chapter 9

# Nuclear Counting Statistics

All measurements are subject to measurement error. This includes physical measurements, such as radiation counting measurements used in nuclear medicine procedures, as well as in biologic and clinical studies, such as evaluation of the effectiveness of an imaging technique. In this chapter, we discuss the type of errors that occur, how they are analyzed, and how, in some cases, they can be minimized.

## A. TYPES OF MEASUREMENT ERROR

Measurement errors are of three general types: blunders, systematic errors, and random errors.

*Blunders* are errors that are adequately described by their name. Usually they produce grossly inaccurate results and their occurrence is easily detected. Examples in radiation measurements include the use of incorrect instrument settings, incorrect labeling of sample containers, and injecting the wrong radiopharmaceutical into the patient. When a single value in the data seems to be grossly out of line with others in an experiment, statistical tests are available to determine whether the suspect value may be discarded (see Section E.3). Apart from this there is no way to "analyze" errors of this type, only to avoid them by careful work.

*Systematic errors* produce results that differ consistently from the correct result by some fixed amount. The same result may be obtained in repeated measurements, but it is the wrong result. For example, length measurements with a warped ruler, or activity measurements with a radiation detector that was miscalibrated or had some other persistent malfunction, could contain systematic errors. Observer bias in the subjective interpretation of data (e.g., scan reading) is another example of systematic error, as is the use for a clinical study of two population groups having underlying differences in some important characteristic, such as different average ages. Measurement results having systematic errors are said to be *inaccurate.*

It is not always easy to detect the presence of systematic error. Measurement results affected by systematic error may be very repeatable and not too different from the expected results, which may lead to a mistaken sense of confidence. One way to detect systematic error in physical measurements is by the use of measurement *standards,* which are known from previous measurements with a properly operating system to give a certain measurement result. For example, radionuclide standards, containing a known quantity of radioactivity, are used in various quality assurance procedures to test for systematic error in radiation counting systems. Some of these procedures are described in Chapter 11, Section D.

*Random errors* are variations in results from one measurement to the next, arising from physical limitations of the measurement system or from actual random variations of the measured quantity itself. For example, length measurements with an ordinary ruler are subject to random error because of inexact repositioning of the ruler and limitations of the human eye. In clinical or animal studies, random error may arise from differences between individual subjects, for example, in uptake of a radiopharmaceutical. Random error *always* is present in radiation counting measurements because the quantity that is being measured—namely, the rate of emission from the radiation source—is itself a randomly varying quantity.

Random error affects measurement *reproducibility* and thus the ability to detect real differences in measured data. Measurements that are very reproducible—in that nearly the same result is obtained in repeated measurements—are said to be *precise.* It is possible to minimize random error by using careful measurement technique, refined instrumentation, and so forth; however, it is impossible to eliminate it completely. There is always some limit to the precision of a measurement or measurement system. The amount of random error present sometimes is called the *uncertainty* in the measurement.

It also is possible for a measurement to be precise (small random error) but inaccurate (large systematic error), or vice versa. For example, length measurements with a warped ruler may be very reproducible (precise); nevertheless, they still are inaccurate. On the other hand, radiation counting measurements may be imprecise (because of inevitable variations in radiation emission rates) but still they can be accurate, at least in an average sense.

Because random errors always are present in radiation counting and other measured data, it is necessary to be able to analyze them and to obtain estimates of their magnitude. This is done using methods of statistical analysis. (For this reason, they are also sometimes called *statistical errors.*) The remainder of this chapter describes these methods of analysis. The discussion focuses on applications involving nuclear radiation-counting measurements; however, some of the methods to be described also are applicable to a wider class of experimental data as discussed in Section E.

## B. NUCLEAR COUNTING STATISTICS

### 1. The Poisson Distribution

Suppose that a long-lived radioactive sample is counted repeatedly under supposedly identical conditions with a properly operating counting system. Because the disintegration rate of the radioactive sample undergoes random variations from one moment to the next, the numbers of counts recorded in successive measurements ($N_1$, $N_2$, $N_3$, etc.) are not the same. Given that different results are obtained from one measurement to the next, one might question if a "true value" for the measurement actually exists. One possible solution is to make a large number of measurements and use the average $\bar{N}$ as an estimate for the "true value."

$$\text{True Value} \approx \bar{N} \tag{9-1}$$

$$\begin{aligned}\bar{N} &= (N_1 + N_2 + \cdots + N_n)/n \\ &= \frac{1}{n}\sum_{i=1}^{n} N_i\end{aligned} \tag{9-2}$$

where $n$ is the number of measurements taken. The notation $\sum$ indicates a sum that is taken over the indicated values of the parameter with the subscript $i$.

Unfortunately, multiple measurements are impractical in routine practice, and one often must be satisfied with only one measurement. The question then is, how good is the result of a single measurement as an estimate of the true value; that is, what is the uncertainty in this result? The answer to this depends on the *frequency distribution* of the measurement results. Figure 9-1 shows a typical frequency distribution curve for radiation counting measurements. The solid dots show the different possible results (i.e., number of counts recorded) versus the probability of getting each result. The probability is peaked at a *mean value, m,* which is the true value for the measurement. Thus if a large number of measurements were made and their results averaged, one would obtain

$$\bar{N} \approx m \tag{9-3}$$

The solid dots in Figure 9-1 are described mathematically by the *Poisson distribution.* For this distribution, the probability of getting a certain result $N$ when the true value is $m$ is given by

$$P(N;m) = e^{-m} m^N / N! \tag{9-4}$$

where $e$ (= 2.718 …) is the base of natural logarithms and $N!$ (*N factorial*) is the product of all integers up to $N$ (i.e., $1 \times 2 \times 3 \times \cdots \times N$) (Note that, by definition, $0! = 1$). From Figure 9-1 it is apparent that the probability of getting the exact result $N = m$ is rather small; however, one could hope that the result would at least be "close to" $m$. Note that the Poisson distribution is defined only for non-negative integer values of $N$ (0, 1, 2, …).

The probability that a measurement result will be "close to" $m$ depends on the relative width, or dispersion, of the frequency distribution curve. This is related to a parameter called the *variance*, $\sigma^2$, of the distribution.

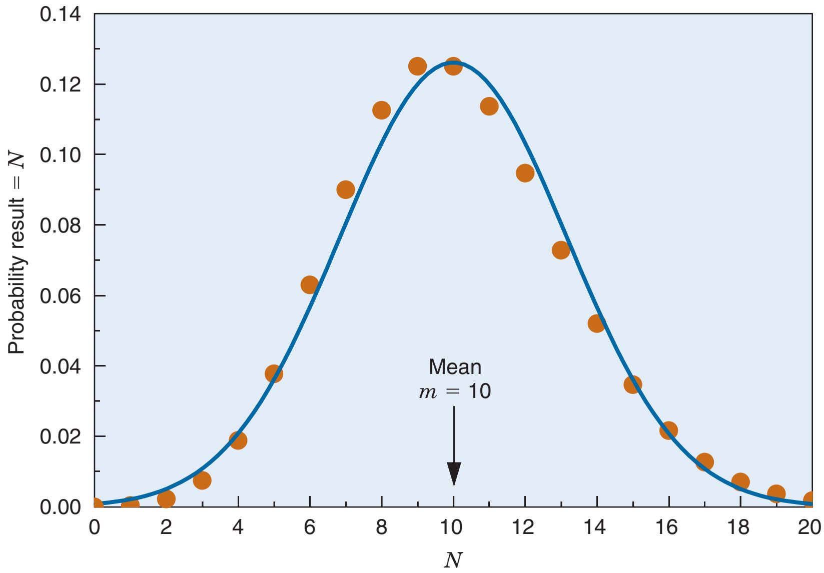


**FIGURE 9-1** Poisson (●) and Gaussian (—) distributions for mean, $m$, and variance, $\sigma^2 = 10$.

The variance is a number such that 68.3% (~2/3) of the measurement results fall within $\pm\sigma$ (i.e., square root of the variance) of the true value $m$. For the Poisson distribution, the variance is given by

$$\sigma^2 = m \tag{9-5}$$

Thus one expects to find approximately 2/3 of the counting measurement results within the range $\pm\sqrt{m}$ of the true value $m$.

