Statistics

Chapter 4

Is My Blood Glucose Reading Correct?

What do varying readings indicate?

- Serious health issues
- Poor technique
- Malfunctioning meter

Seasontime/Shutterstock

- Run a control sample to test one's technique and glucose meter.

Measurement of control sample	Control sample concentration
90 mg/dL	
94 mg/dL	
85 mg/dL	99 mg/dL
95 mg/dL	
93 mg/dL	
Average = 91.4 mg/dL	

Is my blood glucose reading correct?

All Measurements Have Experimental Uncertainty

Statistics gives us tools to:

- Accept conclusions that have a high probability of being correct
- Reject conclusions that have a high probability of being incorrect

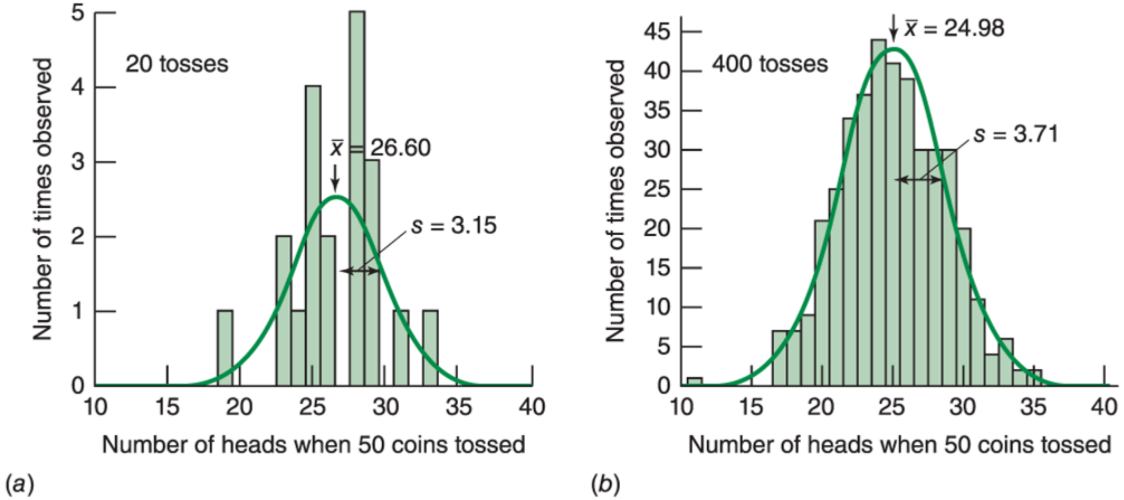
Section 4-1
Gaussian Distribution

Gaussian Distribution

If an experiment is repeated a great many times *and* if the errors are purely random:

- The results tend to cluster symmetrically about the average value.
- The more times the experiment is repeated, the more closely the results approach a **Gaussian distribution.**

Figure 4-1



Harris/Lucy, *Quantitative Chemical Analysis*, 10e, © 2020 W. H. Freeman and Company

Usually we repeat an experiment 3–5 times (not 400 times).

From small data sets we can estimate properties of a hypothetical large set.

Mean Value and Standard Deviation

Mean (average): the sum of a set of results divided by the number of values in the set

$$\bar{x} = \frac{\sum_i x_i}{n}$$

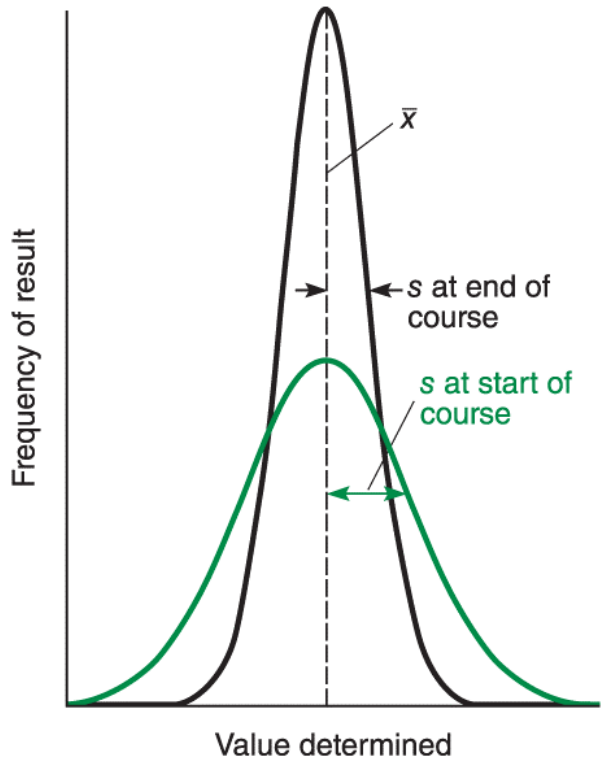
as n increases, $\bar{x} \rightarrow \mu$

Standard deviation: measures how closely data are clustered about the mean

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

as n increases, $s \rightarrow \sigma$

Figure 4-2



Standard Deviation

The smaller the standard deviation, *s*, the more closely the data are clustered about the mean.

Precision: reproducibility

Accuracy: nearness to the "truth"

- Experiments with a small standard deviation are more **precise** than experiments with a large standard deviation.
- Greater precision does not necessarily imply greater **accuracy**.

- Express the mean and standard deviation in the form $\bar{x} \pm s\ (n = \_\_\_)$.
- The average and the standard deviation should both end in the *same decimal place.*

Other Statistical Parameters

Degrees of freedom:

$$\text{Degrees of freedom} = n - 1$$

Variance: square of the standard deviation

$$\text{Variance} = \sigma^2$$

Relative standard deviation (coefficient of variation): standard deviation expressed as a percentage of the mean

$$\text{Relative standard deviation} = 100 \times \frac{s}{\bar{x}}$$

Example: Mean and Standard Deviation (1 of 4)

Find the average, standard deviation, and relative standard deviation for 821, 783, 834, and 855.

