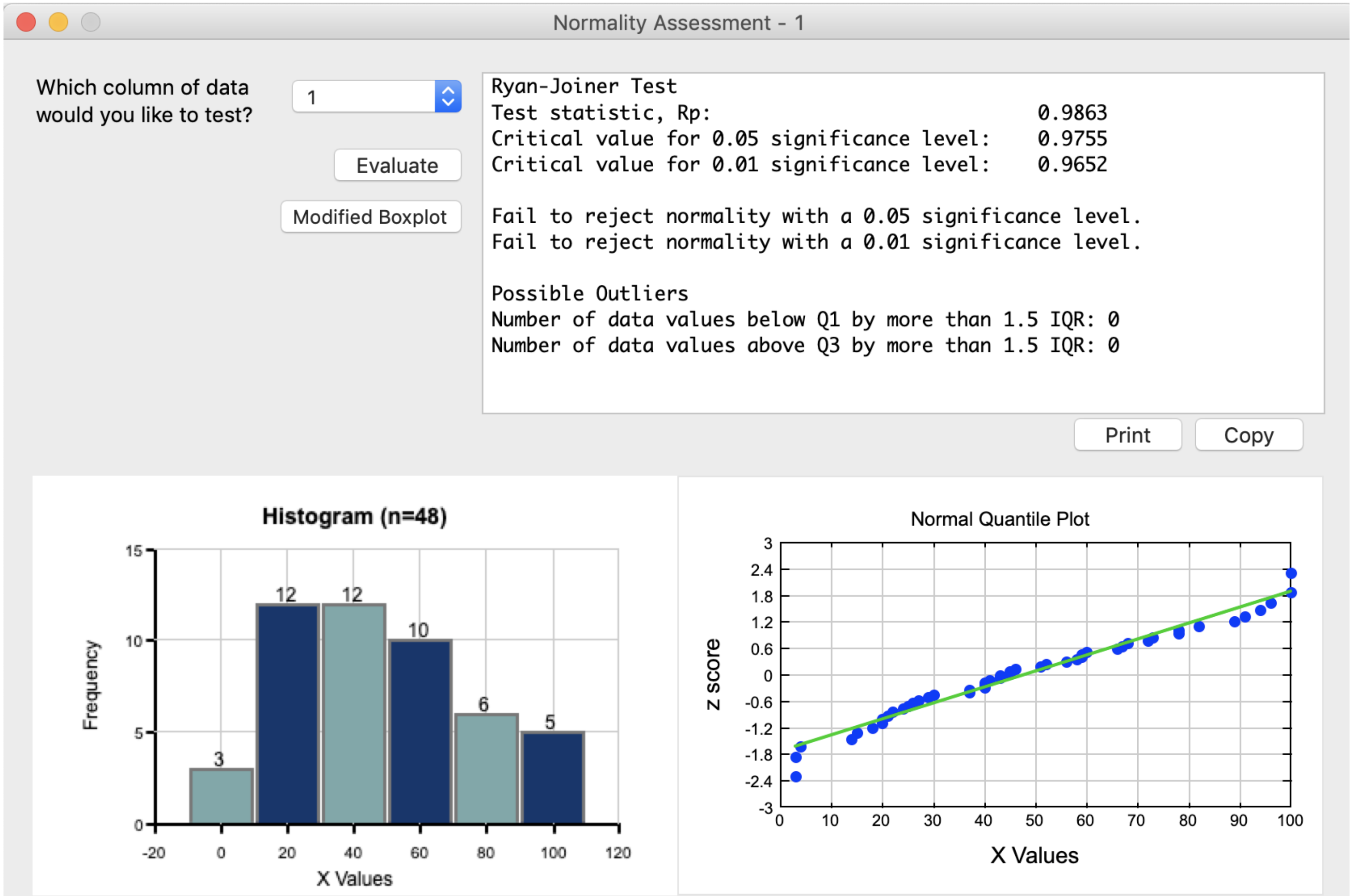
significantly low

significantly high

Previous

Next

Normality Assessment



Previous

Next

Normality Assessment

Shapiro-Wilk Normality Test

Shapiro, S. S. and Wilk, M. B. (1965). "Analysis of variance test for normality (complete samples)", *Biometrika* 52: 591–611.
Online version implemented by Simon Dittami (2009)

Paste data here: (results below)

```
59
51
43
96
78
14
100
45
26
58
30
43
100
24
72
3
15
56
```

Calculate | Clear all

Results:

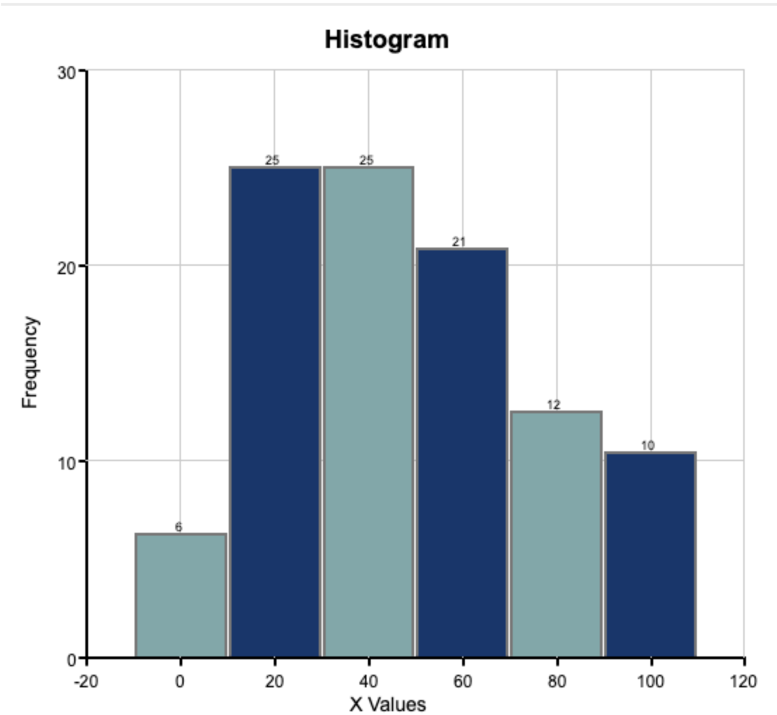
```
n = 48
Mean = 48.06250000000001
SD = 27.112379405380054
W = 0.9506724048156191

Threshold (p=0.01) = 0.9290000200271606 --> HO accepted
Threshold (p=0.05) = 0.9470000267028809 --> HO accepted
Threshold (p=0.10) = 0.9539999961853027 --> HO rejected
```

--> Your data is not normally distributed p<0.10

Previous

Next

Relative Frequency Histograms and Probability Distributions



Although this text does not discuss the concept of probability density in detail, you should keep the following ideas in mind about the curve that describes a continuous distribution (like the normal distribution).

- **First, the area under the curve equals 1.**
- **Second, the probability of any exact value of *X* is 0.**
- **Finally, the area under the curve and bounded between two given points on the *X*-axis is the probability that a number chosen at random will fall between the two points.**

Previous

Next

125

How many?

stemPlot

10	
9	
8	
7	
6	
5	
4	
3	
2	
1	
0	

6
4
3
1
2
4
5
6
5
1
6
5
3
6
4
2
4
2
4
4
4
2
1
3
5
1

	a	*b*	*c*	
uniform	1	100	1	
	n	*p*		
binomial	200	.25		
	λ			
Poisson	40			exponential
normal	65	15		
	μ	σ		

Previous

Next

Histogram

250

How many?

6
4
3
1
2
4
5
6
5
1
6
5
3
6
4
2
4
2
4
4
4
2
1
3
5
1

uniform a -> 1 b -> 6 c -> 1

binomial n -> 200 p -> .4

Poisson λ -> 40

normal μ -> 65 σ -> 20

exponential

Previous

Next

Histogram

Population Size: 1000

create_new Populsation

use given Populsation

Number of intervals: 20

6
4
3
1
2
4
5
6
5
1
6
5
3
6
4
2
4
2
4
4
4
2
1

42
46
53
58
48
56
51
52
46
58
44
44
46
52
57
48
48
51
42
50
43
51
49
46
50
48
48
48
55
52
51
44
45
45
51
44
45
42
46
48
51
47
40
47
54

1,0
2,0
3,0
4,0
5,0
6,0
7,0
8,0
9,0
10,0
11,0
12,0
13,0
14,0
15,0
16,0
17,0
18,0
19,0
20,0
21,0
22,0
23,0
24,0
25,0
26,0
27,0
28,0
29,0
30,0
31,0
32,0
33,0
34,0
35,2
36,1
37,2
38,2
39,2
40,6
41,6
42,14
43,21
44,27
45,32

SAMPLE

Sample. Size: 50

Number of Samples: 500

Mean of Sampling Distribution: 48.538

Standard Deviation of Sampling Distribution: 4.224755

Standard Error: (Calculated) 4.130235

MU : 48.599

SIGMA : 29.205174

VAR : 852.94219

Previous

Next