Given only the result of a single measurement, $N$, one does not know the exact value of $m$ or of $\sigma$; however, one can reasonably assume that $N \approx m$, and thus that $\sigma \approx \sqrt{N}$. One can therefore say that if the result of the measurement is $N$, there is a 68.3% chance that the true value of the measurement is within the range $N \pm \sqrt{N}$. This is called the "68.3% *confidence interval*" for $m$; that is, one is 68.3% confident that $m$ is somewhere within the range $N \pm \sqrt{N}$.

The range $\pm\sqrt{N}$ is the uncertainty in $N$. The *percentage uncertainty* in $N$ is

$$\begin{aligned} V &= (\sqrt{N}/N) \times 100\% \\ &= 100\%/\sqrt{N} \end{aligned} \tag{9-6}$$

## EXAMPLE 9-1

Compare the percentage uncertainties in the measurements $N_1 = 100$ counts and $N_2 = 10{,}000$ counts.

***Answer***

For $N_1 = 100$ counts, $V_1 = 100\% / \sqrt{100} = 10\%$ (Equation 9-6). For $N_2 = 10{,}000$ counts, $V_2 = 100\% / \sqrt{10{,}000} = 1\%$. Thus the percentage uncertainty in 10,000 counts is only 1/10 the percentage uncertainty in 100 counts.

Equation 9-6 and Example 9-1 indicate that *large numbers of counts have smaller percentage uncertainties and are statistically more reliable than small numbers of counts.*

Other confidence intervals can be defined in terms of $\sigma$ or $\sqrt{N}$. They are summarized in Table 9-1. The 50% confidence interval ($0.675\sigma$) is sometimes called the *probable error* in $N$.

**TABLE 9-1**
**CONFIDENCE LEVELS IN RADIATION COUNTING MEASUREMENTS**

| Range | Confidence Level for *m* (True Value) (%) |
|---|---|
| $N \pm 0.675\sigma$ | 50 |
| $N \pm \sigma$ | 68.3 |
| $N \pm 1.64\sigma$ | 90 |
| $N \pm 2\sigma$ | 95 |
| $N \pm 3\sigma$ | 99.7 |

### 2. The Standard Deviation

The variance $\sigma^2$ is related to a statistical index called the *standard deviation* (*SD*). The standard deviation is a number that is calculated for a series of measurements. If $n$ counting measurements are made, with results $N_1, N_2, N_3, \ldots, N_n$, and a mean value $\bar{N}$ for those results is found, the standard deviation is

$$SD = \sqrt{\frac{\sum_{i=1}^{n}(N_i - \bar{N})^2}{(n-1)}} \qquad (9\text{-}7)$$

The standard deviation is a measure of the dispersion of measurement results about the mean and is in fact an estimate of σ, the square root of the variance. For radiation counting measurements, one therefore should obtain

$$SD \approx \sqrt{N} \qquad (9\text{-}8)$$

This can be used to test whether the random error observed in a series of counting measurements is consistent with that predicted from random variations in source decay rate, or if there are additional random errors present, such as from faulty instrument performance. This is discussed further in Section E.

### 3. The Gaussian Distribution

When the mean value $m$ is "large," the Poisson distribution can be approximated by the *Gaussian distribution* (also called the *normal distribution*). The equation describing the Gaussian distribution is

$$P(x; m, \sigma) = (1/\sqrt{2\pi\sigma^2})e^{-(x-m)^2/2\sigma^2} \qquad (9\text{-}9)$$

where $m$ and $\sigma^2$ are again the mean and variance. Equation 9-9 describes a symmetrical "bell-shaped" curve. As shown by Figure 9-1, the Gaussian distribution for $m$ = 10 and $\sigma = \sqrt{m}$ is very similar to the Poisson distribution for $m$ = 10. For $m \gtrsim 20$, the distributions are virtually indistinguishable. Two important differences are that the Poisson distribution is defined only for nonnegative integers, whereas the Gaussian distribution is defined for any value of $x$, and that for the Poisson distribution, the variance $\sigma^2$ is equal to the mean, $m$, whereas for the Gaussian distribution, it can have any value.

The Gaussian distribution with $\sigma^2 = m$ is a useful approximation for radiation counting measurements when the only random error present is that caused by random variations in source decay rate. When additional sources of random error are present (e.g., a random error or uncertainty of $\Delta N$ counts caused by variations in sample preparation technique, counting system variations, and so forth), the results are described by the Gaussian distribution with variance given by

$$\sigma^2 \approx m + (\Delta N)^2 \qquad (9\text{-}10)$$

The resulting Gaussian distribution curve would be wider than a Poisson curve with $\sigma^2 = m$. The confidence intervals given in Table 9-1 may be used for the Gaussian distribution with this modified value for the variance. For example, the 68.3% confidence interval for a measurement result $N$ would be $\pm\sqrt{N + \Delta N^2}$ (assuming $N \approx m$).

**EXAMPLE 9-2**

A 1-mL radioactive sample is pipetted into a test tube for counting. The precision of the pipette is specified as "±2%," and 5000 counts are recorded from the sample. What is the uncertainty in sample counts per mL?

***Answer***

The uncertainty in counts arising from pipetting precision is 2% × 5000 counts = 100 counts. Therefore,

$$\sigma^2 = 5000 + (100)^2 \approx 15{,}000,$$

and the uncertainty is $\sqrt{15{,}000} \approx 122$ counts. Compare this with the uncertainty of $\sqrt{5000} \approx 71$ counts that would be obtained without the pipetting uncertainty.

## C. PROPAGATION OF ERRORS

The preceding section described methods for estimating the random error or uncertainty in a single counting measurement; however, most nuclear medicine procedures involve multiple counting measurements, from which ratios, differences, and so on are used to compute the final result. In the following four sections we describe equations and methods that apply when a result is obtained from a set of counting measurements, $N_1, N_2, N_3 \ldots$. In some cases, we present first the general equation applicable for measurements of any type, $M_1$, $M_2$, $M_3$, …, having individual variances, $\sigma(M_1)^2$, $\sigma(M_2)^2$, $\sigma(M_3)^2$, … The general equations can be used to compute the uncertainty in the result for whatever $M$ might represent (e.g., a series of readings from a scale or a thermometer). We then apply these general equations to nuclear counting measurements. Note that in the following subsections it is assumed that random fluctuations in counting measurements arise *only* from random fluctuations in sample decay rate and that the individual measurements are *statistically independent* from one another. The latter condition would be violated if $N_1$ was in some way correlated with $N_2$, for example, if $N_1$ was the result for the first half of the counting period for the measurement of $N_2$.

