Counting Statistics

Applied Health Physics

Oak Ridge Associated Universities

Objectives

- To review the need for statistics in health physics.
- To review measures of central tendency and dispersion.
- To review common statistical models (distributions).
- To calculate the uncertainty for a count and count rate.
- To review the statistics for a single count (rate) versus those for a series of counts of the same sample.

Introduction

- It is impossible to directly measure radiation in every situation at every point in space and time.
- It is impossible to know with complete certainty the level of radiation when it was not directly measured.

Introduction

- One of the goals of statistics is to make accurate inferences based on incomplete data.

- Using statistics, it is possible to extrapolate from a set of measurements to all measurements in a scientifically valid way.

Introduction

Statistical tests allow us to answer questions such as:

- How much radioactivity is present?
- How accurate is the measured result?
- Is the detector working properly? (Are the detector results reproducible (precise))?
- Is there bias error in addition to random error in a measurement?
- Is this radioactive?

Introduction

- Without statistics, it is difficult to answer these questions.

- Statistics allow us to answer questions with a degree of confidence that we are drawing the right conclusion.

PART I:

DEFINE THE TERMS

Uncertainty (Error)

Uncertainty (or error) in health physics is the difference between the true average rate of transformation (decay) and the measured rate.

- Accuracy refers to how close the measurement is to the true value.
- Precision refers to how close (reproducible) a repeated set of measurements are.

Uncertainty

- Random errors fluctuate from one measurement to the next and yield results distributed about some mean value.

- Systematic (bias) errors tend to shift all measurements in a systematic way so the mean value is displaced.

Uncertainty

- An example of random error results from the measurement of radioactive decay.
- The process of radioactive decay is random in time.
- An estimate of how many atoms will decay can be made but not the ones that will decay.

Uncertainty

Some examples of bias error are:

- Dose-rate dependence of instruments
- Throwing out high and/or low data points (outliers)
- Uncertainty of a radioactivity standard