Example: Mean and Standard Deviation (2 of 4)

Solution: The average is

$$\bar{x} = \frac{821 + 783 + 834 + 855}{4} = 823._2$$

To avoid accumulating round-off errors, retain one more digit in the mean than was present in the original data. The standard deviation is

$$s = \sqrt{\frac{(821 - 823._2)^2 + (783 - 823._2)^2 + (834 - 823._2)^2 + (855 - 823._2)^2}{(4 - 1)}} = 30._3$$

Example: Mean and Standard Deviation (3 of 4)

Solution: The average and the standard deviation should both end at the *same decimal place*. For $\bar{x} = 823._2$, we will write $s = 30._3$. The relative standard deviation is the percent relative uncertainty:

$$\text{Relative standard deviation} = 100 \times \frac{s}{\bar{x}} = 100 \times \frac{30._3}{823._2} = 3.7\%$$

Example: Mean and Standard Deviation (4 of 4)

Test Yourself: If each of the four numbers 821, 783, 834, and 855 in the example is divided by 2, how will the mean, standard deviation, and relative standard deviation be affected?

Spreadsheets for Average and Standard Deviation

Spreadsheets have built-in statistical functions

- Average:

  =AVERAGE(B1:B4)

- Standard deviation:

  =STDEV.S(B1:B4)

	A	B
1		821
2		783
3		834
4		855
5	Average =	823.25
6	Std dev =	30.27
7	B5 = AVERAGE(B1:B4)	
8	B6 = STDEV.S(B1:B4)	

Standard Deviation and Probability (1 of 2)

Gaussian curve:

$$y = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

The probability of observing a value within a certain range is proportional to the area of that range.

Express deviations from the mean value in multiples, *z*, of the standard deviation. We transform *x* into *z*

$$z \approx \frac{x-\mu}{\sigma} \approx \frac{x-\bar{x}}{s}$$

Figure 4-3

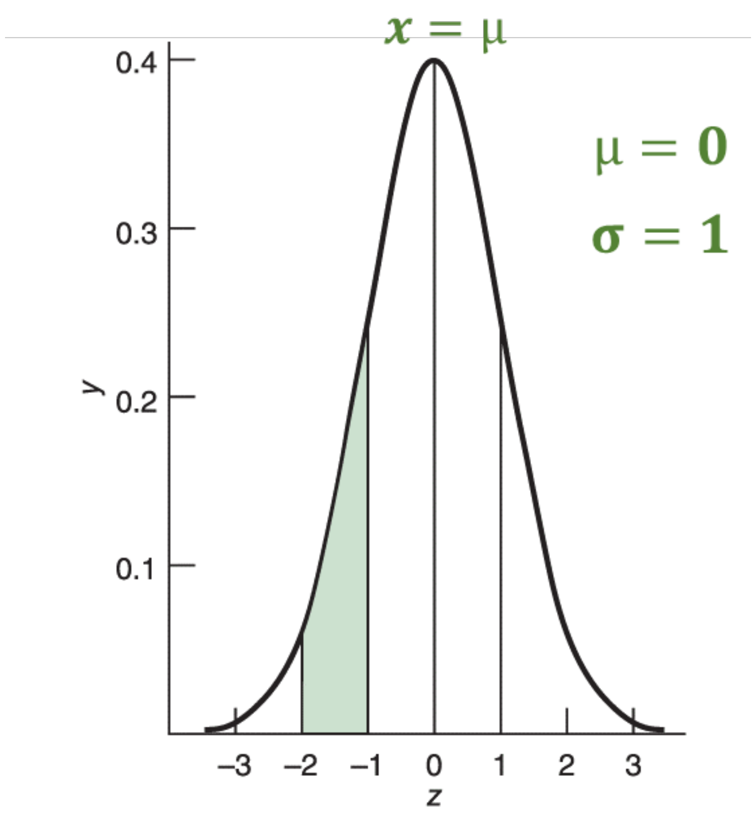


Table 4-1 Ordinate and area for the normal (Gaussian) error curve, $y = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$

$\|z\|$[a]	y	Area[b]	$\|z\|$	y	Area	$\|z\|$	y	Area
0.0	0.398 9	0.000 0	1.4	0.149 7	0.419 2	2.8	0.007 9	0.497 4
0.1	0.397 0	0.039 8	1.5	0.129 5	0.433 2	2.9	0.006 0	0.498 1
0.2	0.391 0	0.079 3	1.6	0.110 9	0.445 2	3.0	0.004 4	0.498 650
0.3	0.381 4	0.117 9	1.7	0.094 1	0.455 4	3.1	0.003 3	0.499 032
0.4	0.368 3	0.155 4	1.8	0.079 0	0.464 1	3.2	0.002 4	0.499 313
0.5	0.352 1	0.191 5	1.9	0.065 6	0.471 3	3.3	0.001 7	0.499 517
0.6	0.333 2	0.225 8	2.0	0.054 0	0.477 3	3.4	0.001 2	0.499 663
0.7	0.312 3	0.258 0	2.1	0.044 0	0.482 1	3.5	0.000 9	0.499 767
0.8	0.289 7	0.288 1	2.2	0.035 5	0.486 1	3.6	0.000 6	0.499 841
0.9	0.266 1	0.315 9	2.3	0.028 3	0.489 3	3.7	0.000 4	0.499 904
1.0	0.242 0	0.341 3	2.4	0.022 4	0.491 8	3.8	0.000 3	0.499 928
1.1	0.217 9	0.364 3	2.5	0.017 5	0.493 8	3.9	0.000 2	0.499 952
1.2	0.194 2	0.384 9	2.6	0.013 6	0.495 3	4.0	0.000 1	0.499 968
1.3	0.171 4	0.403 2	2.7	0.010 4	0.496 5	∞	0	0.5