## 1. Sums and Differences

For either sums or differences of a series of individual measurements, $M_1$, $M_2$, $M_3$, …, with individual variances, $\sigma(M_1)^2$, $\sigma(M_2)^2$, $\sigma(M_3)^2$, …, the general equation for the square root of the variance of the result is given by

$$\mathrm{SD}(M_1 \pm M_2 \pm M_3 \pm \cdots) \approx \sqrt{\mathrm{SD}(M_1)^2 + \mathrm{SD}(M_2)^2 + \mathrm{SD}(M_3)^2 + \cdots} \tag{9-11}$$

Thus, for a series of counting measurements with individual results $N_1$, $N_2$, $N_3$, … an estimate for the standard deviation of their sums or differences can be obtained using the standard deviations for individual results (Eq. 9-8) as follows:

$$\mathrm{SD}(N_1 \pm N_2 \pm N_3 \pm \cdots) \approx \sqrt{N_1 + N_2 + N_3 + \cdots} \tag{9-12}$$

and the percentage uncertainty is

$$V(N_1 \pm N_2 \pm N_3 \pm \cdots) = \frac{\sqrt{N_1 + N_2 + N_3 + \cdots}}{N_1 \pm N_2 \pm N_3 \cdots} \times 100\% \tag{9-13}$$

Note that these equations apply to mixed combinations of sums and differences.

## 2. Constant Multipliers

If a measurement $M$ having variance $\sigma_M{}^2$ is multiplied by a constant $k$, the general equation for the variance of the product is

$$\mathrm{SD}\,(kM) \approx k\,\mathrm{SD}_M \tag{9-14}$$

Substituting the appropriate quantities for counting measurements, with $M = N$

$$\mathrm{SD}\,(kN) \approx k\sqrt{N} \tag{9-15}$$

The percentage uncertainty $V$ in the product $kN$ is

$$\begin{aligned} V(kN) &= [\sigma\,(kN)/kN] \times 100\% \\ &= 100\%/\sqrt{N} \end{aligned} \tag{9-16}$$

which is the same result as Equation 9-6. Thus there is no statistical advantage gained or lost in multiplying the number of counts recorded by a constant. The percentage uncertainty still depends on the actual number of counts recorded.

## 3. Products and Ratios

The uncertainty in the product or ratio of a series of measurements $M_1$, $M_2$, $M_3$, … is most conveniently expressed in terms of the *percentage* uncertainties in the individual results, $V_1$, $V_2$, $V_3$, … The general equation is given by

$$V(M_1 \overset{\times}{\div} M_2 \overset{\times}{\div} M_3 \overset{\times}{\div} \cdots) = \sqrt{V_1^2 + V_2^2 + V_3^2 + \cdots} \tag{9-17}$$

For counting measurements, this becomes

$$V(N_1 \overset{\times}{\div} N_2 \overset{\times}{\div} N_3 \overset{\times}{\div} \cdots) = \sqrt{\frac{1}{N_1} + \frac{1}{N_2} + \frac{1}{N_3} + \cdots} \times 100\% \tag{9-18}$$

Again, this expression applies to mixed combinations of products and ratios.

## 4. More Complicated Combinations

Many nuclear medicine procedures, such as thyroid uptakes and blood volume determinations, use equations of the following general form.

$$Y = \frac{k(N_1 - N_2)}{(N_3 - N_4)} \tag{9-19}$$

The uncertainty in $Y$ is expressed most conveniently in terms of its percentage uncertainty. Using the rules given previously, one can show that

$$V_Y = \sqrt{\frac{(N_1 + N_2)}{(N_1 - N_2)^2} + \frac{(N_3 + N_4)}{(N_3 - N_4)^2}} \times 100\% \tag{9-20}$$

**EXAMPLE 9-3**

A patient is injected with a radionuclide. At some later time a blood sample is withdrawn for counting in a well counter and $N_p = 1200$ counts are recorded. A blood sample withdrawn prior to injection gives a blood background of $N_{pb} = 400$ counts. A standard prepared from the injection preparation records $N_s = 2000$ counts, and a "blank" sample records an instrument background of $N_b = 200$ counts. Calculate the ratio of net patient sample counts to net standard counts, and the uncertainty in this ratio.

***Answer***

The ratio is

$$\begin{aligned} Y &= (N_p - N_{pb})/(N_s - N_b) \\ &= (1200-400)/(2000-200) \\ &= 800/1800 = 0.44 \end{aligned}$$

The percentage uncertainty in the ratio is (Equation 9-20)

$$\begin{aligned} V_Y &= \sqrt{\frac{(1200+400)}{(1200-400)^2} + \frac{(2000+200)}{(2000-200)^2}} \times 100\% \\ &= 5.6\% \end{aligned}$$

The uncertainty in $Y$ is $5.6\% \times 0.44 \approx 0.02$; thus the ratio and its uncertainty are $Y = 0.44 \pm 0.02$.

## D. APPLICATIONS OF STATISTICAL ANALYSIS

### 1. Effects of Averaging

If $n$ counting measurements are used to compute an average result, the average $\bar{N}$ is a more reliable estimate of the true value than any one of the individual measurements. The uncertainty in $\bar{N}$, $\sigma_{\bar{N}}$, can be obtained by combining the rules for sums (Equation 9-11) and constant multipliers (Equation 9-14).

$$\sigma_{\bar{N}} = \sqrt{\bar{N}/n} \tag{9-21}$$

The uncertainty in $\bar{N}$ as an estimator of $m$ therefore is smaller than the uncertainty in a single measurement by a factor $1/\sqrt{n}$.

### 2. Counting Rates

If $N$ counts are recorded during a measuring time $t$, the average counting rate during that interval is $R = N/t$. Using Equation 9-15, the uncertainty in the counting rate $R$ is

$$\begin{aligned} \sigma_R &= (1/t)\sqrt{N} \\ &= \sqrt{N/t^2} \\ &= \sqrt{R/t} \end{aligned} \tag{9-22}$$

The percentage uncertainty in R is

$$\begin{aligned} V_R &= (\sigma_R/R) \times 100\% \\ &= 100\%/\sqrt{Rt} \end{aligned} \tag{9-23}$$

**EXAMPLE 9-4**

In a 2-min counting measurement, 4900 counts are recorded. What is the average counting rate $R$(cpm) and its uncertainty?

***Answer***

$$R = 4900/2 = 2450 \text{ cpm}$$

From Equation 9-22

$$\sigma_R = \sqrt{2450/2} = 35 \text{ cpm}$$

and from Equation 9-23

$$V_R = 100\%/\sqrt{2450 \times 2} \approx 1.4\%$$

Note from Equations 9-22 and 9-23 that *longer counting times produce smaller uncertainties in estimated counting rates.*

### 3. Significance of Differences Between Counting Measurements

Suppose two samples are counted and that counts $N_1$ and $N_2$ are recorded. The difference $(N_1 - N_2)$ may be due to an actual difference between sample activities or may be simply the result of random variations in counting rates. There is no way to state with absolute certainty that a given difference is or is not caused by random error; however, one can assess the "statistical significance" of the difference by comparing it with the expected random error. In general, differences of less than 2 σ [i.e., $(N_1 - N_2) < 2\sqrt{N_1 + N_2}$] are considered to be of marginal or no statistical significance because there is at least a 5% chance that such a difference is simply caused by random error (see Table 9-1). Differences greater than 3σ are considered significant (<1% chance caused by random error), whereas differences between 2σ and 3σ are in the questionable category, perhaps deserving repeat measurement or longer measuring times to determine their significance.

If two counting rates $R_1$ and $R_2$ are determined from measurements using counting times $t_1$ and $t_2$, respectively, the uncertainty in their difference $R_1 - R_2$, can be obtained by applying Equations 9-11 and 9-22.

$$\sigma(R_1 - R_2) = \sqrt{R_1/t_1 + R_2/t_2} \qquad (9\text{-}24)$$

Comparison of the observed difference to the expected random error difference can again be used to assess statistical significance, as described in Section B.