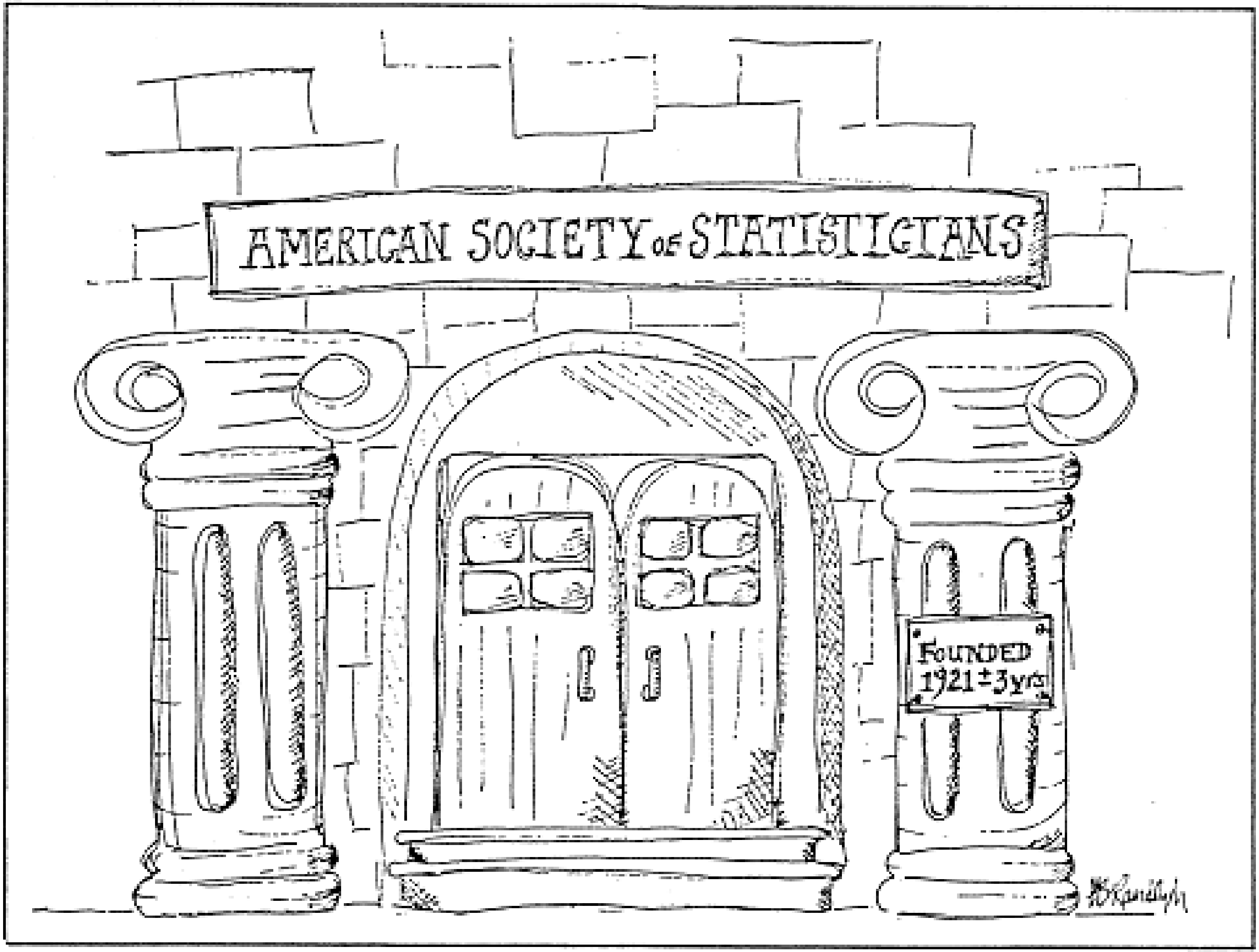
Uncertainty

Uncertainty results from:

- Measurement errors
- Sample collection errors
- Sample preparation errors
- Analysis errors
- Survey design errors
- Modeling errors

Assumptions

- Nuclear transformations are random, independent, and occur at a constant rate.
- The physical half-life is long compared to the counting time.



Science, vol. 234

Descriptive Statistics

- Descriptive statistics characterize a set of data.
- Descriptive statistics include:
  - Measures of central tendency.
  - Measures of dispersion.
- Descriptive statistics should always be performed on a set of data.

Measures Of Central Tendency

- The mean is the arithmetic midpoint or more commonly, the average.
- If the true mean is known, it is designated as μ.
- For example, the true mean age of the people in this classroom could be determined.

Central Tendency

- However, the true mean of radioactive samples is not usually known.
- An estimate of the mean is made for radioactive samples.
- The estimated mean is designated $\overline{X}$.