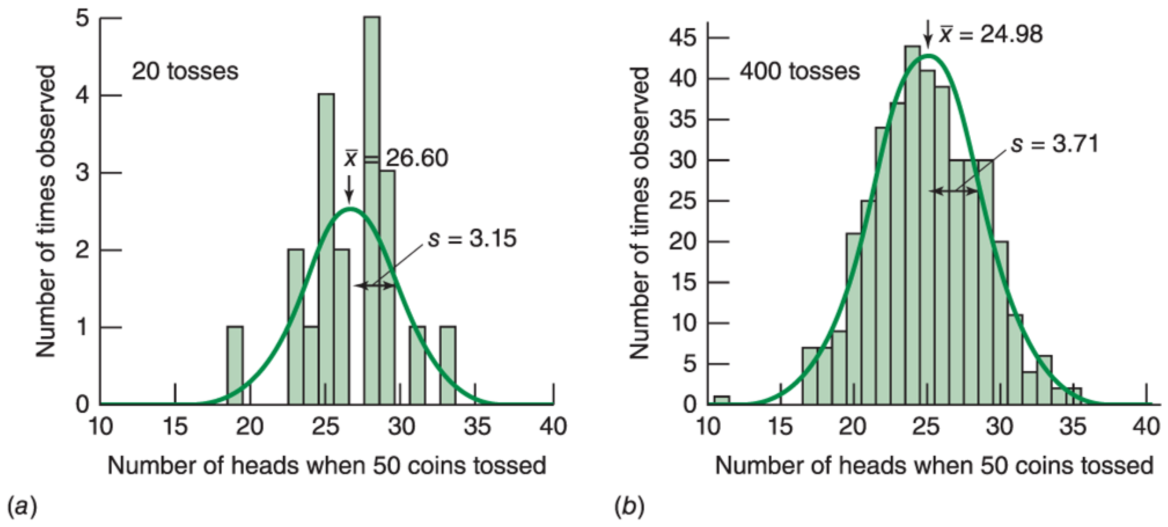
a. $z = (x - \mu)/\sigma$

b. *The area refers to the area between z = 0 and z = the value in the table. Thus the area from z = 0 to z = 1.4 is 0.419 2. The area from z = −0.7 to z = 0 is the same as from z = 0 to z = 0.7. The area from z = −0.5 to z = +0.3 is (0.191 5 + 0.117 9) = 0.309 4. The total area between z = −∞ and z = +∞ is unity.*

Example: Area Under a Gaussian Curve (1 of 3)

For many tosses of a set of 50 coins in Figure 4-1, probability theory predicts a mean of 25.00 heads and a standard deviation of 3.54. How many tosses are expected to have fewer than 15 heads if the 50 coins were tossed 400 times?

Figure 4-1



Example: Area Under a Gaussian Curve (2 of 3)

Solution: We express the desired interval in multiples of the standard deviation and then find the area of the interval in Table 4-1.

Because $\mu = 25.00$ and $\sigma = 3.54$, $z = (15 - 25.00) / 3.54 = -2.82 \approx -2.8$.

From Table 4-1 the area between the mean and z = −2.8 is 0.497 4.

The entire area from −∞ to the mean value is 0.500 0, so the area from −∞ to −2.8 is 0.500 0 − 0.497 4 = 0.002 6.

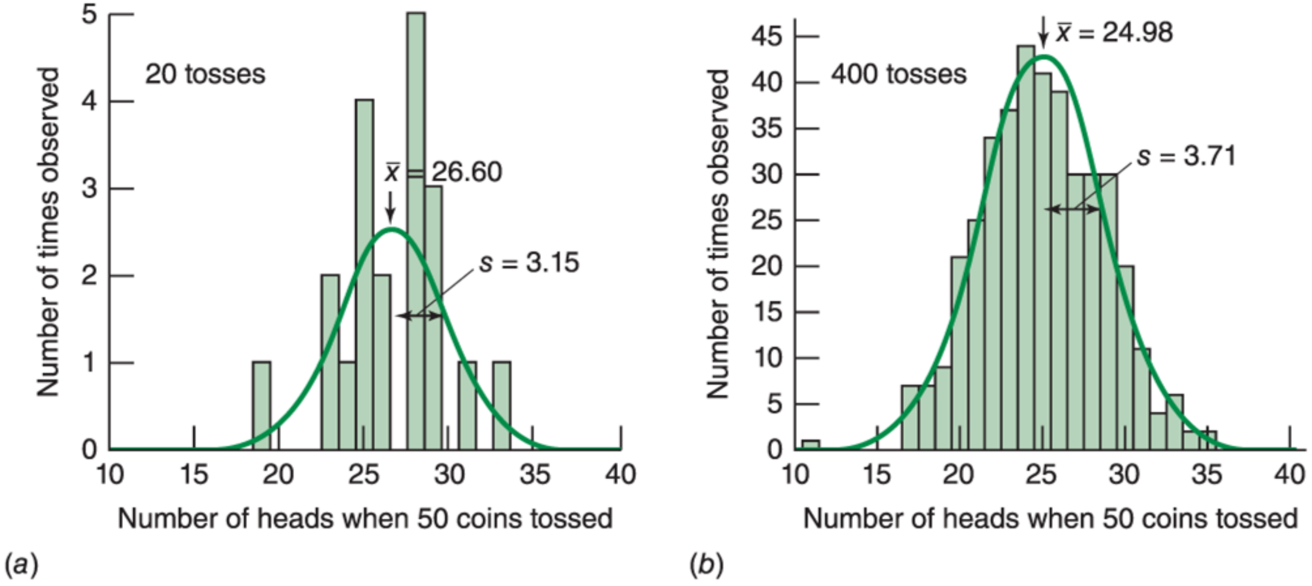
The area to the left of 15 heads in Figure 4-1b is only 0.26% of the entire area under the curve.