## 4. Effects of Background

All nuclear counting instruments have background counting rates, caused by electronic noise, detection of cosmic rays, natural radioactivity in the detector itself (e.g., $^{40}$K), and so forth. If the background counting rate, measured with no sample present, is $R_b$ and the gross counting rate with the sample is $R_g$, then the net sample counting rate is

$$R_s = R_g - R_b \qquad (9\text{-}25)$$

The uncertainty in $R_s$ is (from Equation 9-24)

$$\sigma_{R_s} = \sqrt{R_g/t_g + R_b/t_b} \qquad (9\text{-}26)$$

The percentage uncertainty in $R_s$ is

$$V_{R_s} = \left[\sqrt{R_g/t_g + R_b/t_b}\,/(R_g - R_b)\right] \times 100\% \qquad (9\text{-}27)$$

If the same counting time $t$ is used for both sample and background counting,

$$\begin{aligned}\sigma_{R_s} &= \sqrt{R_g + R_b}\,/\sqrt{t} \\ &= \sqrt{R_s + 2R_b}\,/\sqrt{t}\end{aligned} \qquad (9\text{-}28)$$

### EXAMPLE 9-5

In 4-min counting measurements, gross sample counts are 6000 counts and background counts are 4000 counts. What are the net sample counting rate and its uncertainty?

***Answer***

$$R_g = 6000/4 = 1500 \text{ cpm}$$

$$R_b = 4000/4 = 1000 \text{ cpm}$$

$$R_s = 1500 - 1000 = 500 \text{ cpm}$$

From Equation 9-28

$$\begin{aligned}\sigma_{R_s} &= \sqrt{500 + (2 \times 1000)}\,/\sqrt{4} \\ &= \sqrt{2500}\,/\sqrt{4} = 50/2 \\ &= 25 \text{ cpm}\end{aligned}$$

Therefore $R_s = 500 \pm 25$ cpm ($\pm 5\%$). Compare this with the uncertainty in the gross counting rate $R_g$ (from Equation 9-22)

$$\sigma_{R_g} = \sqrt{1500/4} \approx 19 \text{ cpm } (\sim 1\%)$$

and to the uncertainty in $R_s$ that would be obtained if there were negligible background ($R_b \approx 0$),

$$\sigma_{R_s} = \sqrt{500/4} \approx 11 \text{ cpm } (\sim 2\%)$$

---

Example 9-5 illustrates two important points:

1. *High background counting rates are undesirable because they increase uncertainties in net sample counting rates.*
2. *Small differences between relatively high counting rates can have relatively large uncertainties.*

## 5. Minimum Detectable Activity

The minimum detectable activity (MDA) of a radionuclide for a particular counting system and counting time $t$ is that activity that increases the counts recorded by an amount that is "statistically significant" in comparison with random variations in background counts that would be recorded during the same measuring time. In this instance, statistically significant means a counting rate increase of $3\sigma$. Therefore, from Equation 9-22, the counting rate for the MDA is

$$\text{MDA} = 3\sqrt{R_b/t} \qquad (9\text{-}29)$$

### EXAMPLE 9-6

A standard NaI(Tl) well counter has a background counting rate (full spectrum) of approximately 200 cpm. The sensitivity of the well counter for $^{131}$I is approximately 29 cpm/Bq (see Table 12-2). What is the MDA for $^{131}$I, using 4-min counting measurements?

***Answer***

The MDA is that amount of $^{131}$I giving $3 \times \sqrt{200 \text{ cpm}/4} \approx 3 \times 7$ cpm = 21 cpm. Thus

$$\begin{aligned}\text{MDA} &= 21 \text{ cpm}/(29 \text{ cpm/Bq}) \\ &\approx 0.7 \text{ Bq (i.e., } <1 \text{ dps)}\end{aligned}$$

In traditional units (1 μCi = 37 kBq), the MDA is ~ 0.00002 μCi.

---

## 6. Comparing Counting Systems

In Section B.1 it was noted that larger numbers of counts have smaller percentage uncertainties. Thus in general it is desirable from a statistical point of view to use a counting system with maximum sensitivity (i.e., large detector, wide pulse-height analyzer window) so that a maximum number of counts is obtained in a given measuring time; however, such systems are also more sensitive to background radiation and give higher background counting rates as well, which, as shown by Example 9-5, tends to increase statistical uncertainties. The tradeoff between sensitivity and background may be analyzed as follows:

Suppose a counting system provides gross sample counts $G_1$, background counts $B_1$, and net sample counts $S_1 = G_1 - B_1$ and that a second system provides gross, background, and net counts $G_2$, $B_2$, and $S_2$ in the same counting time. One can compare the uncertainties in $S_1$ and $S_2$ to determine which system is statistically more reliable. The percentage uncertainty in $S_1$ is given by

$$\begin{aligned} V_1 &= \frac{\sqrt{G_1 + B_1}}{S_1} \times 100\% \\ &= \frac{\sqrt{S_1 + 2B_1}}{S_1} \times 100\% \end{aligned} \tag{9-30}$$

Corresponding equations apply to the second system. The ratio of the percentage uncertainties for the net sample counts obtained with two systems is therefore

$$\frac{V_1}{V_2} = \frac{S_2}{S_1} \times \frac{\sqrt{S_1 + 2B_1}}{\sqrt{S_2 + 2B_2}} \tag{9-31}$$

If $V_1/V_2 < 1$, then $V_1 < V_2$, in which case system 1 is the statistically preferred system. Conversely, if $V_1/V_2 > 1$, system 2 is preferred.

If background counts are relatively small ($B_1 \ll S_1$, $B_2 \ll S_2$), Equation 9-31 can be approximated by

$$\begin{aligned} \frac{V_1}{V_2} &\approx \frac{S_2\sqrt{S_1}}{S_1\sqrt{S_2}} \\ &\approx \sqrt{\frac{S_2}{S_1}} \end{aligned} \tag{9-32}$$

Thus when background levels are "small," only relative sensitivities are important. The system with the higher sensitivity gives the smaller uncertainty. Conversely, if background counts are large ($B_1 \gg S_1$, $B_2 \gg S_2$), Equation 9-31 is approximated by

$$\frac{V_1}{V_2} \approx \frac{S_2}{S_1}\sqrt{\frac{B_1}{B_2}} \tag{9-33}$$

Both sensitivity and background are important in this case. Note that Equations 9-31 through 9-33 also can be used with counting *rates* (cpm, cps) substituted for counts when equal counting times are used for all measurements.

### EXAMPLE 9-7

A sample is counted in a well counter using a "narrow" pulse-height analyzer window and net sample and background counts are $S_N = 500$ counts and $B_N = 200$ counts, respectively. The sample is counted with the same system but using a "wide" window and the net sample and background counts are $S_W = 800$ counts and $B_W = 400$ counts, respectively. Which window setting offers the statistical advantage?

***Answer***

Background counts are neither "very small" nor "very large" in comparison with net sample counts; thus Equation 9-31 must be used:

$$\begin{aligned} \frac{V_N}{V_W} &= \frac{800}{500} \times \frac{\sqrt{500 + (2 \times 200)}}{\sqrt{800 + (2 \times 400)}} \\ &= (8/5) \times \sqrt{9/16} = (8/5) \times (3/4) \\ &= 1.2 \end{aligned}$$

Thus $V_N/V_W > 1$ and the statistical advantage belongs to the wider window setting, in spite of its higher background counting rate.

## 7. Estimating Required Counting Times

Suppose it is desired to determine net sample counting rate $R_s$ to within a certain percentage uncertainty $V$. Suppose further that the approximate net sample and background counting rates are known to be $R_s'$ and $R_b'$, respectively (e.g., from quick preliminary measurements). If a counting time $t$ is to be used for both the sample and background counting measurements, then the time required to achieve the desired level of statistical reliability is given by

$$t = [(R_s' + 2R_b')/R_s'^2](100\%/V)^2 \tag{9-34}$$

**EXAMPLE 9-8**

Preliminary measurements in a sample counting procedure indicate gross and background counting rates of $R_g = 900$ cpm and $R_b = 100$ cpm, respectively. What counting time is required to determine net sample counting rate to within 5%?

