Statistics - Lying without sinning?

In North Dakota, 54 Million Beer Bottles by the side of the Road
April 01 2002
VitalSTATS
When you come across an implausible claim, do the math
South Dakota's Pierre Capital Journal reports (Mar. 1) that "an average of 650 beer cans and bottles are tossed per mile of road annually." The statistic is attributed to Dennis W. Brezina, an activist against drunk-driving.

But how did he come up with his data? According to the Journal, Brezina traveled "highways across the nation to determine whether the problem he perceived was widespread. He made two trips to South Dakota, one in 1998 and another in 2000." He counted "cans and bottles in ditches in May of both years" and claimed to have found an average of "one beer can or bottle every 16 feet when walking randomly selected stretches of ditch."

But the math appears a little blurry. The web site of the South Dakota Department of Transportation claims that the state "has 83,472 miles of highways, roads and streets." Assuming Brezina's estimate is correct, South Dakotans appear to be world-class litterbugs, tossing aside approximately 54,256,800 bottles or cans every year. According to the Census Bureau there are 754,844 people in South Dakota. So, according to Brezina, the average resident throws at least 71 beer bottles or cans on the side of the road every year.

For more
Check out
www.STATS.org

Statistics Equations

mean $$\bar{x} = \frac{\sum x_i}{n}$$

Standard Deviation $$s = \sqrt{\frac{\sum (\bar{x} - x_i)^2}{n-1}}$$

$$\mathrm{var}\,iance = s^2$$

Relative Standard Deviation $$\%RSD = \frac{s}{\bar{x}} \cdot 100$$

Student's t $$\mu = x_i \pm t \cdot s$$

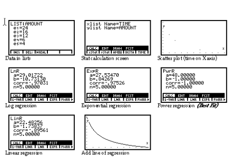
Standard Deviation of the Mean $$s_m = \frac{s}{\sqrt{n}}$$

Confidence Interval $$\mu = \bar{x} \pm t \cdot \frac{s}{\sqrt{n}}$$

Learn to use your calculator's statistical functions to calculate mean and standard deviation. You'll save yourself a lot of work.

http://www.willamette.edu/~mjaneba/help/TI-85-stats.htm

http://www2.ohlone.edu/people2/joconnell/ti/



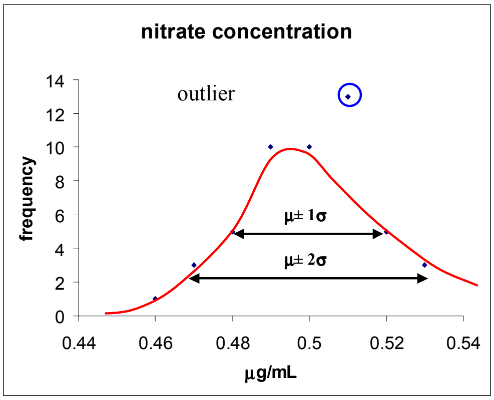
Nitrate Concentrations (μg/mL)

Trial 1	Trial 2	Trial 3	Trial 4	Trial 5	Trial 6	Trial 7	Trial 8	Trial 9	Trial 10	
0.51	0.51	0.51	0.5	0.51	0.49	0.52	0.53	0.5	0.47	
0.51	0.52	0.53	0.48	0.49	0.5	0.52	0.49	0.49	0.5	
0.49	0.48	0.46	0.49	0.49	0.48	0.49	0.49	0.51	0.47	
0.51	0.51	0.51	0.48	0.5	0.47	0.5	0.51	0.49	0.48	
0.51	0.5	0.5	0.53	0.52	0.52	0.5	0.5	0.51	0.51	
0.506	**0.504**	**0.502**	**0.496**	**0.502**	**0.492**	**0.506**	**0.504**	**0.5**	**0.486**	**mean**

average	**0.4998**
stdev	**0.01647**

mg/mL	frequency
0.53	**3**
0.52	**5**
0.51	**13**
0.5	**10**
0.49	**10**
0.48	**5**
0.47	**3**
0.46	**1**

Let's Graph the Data!