If the class tosses the 50 coins 400 times, they would expect to see 15 or fewer heads only once (0.26% of 400 = 1.04).

Example: Area Under a Gaussian Curve (3 of 3)

Test Yourself: If 50 coins are tossed 400 times, how many times would 32 or more heads be expected? How many times are observed in Figure 4-1b?

Figure 4-1

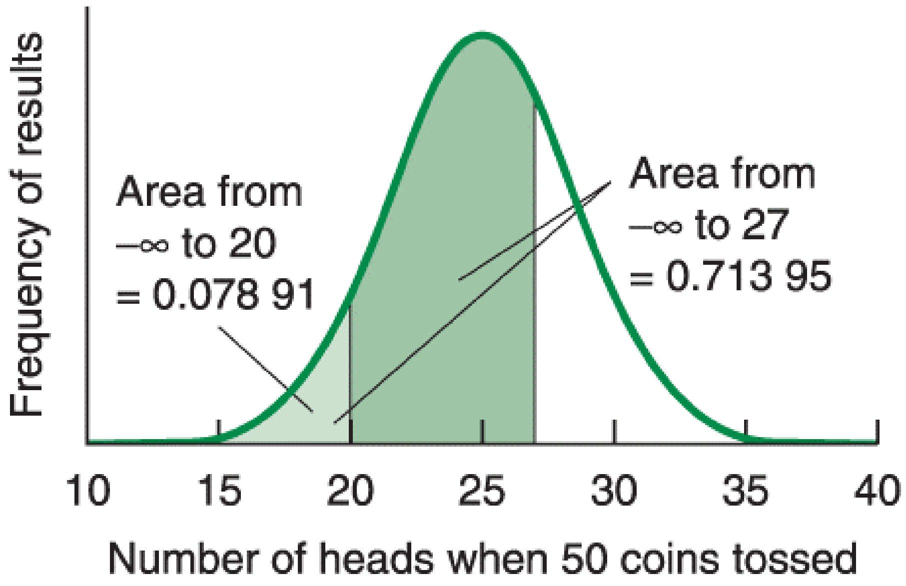


Harris/Lucy, *Quantitative Chemical Analysis*, 10e, © 2020 W. H. Freeman and Company

Example: Using a Spreadsheet to Find Area Under a Gaussian Curve (1 of 5)

For 400 tosses of 50 coins, how many tosses are expected to have between 20 and 27 heads?

Figure 4-4



Example: Using a Spreadsheet to Find Area Under a Gaussian Curve (2 of 5)

Solution: *We need to find the fraction of the area of the Gaussian curve between $x = 20$ and $x = 27$ heads and then multiply this fraction by 400 tosses.* The function NORM.DIST in Excel gives the area under the curve from $-\infty$ to a chosen value of x. We will find the area from $-\infty$ to 27 heads, which is all the shaded area to the left of 27 heads in Figure 4-4. Then we will find the area from $-\infty$ to 20 heads, which is the shaded area to the left of 20 heads. The difference between the two is the area from 20 to 27 heads:

$$\text{Area from 20 to 27} = (\text{area from } -\infty \text{ to } 27) - (\text{area from } -\infty \text{ to } 20)$$

Example: Using a Spreadsheet to Find Area Under a Gaussian Curve (3 of 5)

Solution: In a spreadsheet, enter the mean $\mu = 25.00$ in cell A2 and the standard deviation $\sigma = 3.54$ in cell B2. To find the area under the Gaussian curve from $-\infty$ to 27, select cell C4, then go to Formulas and Insert Function. Select Statistical functions and double click NORM.DIST. A window appears asking for four values that will be used by NORM.DIST.

	A	B	C
1	Mean =	Std dev =	
2	25.00	3.54	
3			
4	Area from –∞ to 27 =		0.7140
5	Area from –∞ to 20 =		0.0789
6	Area from 20 to 27 =		0.6350
7			
8	C4 = NORM.DIST(27,A2,B2,TRUE)		
9	C5 = NORM.DIST(20,A2,B2,TRUE)		
10	C6 = C4-C5		

Example: Using a Spreadsheet to Find Area Under a Gaussian Curve (4 of 5)

Solution: The values provided to NORM.DIST(*x*,mean,standard_dev,cumulative) are called *arguments* of the function.

The first argument is *x*, which is 27. The second argument is the mean. You can enter 25 for the mean or enter A2, which is the cell containing 25. We use dollar signs in A2 so that we can move the formula to other cells and still refer to cell A2.

The third argument is the standard deviation, for which we enter B2. The last argument is "cumulative." When cumulative is TRUE, NORM.DIST gives the area under the Gaussian curve. When cumulative is FALSE, NORM.DIST gives the ordinate (the *y*-value) of the Gaussian curve. We want area, so enter TRUE. The formula "= NORM.DIST(27,A2,B2,TRUE)" in cell C4 returns 0.714 0 for the area under the curve from $-\infty$ to 27. To get the area from $-\infty$ to 20, enter "= NORM.DIST(20,A2,B2,TRUE)" in cell C5. The value returned is 0.078 9. Then subtract the areas (C6 = C4 − C5) to obtain 0.635 0, which is the area from 20 to 27 heads. We expect 63.50% of 400 tosses (= 254 tosses) to have 20 to 27 heads.

Example: Using a Spreadsheet to Find Area Under a Gaussian Curve (5 of 5)

Test Yourself: How many tosses out of 400 tosses of 50 coins are expected to have 23 to 28 heads?

Standard Deviation and Probability (2 of 2)

The sum of the probabilities of all measurements must be unity.

- The area under the whole curve from $z = -\infty$ to $+\infty$ adds up to 1.

The standard deviation measures the width of the Gaussian curve.

- The larger σ, the broader the curve.