***Answer***

$$R'_s = 900 - 100 = 800 \text{ cpm}$$

$$\begin{aligned} t &= \{[800 + (2 \times 100)]/800^2\} \times (100/5)^2 \\ &= (1000/800^2) \times (100/5)^2 \\ &= 0.625 \text{ min} \end{aligned}$$

This time is used for both sample and background counting. Therefore the total counting time required is 1.25 min.

## 8. Optimal Division of Counting Times

In the preceding section it was assumed that equal counting times were used for the sample and background measurements. This is not necessary; in fact, statistically advantageous results may be obtainable by using unequal times. The difference between two counting rates $R_1$ and $R_2$ is determined with the smallest statistical error if the total counting time $t = t_1 + t_2$ is divided according to

$$t_1 / t_2 = \sqrt{R'_1 / R'_2} \tag{9-35}$$

where $R'_1$ and $R'_2$ are counting rates estimated from preliminary measurements. Applying this to gross sample and background counting rate estimates, one obtains

$$t_g/t_b = \sqrt{R'_g/R'_b} \tag{9-36}$$

If ($R'_g \approx R'_b$), approximately equal counting times are preferred; however, if the background counting rate is small ($R'_b \ll R'_g$), it is better to devote most of the available time to counting the sample.

**EXAMPLE 9-9**

In Example 9-8, what is the optimal division of a 1.25-min total counting time and the resulting uncertainty in the net sample counting rate?

***Answer***

Applying Equation 9-36, with $R'_g = 900$ cpm and $R'_b = 100$ cpm,

$$t_g/t_b = \sqrt{900/100} = 3$$

$$t_g = 3t_b$$

$$t_g + t_b = 3t_b + t_b = 1.25 \text{ min}$$

$$t_b = 1.25 \text{ min}/4 \approx 0.3 \text{ min}$$

$$t_g \approx 1.25 - 0.3 = 0.95 \text{ min}$$

The percentage uncertainty in $R_s$ given by Equation 9-27 is

$$\begin{aligned} V_{R_s} &= \left[\sqrt{(R_g/t_g) + (R_b/t_b)} \times 100\%\right]/(R_g - R_b) \\ &= \left[\sqrt{(900/0.95) + (100/0.3)} \times 100\%\right]/800 \\ &\approx 4.5\% \end{aligned}$$

Thus a small statistical advantage (4.5% vs. 5%) is gained by using an optimal division rather than equal counting times in this example.

# E. STATISTICAL TESTS

In Section D.3, an example was given of a method for testing the statistical significance of the difference between two counting measurements. The test was based on the assumption of underlying Poisson distributions for the two individual measurements, with variances $\sigma^2 \approx N$. In this section we consider a few other tests for evaluating statistical parameters of sets of counting measurements. The discussion focuses on applications of these tests to nuclear counting data; however, as noted in the discussion, the tests also are applicable to other experimental data for which the underlying random variability is described by a Poisson or Gaussian distribution. More detailed discussions of statistical tests are found in the references and suggested readings at the end of this chapter.

## 1. The $\chi^2$ Test

The $\chi^2$ (*chi-square*) test is a means for testing whether random variations in a set of measurements are consistent with what would be expected for a Poisson distribution. This is a particularly useful test when a set of counting measurements is suspected to contain sources of random variation in addition to Poisson counting statistics, such as those resulting from faulty instrumentation or other random variability between samples, animals, patients, and measurement techniques. The test is performed as follows:

1. Obtain a series of counting measurements (at least 20 measurements is desirable).
2. Compute the mean

$$\bar{N} = \sum_{i=1}^{n} N_i / n \qquad (9\text{-}37)$$

and the quantity

$$\begin{aligned} \chi^2 &= \sum_{i=1}^{n} (N_i - \bar{N})^2 / \bar{N} \\ &= (n-1)SD^2/\bar{N} \end{aligned} \qquad (9\text{-}38)$$

where $SD$ = standard deviation (Equation 9-7). Many pocket calculators have programs for calculating standard deviations; thus the second form in Equation 9-38 may be more convenient to use.
3. Refer to a $\chi^2$ table or graph (Fig. 9-2). Locate the value corresponding to the number of measurements, $n$, on the horizontal axis.
4. Compare the computed value of $\chi^2$ to the most closely corresponding $P$-value curve.

$P$ is the probability that random variations observed in a series of $n$ measurements from a Poisson distribution would *equal or exceed* the calculated $\chi^2$ value. Conversely, $1 - P$ is the probability that smaller variations would be observed. A $P$ value of 0.5 (50%) would be "perfect." It indicates that the observed $\chi^2$ value is in the middle of the range expected for a Poisson distribution. (Note that this corresponds to $\chi^2 \approx n - 1$.)

A low $P$ value (<0.01) indicates that there is only a small probability that a Poisson distribution would give the $\chi^2$ value as large as the value observed and suggests that additional sources of random error are present. A high $P$ value (>0.99) indicates that random variations are much smaller than expected and also is cause for concern. For example, it could indicate that periodic noise (e.g., 60-Hz line frequency) is being counted. Such signals are not subject to the same degree of random variation as are radiation counting measurements and therefore have very small $\chi^2$ values. In general, a range $0.05 < P < 0.95$ is considered an acceptable result. If $P$ falls outside the range (0.01-0.99), one usually can conclude that something is wrong with the measurement system. If $0.01 < P < 0.05$ or $0.95 < P < 0.99$, the results are suspicious but the experiment is considered inconclusive and should be repeated.

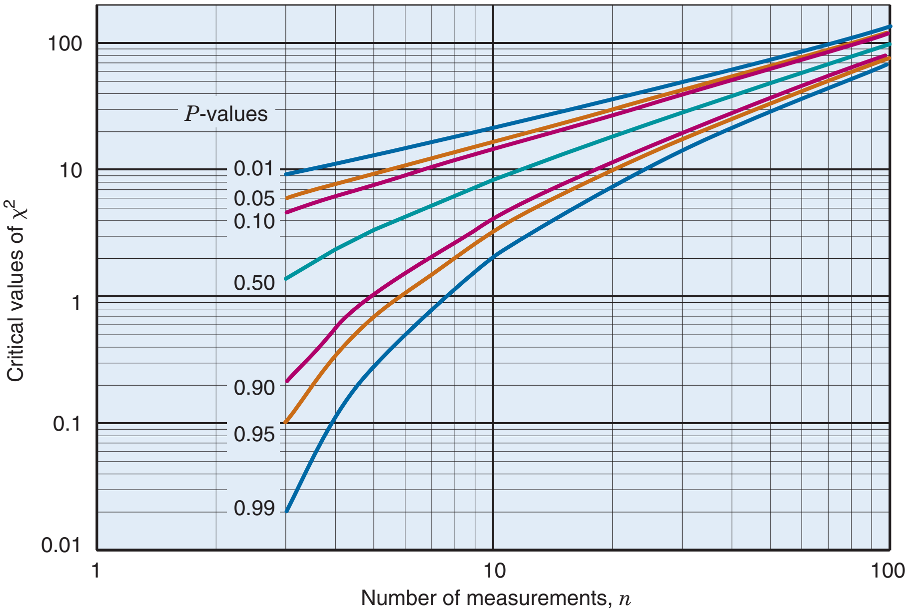


**FIGURE 9-2** Critical values of $\chi^2$ versus number of measurements, $n$. For a properly operating system, one should obtain $\chi^2 \approx (n - 1)$. $P$ values indicate probability of obtaining $\chi^2$ larger than associated curve value. —, $P$ = 0.01-0.99; —, $P$ = 0.05-0.95; —, $P$ = 0.10-0.90; —, $P$ = 0.5.