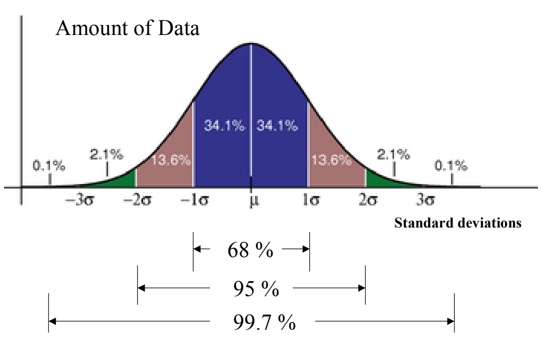


Statistics in the News

Outliers Disrupt the Mean
January 01 1999

In 1984, according to Larry Gonick and Woollcott Smith, the University of Virginia announced that the mean starting salary of its graduates from the Department of Rhetoric and Communications was a very hefty $55,000 per year. But before you abandon your computer science training for speech classes, you should know that the graduating class contained a significant "outlier," or extreme data point not typical of the rest of the data set - Ralph Sampson, future NBA All-Star, who majored in speech. It would have been better to learn the median salary, the data point in the middle of the set.

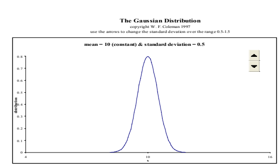
The Normal Distribution

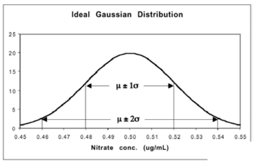


Total % of the data covered by distribution

The Normal Distribution

The effect of standard deviation on width





What's the purpose of the t-table?

t-table

Degrees of freedom (n-1)	95% confidence	98% confidence	99% confidence
1	12.71	31.82	63.66
2	4.30	6.96	9.92
3	3.18	4.54	5.84
4	2.78	3.75	4.60
5	2.57	3.36	4.03
6	2.45	3.14	3.71
7	2.36	3.00	3.50
8	2.31	2.90	3.36
9	2.26	2.82	3.25
10	2.23	2.76	3.17
12	2.18	2.68	3.05
14	2.14	2.62	2.98
16	2.12	2.58	2.92
18	2.10	2.55	2.88
20	2.09	2.53	2.85
30	2.04	2.46	2.75
50	2.01	2.40	2.68

Confidence Interval

$$\overline{x} = \mu \pm t \cdot s_m = \mu \pm t \cdot \frac{s}{\sqrt{n}}$$

95 % confidence interval

The mean falls within the true value 95 % of the time

Confidence Interval Exercise

$$\overline{x} = \mu \pm t \cdot s_m = \mu \pm t \cdot \frac{s}{\sqrt{n}}$$

Calculate the 95, 98 and 99 % confidence intervals

For the nitrate concentration data

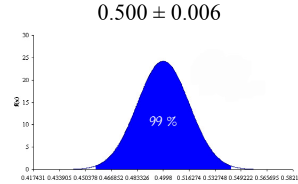
95 % 0.500 ± 0.005

98 % 0.500± 0.006

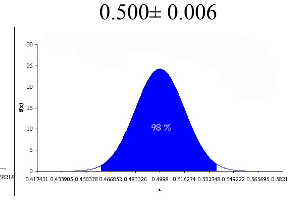
99 % 0.500 ± 0.006

50 % 0.500 ± 0.002

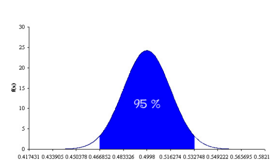
0.500 ± 0.006



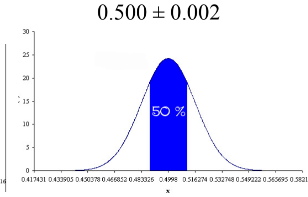
0.500± 0.006



0.500 ± 0.005



0.500 ± 0.002



Student's *t* and the Law

Clearly, the meanings of 1.083 ± 0.007 and 1.0 ± 0.4 are very different. As a person who will either derive or use analytical results, you should be aware of this warning published in a report entitled "Principles of Environmental Analysis":

Analytical chemists must always emphasize to the public that the single most important characteristic of any result obtained from one or more analytical measurements is an adequate statement of its uncertainty interval. Lawyers usually attempt to dispense with uncertainty and try to obtain unequivocal statements: therefore, an uncertainty interval must be defined in cases involving litigation and or enforcement proceedings. Otherwise, a value of 1.001 without a specified uncertainty, for example may be views as legally exceeding a permissible level of 1.

L. K. Keith, W. Crummett, J. Deegan Jr., R. A. Libby, J. K. Taylor, and G. Wentler, *Analytical Chemis*try, **55**, 2210 (1983).