For any Gaussian curve:

Range	Percentage of measurements
$\mu \pm 1\sigma$	68.3
$\mu \pm 2\sigma$	95.5
$\mu \pm 3\sigma$	99.7

Standard Deviation of the Mean

The more times a quantity is measured, the more confident you can be that the mean is close to the population mean.

$$\boldsymbol{u_x = \frac{s_x}{\sqrt{n}}}$$

Standard deviation of the mean → u_x

Standard deviation → s_x

Number of measurements → n

- s_x approaches a *constant* value as $n \to \infty$.
- u_x approaches 0 as $n \to \infty$.

Section 4-2
Comparison of Standard Deviations with the *F* Test

Statistical Tests (1 of 2)

Two sets of measurements (on the same quantity) will generally differ in $\bar{x}$ and s. Statistical tools determine the probability of a conclusion.

- Accept conclusions with a high probability of being correct.
- Reject conclusions with a high probability of being incorrect.

Null hypothesis: states that two sets of data are drawn from populations with the *same* properties

- Observed differences arise only from random variation in measurements.
- Reject the null hypothesis if there is <5% probability of observing experimental results from two populations with the same value.

Statistical Tests (2 of 2)

The **null hypothesis** is used with several tests.

“Two sets of data are drawn from populations with the same properties.”

- ***F*-test** compares standard deviation σ
- ***t* test** compares mean μ

Must complete F-test before t test

Note: Statistical tables (Tables 4-3 and 4-4) include degrees of freedom.

F-Test: Comparison of Standard Deviation

Are the *standard deviations* of two sets of measurements "statistically different"?

$$F_{\text{calculated}} = \frac{s_1^2}{s_1^2}$$

where $s_1 \geq s_2$

Put the larger standard deviation in the numerator so that $F \geq 1$.

If $F_{\text{calculated}} > F_{\text{table}}$, then reject the null hypothesis.

- There is <5% chance that the two data sets came from populations with the same population standard deviation.
- The difference is considered significant.

Degrees of freedom for *n* measurements are $n - 1$.

Table 4-2 Measurement of HCO_3^- in horse blood

- Unethical trainers inject $NaHCO_3$ into a horse prior to a race to neutralize lactic acid that accumulates during strenuous activity.
- To enforce the ban on this practice, HCO_3^- in horse blood is measured after each race.

Authorities need to certify a new instrument:

	Original instrument	Substitute instrument
Mean ($\bar{x}$, mM)	36.14	36.20
Standard deviation (s, mM)	0.28	0.47
Number of measurements (n)	10	4

Table 4-3 Critical values of $F = s_1^2 / s_2^2$ at 95% confidence level for two-tailed F test

Degrees of freedom for s_2	Degrees of freedom for s_1													
	2	3	4	5	6	7	8	9	10	12	15	20	30	∞
2	39.00	39.17	39.25	39.30	39.33	39.36	39.37	39.39	39.40	39.41	39.43	39.45	39.46	39.50
3	16.04	15.44	15.10	14.88	14.73	14.62	14.54	14.47	14.42	14.34	14.25	14.17	14.08	13.90
4	10.65	9.98	9.60	9.36	9.20	9.07	8.98	8.90	8.84	8.75	8.66	8.56	8.46	8.26
5	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62	6.52	6.43	6.33	6.23	6.02
6	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	5.46	5.37	5.27	5.17	5.07	4.85
7	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82	4.76	4.67	4.57	4.47	4.36	4.14
8	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.30	4.20	4.10	4.00	3.89	3.67
9	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03	3.96	3.87	3.77	3.67	3.56	3.33
10	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	3.72	3.62	3.52	3.42	3.31	3.08
11	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.59	3.53	3.43	3.33	3.23	3.12	2.88
12	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44	3.37	3.28	3.18	3.07	2.96	2.72
13	4.97	4.35	4.00	3.77	3.60	3.48	3.39	3.31	3.25	3.15	3.05	2.95	2.84	2.60
14	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.21	3.15	3.05	2.95	2.84	2.73	2.49

Table 4-3 Critical values of $F = s_1^2 / s_2^2$ at 95% confidence level for two-tailed F test

Degrees of freedom for s_2	Degrees of freedom for s_1													
	2	3	4	5	6	7	8	9	10	12	15	20	30	∞
15	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12	3.06	2.96	2.86	2.76	2.64	2.40
16	4.69	4.08	3.73	3.50	3.34	3.22	3.12	3.05	2.99	2.89	2.79	2.68	2.57	2.32
17	4.62	4.01	3.66	3.44	3.28	3.16	3.06	2.98	2.92	2.82	2.72	2.62	2.50	2.25
18	4.56	3.95	3.61	3.38	3.22	3.10	3.01	2.93	2.87	2.77	2.67	2.56	2.44	2.19
19	4.51	3.90	3.56	3.33	3.17	3.05	2.96	2.88	2.82	2.72	2.62	2.51	2.39	2.13
20	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84	2.77	2.68	2.57	2.46	2.35	2.09
30	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57	2.51	2.41	2.31	2.20	2.07	1.79
∞	3.69	3.12	2.79	2.57	2.41	2.29	2.19	2.11	2.05	1.94	1.83	1.71	1.57	1.00

Critical values for a two-tailed test of the null hypothesis that $\sigma_1 = \sigma_2$. Tails are explained in Figure 4-9. There is a 5% probability of observing F at the tabulated value if the two sets of data come from populations with the same population standard deviation.
You can compute F for a chosen confidence with the Excel function F.INV.RT(probability,degree_freedom1,degree_freedom2). The statement "=F.INV.RT(0.025,7 ,6)" reproduces the value F = 5.70 in this table. The statement "F.INV.RT(0.05,7,6)" gives F = 4.21 for 90% confidence for a two-tailed test, which is also the value of F for a one-tailed test at 95% confidence.