**EXAMPLE 9-10**

Use the $\chi^2$ test to determine the likelihood that the following set of 20 counting measurements were obtained from a Poisson distribution.

| | |
|---|---|
| 3875 | 3575 |
| 3949 | 4023 |
| 3621 | 3314 |
| 3817 | 3612 |
| 3790 | 3705 |
| 3902 | 3412 |
| 3851 | 3520 |
| 3798 | 3743 |
| 3833 | 3622 |
| 3864 | 3514 |

***Answer***

Using a pocket calculator or by direct calculation, it can be shown that the mean and standard deviation of the counting measurements are

$$\bar{N} = 3717$$

$$SD = 187.4$$

Thus, from Equation 9-38,

$$\begin{aligned}\chi^2 &= 19 \times (187.4)^2/3717 \\ &\approx 179.5\end{aligned}$$

Using Figure 9-2, the calculated value for $\chi^2$ far exceeds the largest critical value shown for $n = 20$; (critical value $\approx 35$ for $P = 0.01$). Hence, we conclude that the probability is very small that the observed set of counting measurements were obtained from a Poisson distribution ($P \ll 0.01$). The observed standard deviation, $SD = 187.4$, also far exceeds what would be expected for a Poisson distribution, $\sqrt{N} = 61$. These results suggest the presence of additional sources of random variation beyond simple counting statistics in the data.

Tables of $\chi^2$ values are provided in most statistics textbooks. It is possible to determine more precise $P$ values from these tables than can be read from Figure 9-2, especially for large values of $n$. However, it should be noted that $\chi^2$ is itself a statistically variable quantity, having a standard deviation ranging from approximately 25% for $n \sim 30$ to approximately 15% for $n \sim 100$. Thus it is unwise to place too much confidence in $\chi^2$ values that are within approximately 10% of a critical value, which is about the accuracy to which values can be read from Figure 9-2. When $\chi^2$ values are close to critical values, it is recommended that the experiment be repeated. A useful discussion of the $\chi^2$ statistic and applications to a variety of tests of nuclear counting systems can be found in reference 1.

## 2. The *t*-Test

The *t*-test (also sometimes called the *Student t-test*) is used to determine the significance of the difference between the means of two sets of data. In essence, the test compares the difference in means relative to the observed random variations in each set. Strictly speaking, the test is applicable only to Gaussian-distributed data; however, it is reasonably reliable for Poisson-distributed data as well (see Fig. 9-1).

Two different tests are used, depending on whether the two sets represent independent or paired data. *Independent* data are obtained from two *different* sample groups, for example, two different groups of radioactive samples, two different groups of patients or animals, and so forth. *Paired* data are obtained from the *same* sample group but at different times or under different measurement conditions, such as the *same* samples counted on two different instruments or a group of patients or animals imaged “before” and “after” a procedure. The test for paired data assumes that there is some degree of correlation between the two measurements of a pair. For example, in an experiment comparing two different radiopharmaceuticals that supposedly have an uptake proportional to blood flow, a subject with a “high” uptake for one radiopharmaceutical may have a “high” uptake for the other as well.

To test whether the difference between the means of two sets of *independent measurements* is significantly different from zero, the following quantity is calculated

$$t = \frac{|\bar{X}_1 - \bar{X}_2|}{\sqrt{[(n_1-1)SD_1^2 + (n_2-1)SD_2^2]/(n_1+n_2-2)}} \times \frac{1}{\sqrt{(1/n_1)+(1/n_2)}} \tag{9-39}$$

where $\bar{X}_1$ and $\bar{X}_2$ are the means of the two data sets, $SD_1$ and $SD_2$ are their standard deviations (calculated as in Equation 9-7), and $n_1$ = number of data values in set 1 and $n_2$ = number of data values in set 2. The vertical lines bracketing the difference of the means indicates that the absolute value should be used. For $n_1 \approx n_2$ and a reasonably large number of samples in each group ($\gtrsim 10$ in each), Equation 9-39 reduces to

$$t = \frac{\left|\bar{X}_1 - \bar{X}_2\right|}{\sqrt{2(SD_1^2 + SD_2^2)/(n_1 + n_2 - 2)}} \tag{9-40}$$

In either case, the calculated value of $t$ then is compared with critical values of the $t$-distribution for the appropriate number of degrees of freedom, $\text{df} = n_1 + n_2 - 2$.

Figure 9-3 shows values of $t$ that would be exceeded at various probability levels if the two sets of data actually were obtained from the *same* distribution. For example, for df = 10, a value of $t \gtrsim 2.2$ would be obtained by chance with a probability of only 5% ($P = 0.05$) if the underlying distributions actually were the same. This probability is sufficiently small that the difference between the means usually would be considered to be "statistically significant," that is, that the underlying distributions very likely have different means.

The $t$ values given by Equations 9-39 and 9-40, and the derivation of associated $P$ values from Figure 9-3 as described earlier, correspond to a *two-sided test*. The $P$ values so obtained express the probability that the observed difference in means of measured data, whether positive or negative, would be obtained if the true means of the underlying distributions actually were the same, that is, $m_1 = m_2$. If a "statistically significant difference" (i.e., a very low $P$ value) is obtained, one concludes that $m_1 \neq m_2$, which could imply either that $m_1 > m_2$, or that $m_2 > m_1$. For example, a two-sided test would be appropriate if one were concerned only whether the uptakes of two radiopharmaceuticals were different.

A *one-sided test* is used when one is concerned only whether one mean is greater than the other (e.g., whether the uptake of one radiopharmaceutical is greater than that of the other). For example, if the experimental result is $\bar{X}_1 > \bar{X}_2$, one might ask whether this is consistent with $m_1 > m_2$. In this case, for a given $t$ value, the $P$ values in Figure 9-3 are reduced by a factor of 2. The $P$ value then is interpreted as the probability that the observed difference in means of the data would be obtained if $m_1 \leq m_2$, that is, if $m_1 < m_2$ or $m_1 = m_2$. Statisticians generally do not recommend the use of one-sided tests. For example, a nonsignificant one-sided test for $m_1 > m_2$ may overlook the possibility that $m_1 < m_2$, which could be an equally important conclusion.

Note further that, as with the $\chi^2$ statistic, $t$ values have their own statistical variations from one experiment to the next. Thus $t$ values that are within a few percent of a critical value should be interpreted with caution.