Example: Is the Standard Deviation from the Substitute Instrument "Significantly Different" from That of the Original Instrument? (1 of 3)

In Table 4-2, the standard deviation from the new instrument is $s_1 = 0.47$ ($n_1 = 4$ measurements) and the standard deviation from the original instrument is $s_2 = 0.28$ ($n_2 = 10$).

	Original instrument	**Substitute instrument**
Mean ($\bar{x}$, mM)	36.14	36.20
Standard deviation (s, mM)	0.28	0.47
Number of measurements (n)	10	4

Example: Is the Standard Deviation from the Substitute Instrument "Significantly" Different from That of the Original Instrument? (1 of 2)

Solution: Compute $F_{\text{calculated}}$ with Equation 4-6:

$$F_{\text{calculated}} = \frac{s_1^2}{s_2^2} = \frac{(0.47)^2}{(0.28)^2} = 2.8_2$$

In Table 4-3, we find $F_{\text{table}} = 5.08$ in the column with 3 degrees of freedom for s_1 (degrees of freedom = $n - 1$) and the row with 9 degrees of freedom for s_2. *Because* $F_{\text{calculated}} = 2.8_2 < F_{\text{table}} = 5.08$, *we accept the null hypothesis and* conclude that s_1 and s_2 are *not significantly different*.

Example: Is the Standard Deviation from the Substitute Instrument “Significantly” Different from That of the Original Instrument? (2 of 2)

Test Yourself: If there had been $n = 21$ replications in both data sets, would the difference in standard deviations be significant?

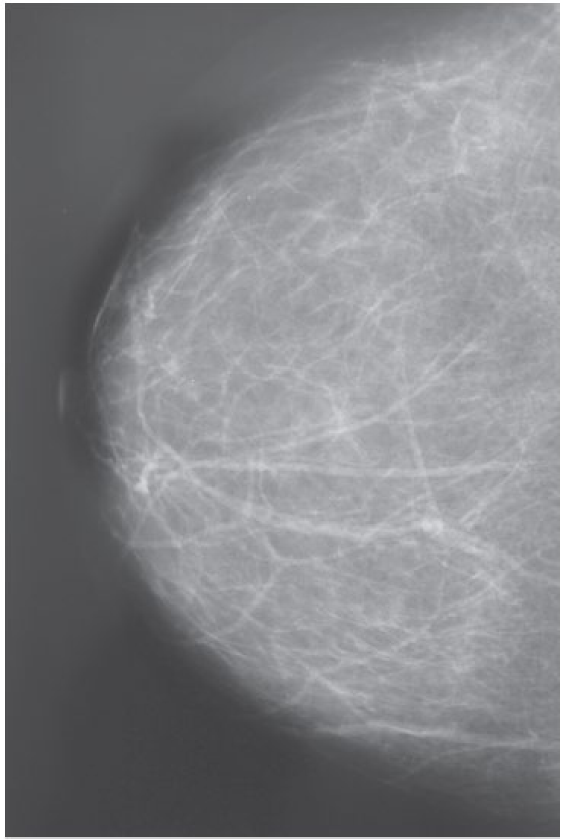
Box 4-1 Choosing the Null Hypothesis in Epidemiology

For a drug's approval, the null hypothesis is that *the treatment does not cause cancer*.

- Similar to the U.S. legal system, the null hypothesis puts the burden of proof on the prosecution: "innocent until proven guilty."
- The null hypothesis (stated above) is assumed to be true. Unless strong evidence is found proving otherwise, the FDA will continue to believe it is true.

Epidemiology Study Conducted at USC

- Studied the relationship between menopausal estrogen-progestin hormone therapy and breast cancer
- Study concluded there was a 7.6% increase in breast cancer risk per year of estrogen-progestin hormone therapy



allOver images/Alamy

Section 4-3
Confidence Intervals

Calculating Confidence Intervals

Student's *t*: used to compare results from different experiments

- Evaluate the probability that an observed experimental result agrees with a "known" value.

***t* is taken from Table 4-4 for desired confidence interval.**

$$\textbf{Confidence interval} = \bar{x} \pm \frac{ts}{\sqrt{n}}$$

If we were to repeat *n* measurements many times and compute $\bar{x}$ and *s*, the 95% confidence interval would include the true population mean (whose value we do not know) in 95% of the sets of *n* measurements.

We are 95% confident that the true mean lies within the confidence interval.

Table 4-4 Values of Student's *t* (1 of 2)

Degrees of freedom	Confidence level (%)						
	50	90	95	98	99	99.5	99.9
1	1.000	6.314	12.706	31.821	63.656	127.321	636.578
2	0.816	2.920	4.303	6.965	9.925	14.089	31.598
3	0.765	2.353	3.182	4.541	5.841	7.453	12.924
4	0.741	2.132	2.776	3.747	4.604	5.598	8.610
5	0.727	2.015	2.571	3.365	4.032	4.773	6.869
6	0.718	1.943	2.447	3.143	3.707	4.317	5.959
7	0.711	1.895	2.365	2.998	3.500	4.029	5.408
8	0.706	1.860	2.306	2.896	3.355	3.832	5.041
9	0.703	1.833	2.262	2.821	3.250	3.690	4.781
10	0.700	1.812	2.228	2.764	3.169	3.581	4.587
15	0.691	1.753	2.131	2.602	2.947	3.252	4.073
20	0.687	1.725	2.086	2.528	2.845	3.153	3.850
25	0.684	1.708	2.060	2.485	2.787	3.078	3.725
30	0.683	1.697	2.042	2.457	2.750	3.030	3.646

Table 4-4 Values of Student's *t* (2 of 2)

Degrees of freedom	Confidence level (%)						
	50	90	95	98	99	99.5	99.9
40	0.681	1.684	2.021	2.423	2.704	2.971	3.551
60	0.679	1.671	2.000	2.390	2.660	2.915	3.460
120	0.677	1.658	1.980	2.358	2.617	2.860	3.373
∞	0.674	1.645	1.960	2.326	2.576	2.807	3.291

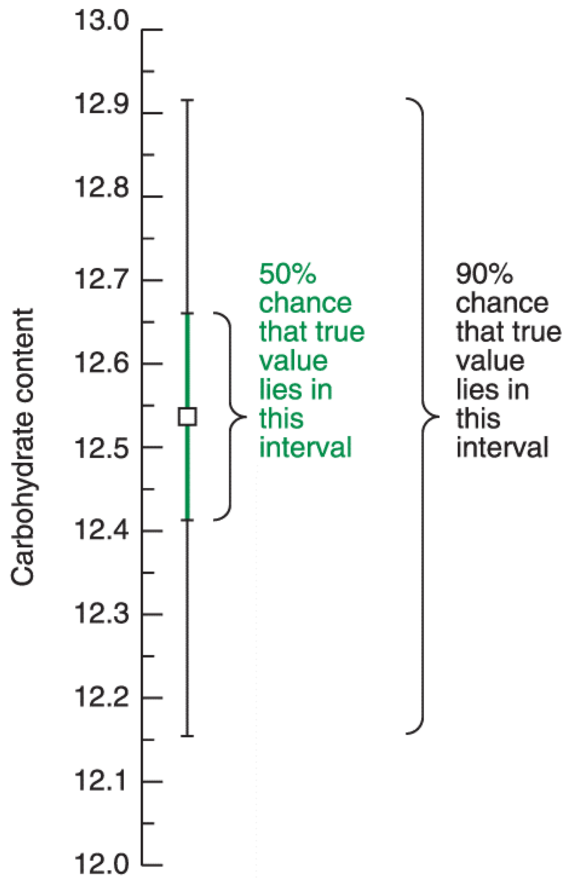
In calculating confidence intervals, σ may be substituted for s in Equation 4-7 if you have a great deal of experience with a particular method and have therefore determined its "true" population standard deviation. If σ is used instead of s, the value of t to use in Equation 4-7 comes from the bottom row of this table.

Values of t in this table apply to two-tailed tests illustrated in Figure 4-9a. The 95% confidence level specifies the regions containing 2.5% of the area in each wing of the curve. For a one-tailed test, we use values of t listed for 90% confidence. Each wing outside of t for 90% confidence contains 5% of the area of the curve.

Find t with the Excel function T.INV.2T. For 12 degrees of freedom and 95% confidence, the function T.INV.2T (0.05,12) gives t = 2.179. Many programmable calculators can provide t. Search "critical t [your calculator model]". Verify that the instructions give t = 2.179 for 12 degrees of freedom and 95% confidence.

Example: Calculating Confidence Intervals (1 of 4)

The carbohydrate content of a glycoprotein (a protein with sugars attached to it) is found to be 12.6, 11.9, 13.0, 12.7, and 12.5 wt% (g carbohydrate/100 g glycoprotein) in replicate analyses. Find the 50% and 90% confidence intervals for the carbohydrate content.



Harris/Lucy, *Quantitative Chemical Analysis*, 10e, © 2020 W. H. Freeman and Company

Example: Calculating Confidence Intervals (2 of 4)

Solution: First calculate $\bar{x}$ $(= 12.5_4)$ and (s = 0.4_0) for the five measurements. For the 50% confidence interval, look up t in Table 4-4 under 50 and across from *four* degrees of freedom (degrees of freedom = n − 1). The value of t is 0.741, so the 50% confidence interval is:

$$50\% \text{ confidence interval} = \bar{x} + \frac{ts}{\sqrt{n}} = 12.5_4 \pm \frac{(0.741)(0.4_0)}{\sqrt{5}} = 12.5_4 \pm 0.1_3 \text{ wt\%}$$

Example: Calculating Confidence Intervals (3 of 4)

Solution: The 90% confidence interval is:

$$90\%\ \text{confidence interval} = \bar{x} + \frac{ts}{\sqrt{n}} = 12.5_4 \pm \frac{(2.132)(0.4_0)}{\sqrt{5}} = 12.5_4 \pm 0.3_8\ \text{wt\%}$$

If we repeat sets of five measurements many times, half of the 50% confidence intervals are expected to include the true mean, μ. Nine-tenths of the 90% confidence intervals are expected to include the true mean, μ.

Example: Calculating Confidence Intervals (4 of 4)

Test Yourself: Carbohydrate measured on one more sample was 12.3 wt%. Using six results, find the 90% confidence interval.