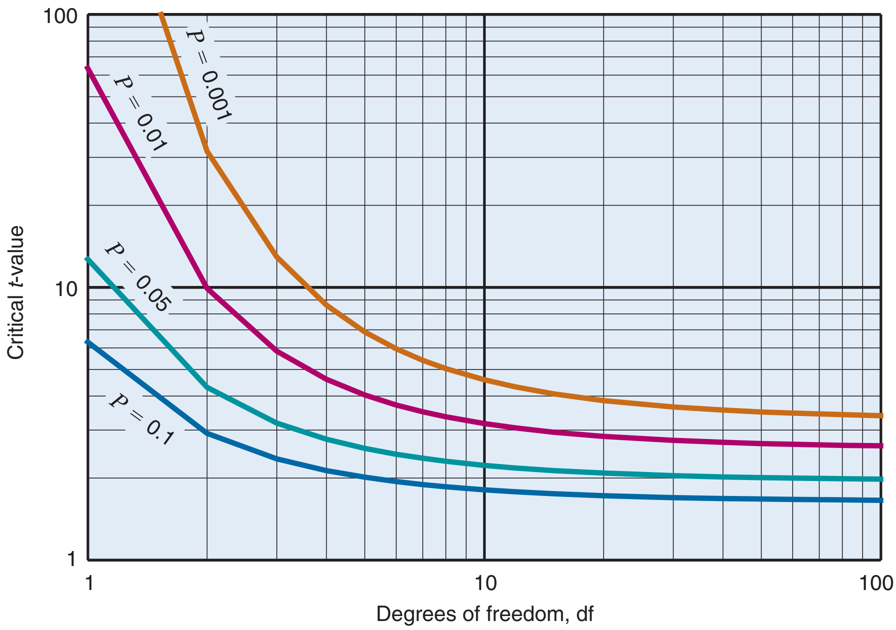


**FIGURE 9-3** Critical values of $t$ versus degrees of freedom (df) for different $P$ values. Curves shown are for a two-sided test of significance.

For most practical situations, $P$ values can be read with sufficient accuracy from Figure 9-3. More precise $t$ and $P$ values are provided in tables in statistics textbooks or by many pocket calculators.

## EXAMPLE 9-11

Suppose the two columns of data in Example 9-10 represent counts measured on two different groups of animals, for the uptake of two different radiopharmaceuticals. Use the $t$-test to determine whether the means of the two sets of counts are significantly different (two-sided test).

***Answer***

Using a pocket calculator or by direct calculation, the means and standard deviations of the two sets of data are found to be (1 = left column, 2 = right column)

$$\bar{X}_1 = 3830$$

$$SD_1 = 87.8$$

$$\bar{X}_2 = 3604$$

$$SD_2 = 195.1$$

Thus, from Equation 9-40,

$$t = \frac{|3830 - 3604|}{\sqrt{2 \times (87.8^2 + 195.1^2)/18}}$$
$$\approx 3.17$$

From Figure 9-3, this comfortably exceeds the critical value of $t$ for df = (10 + 10 – 2) = 18 and $P$ = 0.05 (~2.1) and exceeds as well the value for $P$ = 0.01 (~2.9). Thus we can conclude that it is very unlikely that the means of the two sets of data are the same ($P < 0.01$), and that they are in fact significantly different.

---

For *paired comparisons,* the same table of critical values is used but a different method is used for calculating $t$. In this case, the differences between pairs of measurements are determined, and $t$ is calculated from

$$t = \frac{\left|1/n\sum_{i=1}^{n}(X_{1,i} - X_{2,i})\right|}{(SD_\Delta/\sqrt{n})} \tag{9-41}$$

The numerator is formed by computing the average of the paired differences and taking its absolute value. $SD_\Delta$ is the standard deviation of the paired differences (calculated as in Equation 9-7, with the $N$'s replaced by $\Delta$'s) and $n$ is the number of pairs of measurements. The sign of the difference between individual data pairs is significant and should be used in calculating the mean of the differences. The calculated value of $t$ is compared with critical values in the $t$-distribution table using df = ($n$ – 1). Probability values are interpreted in the same manner as for independent data.

## EXAMPLE 9-12

Suppose that the two columns of data in Example 9-10 represent counts measured on the same group of animals for the uptake of two different radiopharmaceuticals; that is, opposing values in the two columns represent measurements on the same animal. Use the $t$-test to determine whether there is a significant difference in average uptake of the two radiopharmaceuticals in these animals.

***Answer***

The first step is to calculate the difference in counts for each pair of measurements. Subtracting the data value in the right-hand column from that in the left for each pair, one obtains for the differences

$$3875 - 3575 = +300$$

$$3949 - 4023 = -74$$

etc.

The absolute value of the mean difference, $|\bar{\Delta}|$, and standard deviation of the differences are found to be

$$|\bar{\Delta}| = 240.8$$

$$SD_\Delta = 141.0$$

Using Equation 9-41

$$t = \frac{240.8}{(141.0/\sqrt{10})} \approx 5.4$$

From Figure 9-3, the critical value of $t$ for df = $n$ – 1 = 9 and $P$ = 0.01 is $t \approx 3.3$; thus, as in Example 9-11, we can conclude that the means of the two sets of data are significantly different.

---

This discussion of paired data applies for two-sided tests. One-sided tests may be performed using the methods outlined in the discussion of unpaired data.

## 3. Treatment of "Outliers"

Occasionally, a set of data will contain what appears to be a spurious, or "outlier," result, reflecting possible experimental or measurement error. Although generally it is inadvisable to discard data, statistical tests can be used to determine whether it is reasonable, from a *statistical* point of view, to do so. These tests involve calculating the standard deviation of the observed data set and comparing this with the difference between the sample mean $\bar{X}$ and the suspected outlier, $X$. The quantity calculated is

$$T = (X - \bar{X}) / SD \qquad (9\text{-}42)$$

which then is compared with a table of critical values (Table 9-2). The interpretation of the result is the same as for the *t*-test; that is, the critical value is that value of $T$ (also sometimes called the *Thompson criterion*) that would be exceeded by chance at a specified probability level if all the data values were obtained from the same Gaussian distribution. Rejection of data must be done with caution; for example, in a series of 20 measurements, it is likely that at least one of the data values will exceed the critical value at the 5% confidence level.

### EXAMPLE 9-13

In the right-hand column of data in Example 9-10, the value 4023 appears to be an outlier, differing by several standard deviations from the mean of that column (see Example 9-11). Use the Thompson criterion to determine whether this data value may be discarded from the right-hand column of data.

***Answer***

From Example 9-11, the mean and standard deviation of the right-hand column of data are $\bar{X}_2 = 3604$, $SD_2 = 195.1$. Using Equation 9-42

$$\begin{aligned} T &= (4023 - 3604)/195.1 \\ &= 419/195.1 \\ &= 2.15 \end{aligned}$$

According to Table 9-2, for 10 observations and $P = 0.05$, the critical value of $T$ is 2.29. Because the observed value is smaller, we must conclude that there is a relatively high probability ($P > 0.05$) that the value could have been obtained by chance from the observed distribution, and therefore that it should not be discarded.

**TABLE 9-2**
**CRITICAL VALUES OF THE THOMPSON CRITERION FOR REJECTION OF A SINGLE OUTLIER**

| Number of Observations, *n* | Level of Significance, *P* | | |
|---|---|---|---|
| | .1 | .05 | .01 |
| 3 | 1.15 | 1.15 | 1.15 |
| 4 | 1.46 | 1.48 | 1.49 |
| 5 | 1.67 | 1.71 | 1.75 |
| 6 | 1.82 | 1.89 | 1.94 |
| 7 | 1.94 | 2.02 | 2.10 |
| 8 | 2.03 | 2.13 | 2.22 |
| 9 | 2.11 | 2.21 | 2.32 |
| 10 | 2.18 | 2.29 | 2.41 |
| 11 | 2.23 | 2.36 | 2.48 |
| 12 | 2.29 | 2.41 | 2.55 |
| 13 | 2.33 | 2.46 | 2.61 |
| 14 | 2.37 | 2.51 | 2.66 |
| 15 | 2.41 | 2.55 | 2.71 |
| 16 | 2.44 | 2.59 | 2.75 |
| 17 | 2.47 | 2.62 | 2.79 |
| 18 | 2.50 | 2.65 | 2.82 |
| 19 | 2.53 | 2.68 | 2.85 |
| 20 | 2.56 | 2.71 | 2.88 |
| 21 | 2.58 | 2.73 | 2.91 |
| 22 | 2.60 | 2.76 | 2.94 |
| 23 | 2.62 | 2.78 | 2.96 |
| 24 | 2.64 | 2.80 | 2.99 |
| 25 | 2.66 | 2.82 | 3.01 |
| 30 | 2.75 | 2.91 | |
| 35 | 2.82 | 2.98 | |
| 40 | 2.87 | 3.04 | |
| 45 | 2.92 | 3.09 | |
| 50 | 2.96 | 3.13 | |
| 60 | 3.03 | 3.20 | |
| 70 | 3.09 | 3.26 | |
| 80 | 3.14 | 3.31 | |
| 90 | 3.18 | 3.35 | |
| 100 | 3.21 | 3.38 | |

Adapted from Levin S: Statistical Methods. In Harbert J, Rocha AFG (eds): *Textbook of Nuclear Medicine*, Vol 1, ed 2. Philadelphia, 1984, Lea and Febiger, Chapter 4.