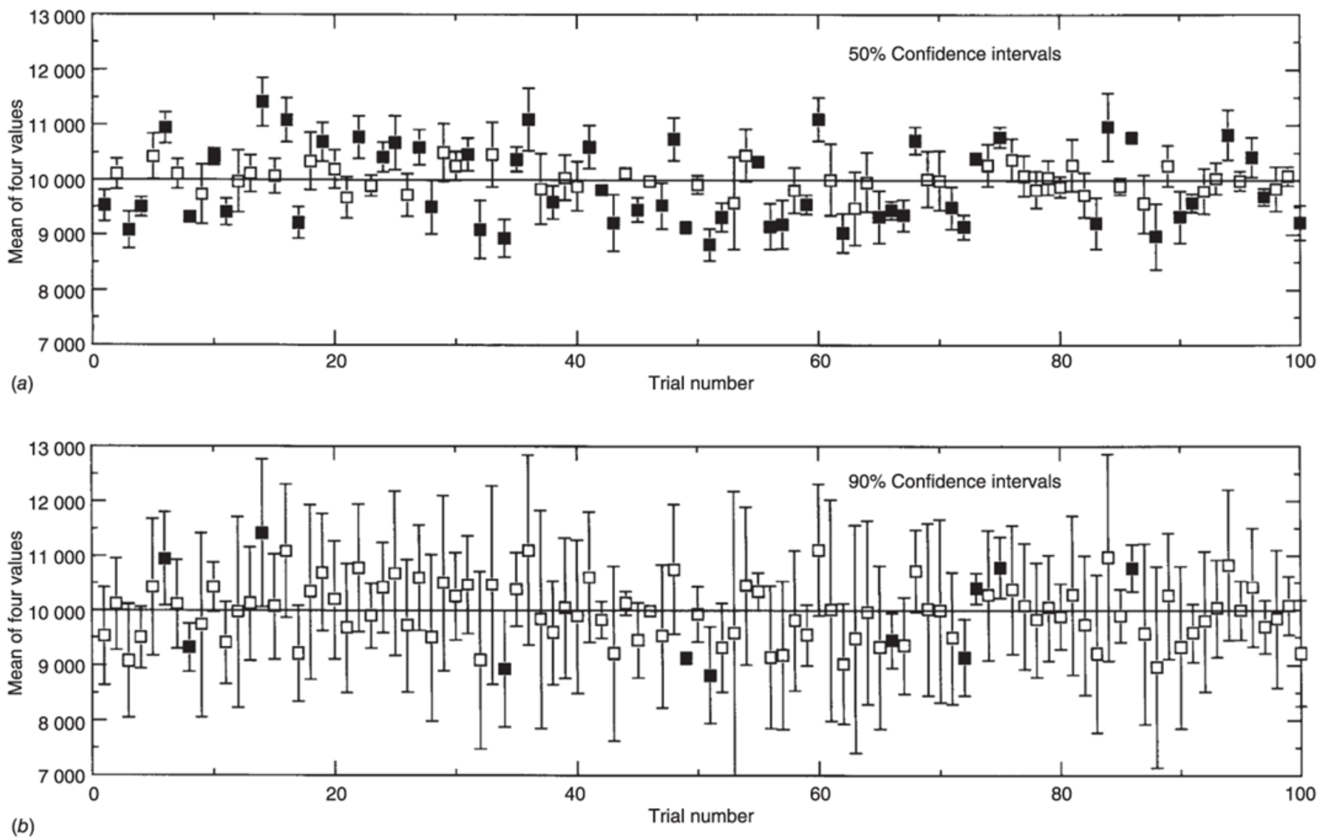
12.6, 11.9, 13.0, 12.7, 12.5 and **12.3** wt% (g carbohydrate/100 g glycoprotein)

50% and 90% Confidence Intervals for the Same Set of Random Data

- Each point represents an average $\overline{x}$ ($n = 4$)
- Error bars represent calculated confidence interval $\overline{x} \pm \frac{ts}{\sqrt{n}}$
- Population mean μ = 10 000

Filled squares: confidence interval does *not* include the true mean μ = 10 000

Figure 4-5



Harris/Lucy, *Quantitative Chemical Analysis*, 10e, © 2020 W. H. Freeman and Company

Always State What Kind of Uncertainty You Are Reporting

You can report uncertainty as

Standard deviation

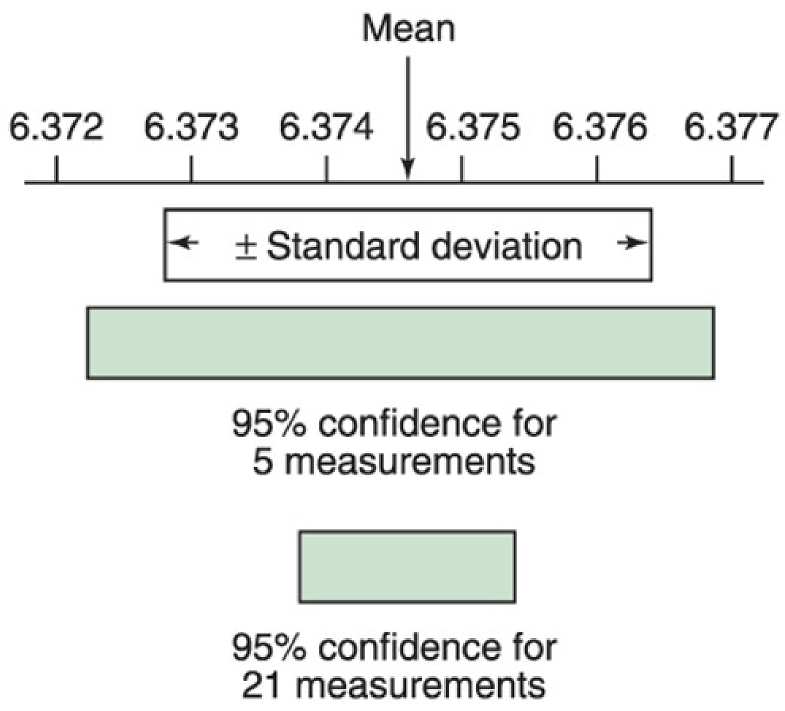
$$\bar{x} \pm s$$

OR

Confidence interval

$$\bar{x} \pm \frac{ts}{\sqrt{n}}$$

Reduce uncertainty by making more measurements.



Harris/Lucy, *Quantitative Chemical Analysis*, 10e, © 2020 W. H. Freeman and Company

Spreadsheets for Confidence Intervals

Finding confidence intervals with Excel

- Count the number of data points:

  =COUNT(A4:A10)

- Find Student's *t*:

  =T.INV.2T(1-C7,C6)

  % probability

  Degrees of freedom

	A	B	C	D	E	F
1	Confidence interval					
2						
3	Data	mean =	6.3746		=AVERAGE(A4:A10)	
4	6.375	stdev =	0.0018		=STDEV.S(A4:A10)	
5	6.372	n =	5		=COUNT(A4:A10)	
6	6.374	degrees of freedom =	4		=C5-1	
7	6.377	confidence level =	0.95		⇐ Enter	
8	6.375	Student's t =	2.776		=T.INV.2T(1-C7,C6)	
9		confidence interval =	0.0023		=C8*C4/SQRT(C5)	
10						