## 4. Linear Regression

Frequently, it is desired to know whether there exists a correlation between a measured quantity and some other parameter (e.g., counts versus time, radionuclide uptake versus organ weight, etc.). The simplest such relationship is described by an equation of the form

$$Y = a + bX \qquad (9\text{-}43)$$

Here, $Y$ is the measured quantity and $X$ is the parameter with which it is suspected to be correlated. The graph of $Y$ versus $X$ is a straight line, with $Y$-axis intercept $a$ and slope $b$ (Fig. 9-4).

To estimate values for $a$ and $b$ from a set of data, the following quantities are calculated.*

$$b = \frac{\left[n\sum X_i Y_i - \sum X_i \sum Y_i\right]}{\left[n\sum X_i^2 - \left(\sum X_i\right)^2\right]} \qquad (9\text{-}44)$$

$$a = \bar{Y} - b\bar{X} \qquad (9\text{-}45)$$

Here $n$ is the number of pairs of data values; $X_i$ and $Y_i$ are individual values of these pairs and $\bar{X}$ and $\bar{Y}$ are their means. The summations $\Sigma$ in Equation 9-44 extend over all values of $i$ (1, 2, … $n$).

The quantity $SD_{Y \cdot X}$ is "the standard deviation of $Y$ given $X$," that is, the standard deviation of data values $Y$ about the regression line. It is computed from

$$SD_{Y \cdot X}^2 = \frac{n-1}{n-2} \times \left(SD_Y^2 - b^2 SD_X^2\right) \qquad (9\text{-}46)$$

where $SD_X$ and $SD_Y$ are the standard deviations of $X$ and $Y$ calculated by the usual methods. The estimated uncertainties (standard deviations) in $b$ and $a$ are given by

$$SD_b = SD_{Y \cdot X} / \left[SD_X \sqrt{n-1}\right]$$
$$SD_a = SD_{Y \cdot X} \sqrt{\frac{1}{n} + \frac{\bar{X}^2}{(n-1)SD_X^2}} \qquad (9\text{-}47)$$

Finally, the *correlation coefficient*, $r$, is computed from

$$r = b\,(SD_X / SD_Y) \qquad (9\text{-}48)$$

The correlation coefficient has a value between ±1, depending on whether the slope $b$ is positive or negative. A value near zero suggests no correlation between $X$ and $Y$, (i.e., $b \approx 0$)

*The equations for regression parameters are interrelated and are expressed in a variety of ways in different textbooks. See recommended additional texts at the end of this chapter.

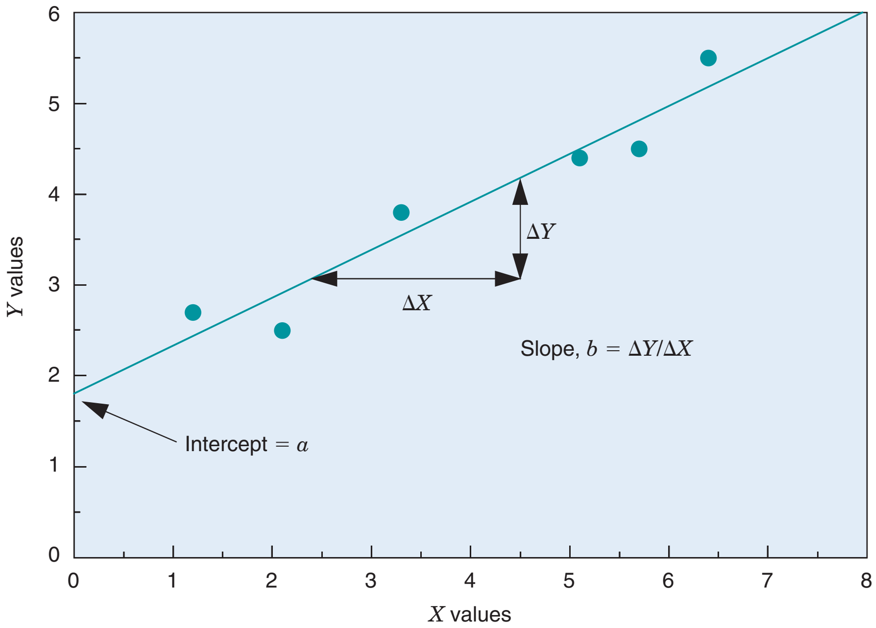


**FIGURE 9-4** Hypothetical example of data and linear regression curve. ● = data values; —— = calculated regression curve; $Y = a + bX$, $a$ = $Y$-axis intercept; $b$ = slope, $\Delta Y/\Delta X$.

and a value near ±1 suggests a strong correlation.*

An alternative method for evaluating the strength of the correlation and its statistical significance is to determine whether $b$ is significantly different from zero. This can be done by calculating

$$t = b/SD_b \tag{9-49}$$

and comparing this to critical values of the $t$-distribution (see Fig. 9-3). The number of degrees of freedom is df = $(n - 2)$ in which $n$ is the number of $(X,Y)$ data pairs. If the calculated value of $t$ exceeds the tabulated critical value at a selected significance level, one can conclude that the data support the hypothesis that $Y$ is correlated with $X$. A similar analysis can be performed (using $SD_a$) to determine whether the intercept, $a$, is significantly different from zero.

*An intuitively attractive interpretation of the correlation coefficient is that $r^2$ is the fraction of the observed variance of the data set $Y$ that actually is attributable to variations in $X$ and the dependence of $Y$ on $X$. Thus, $r^2 = 0.64$ ($r = 0.8$) implies that 64% of the observed variance $SD_Y^2$ actually is caused by the underlying variations in $X$, with the remaining 36% attributable to "other factors" (including random statistical variations).

## REFERENCE

1. Tries MA, Skrable KW, French CS, Chabot GE: Basic applications of the chi-square statistic using counting data. *Health Phys* 77:441-454, 1999.

## BIBLIOGRAPHY

**Many useful general statistics texts are available. Some that were used for this chapter include the following:**

Bevington PR: *Data Reduction and Error Analysis for the Physical Sciences*, New York, 1992, McGraw-Hill. [includes computer disk]

Crow EL, Davis FA, Maxfield MW: *Statistics Manual*, New York, 1960, Dover Publications.

Snedecor GW, Cochran WG: *Statistical Methods*, ed 8, Ames, IA, 1989, Iowa State University Press.

**Additional discussion of nuclear counting statistics may be found in the following:**

Evans RD: *The Atomic Nucleus*, New York, 1972, McGraw-Hill, Chapters 26 and 27.

Knoll GF: *Radiation Detection and Measurement*, ed 4, New York, 2010, John Wiley, Chapter 3.

Leo WR: *Techniques for Nuclear and Particle Physics Experiments*, ed 2, New York, 1994, Springer-Verlag, Chapter 3.