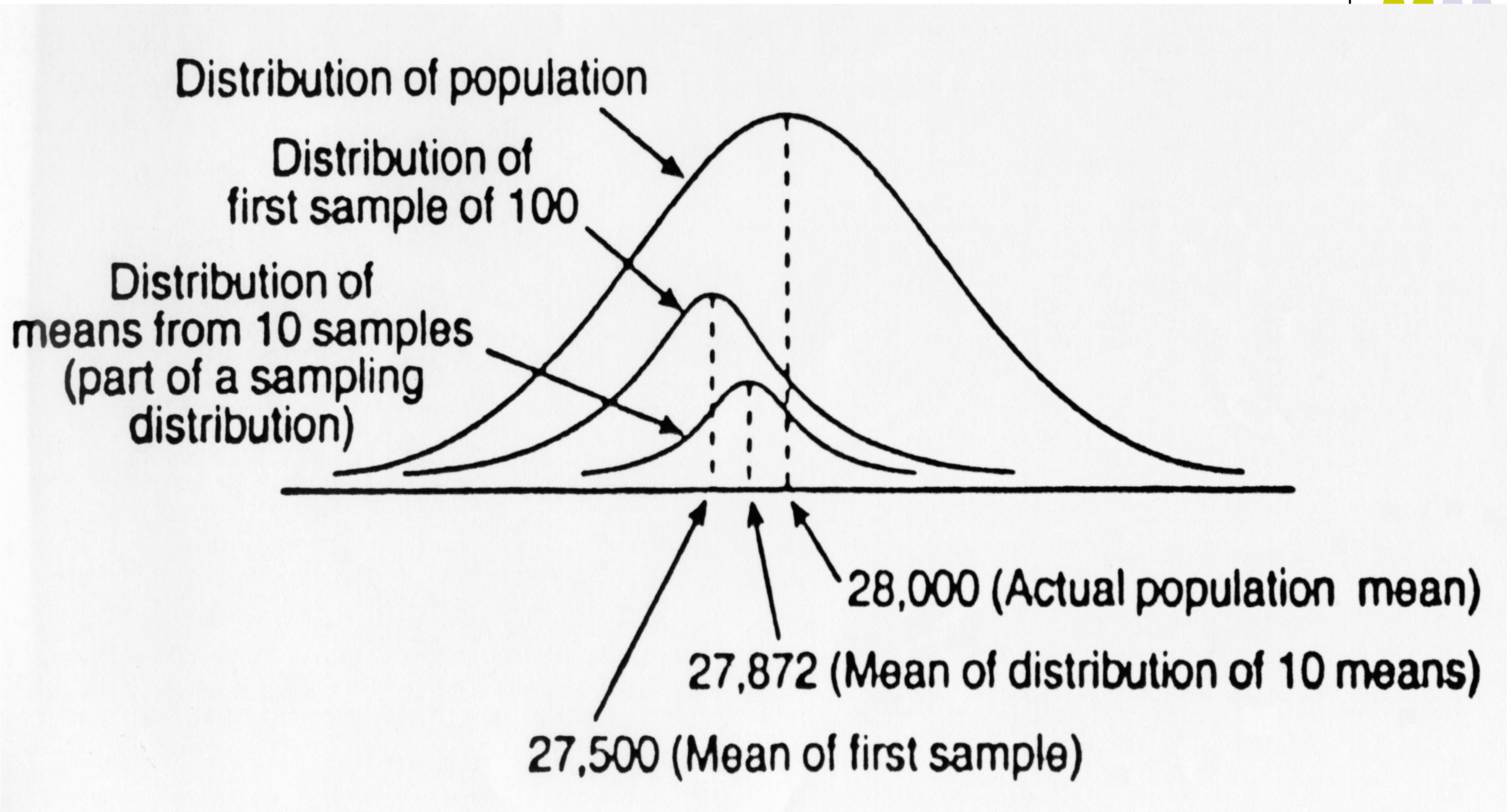
Central Tendency

The estimated mean $\overline{X}$:

$$\overline{X} = \frac{\sum X_i}{n}$$

- **X_i = an observed count (result).**
- **n = the number of observations (counts) made.**

$$\text{as } n \longrightarrow \infty, \text{ then } \overline{X} \longrightarrow \mu$$


Distribution of population
Distribution of
first sample of 100
Distribution of
means from 10 samples
(part of a sampling
distribution)
28,000 (Actual population mean)
27,872 (Mean of distribution of 10 means)
27,500 (Mean of first sample)

Central Tendency

$$\text{as } n \longrightarrow \infty, \text{ then } \bar{X} \longrightarrow \mu$$

This is the basis for pooling data, or counting more than once, etc.

Central Tendency

- The median is the point at which half the measurements are greater and half are less.
- The median is a better measure of central tendency for skewed distributions like dosimetry data or environmental data.

Central Tendency

- For example, assume we wanted to calculate the mean salary of everyone in the classroom.

- The mean would probably be a good measure of central tendency.

Central Tendency

- But suppose Bill Gates walked into the classroom.
- The mean salary would not reflect my salary!
- The median would be a better measure of central tendency for that severely skewed distribution.

Measures of Dispersion

- The range is the difference between the maximum and minimum values.
- The deviation is the difference between a given value and the true mean, and can be positive or negative.

Dispersion

- The mean deviation is of little use since positive and negative values will cancel each other out.

- The variance is the square of the mean deviation, and assures positive values.

Dispersion

- The true variance is:

$$\sigma^2 = \frac{\sum (X_i - \mu)^2}{n}$$

- The true standard deviation is:

$$\sigma = \sqrt{\frac{\sum (X_i - \mu)^2}{n}}$$

Dispersion

In health physics, the sample (group) variance and sample (group) standard deviation is used, since the true mean of radioactive samples is usually not known.

Dispersion

- The sample (group) variance is:

$$S^2 = \frac{\sum (X_i - \overline{X})^2}{n - 1}$$

- The sample (group) standard deviation is:

$$S = \sqrt{\frac{\sum (X_i - \overline{X})^2}{n - 1}}$$

Distributions

- Distributions are statistical models for data.

- They are used as theoretical means to calculate statistics, like the sample (group) standard deviation, for example.

Distributions

- The binomial distribution is a fundamental frequency distribution governing random and dichotomous (2-sided) events.

- This fits radioactivity well, since transformation is random and dichotomous.

Distributions

- The Poisson distribution is a special case of the binomial distribution in which the probability of an event is small and the sample is large.

- The Poisson distribution also fits radioactivity very well, since the probability (p) of any one atom transforming is small, and a sample usually consists of a large number of atoms (>100).

Distributions

The Poisson distribution is simple to use as a model, because it is characterized by one independent parameter, the mean.

$$\mu = np$$

Distributions

- The standard deviation is derived from the mean:

$$\sigma = \sqrt{\mu}$$

- For example, μ is the count, and σ is the square root of the count.

PART II:

SINGLE COUNT

Counting Statistics for a Single Count

- The Poisson distribution is important for counting radioactive samples.
- We assume that a single count is the mean of a Poisson distribution, and the square root of the count is the standard deviation.

Single Count

- For example, a single count of a sample was:

$$X = 209 \,\text{counts}$$

- The standard deviation of that count is:

$$\sigma = \sqrt{209} = 14.5 \text{ counts}$$

Single Count

This means that 68% of any repeated counts of this sample would be between:

$$X = 209 \pm 14.5 \text{ counts}$$

$$194.5 \le X \le 223.5$$

Single Count Rate

Many times in health physics, counting results are expressed as a count rate (R) rather than a total count (X) by dividing by the count time (t).

$$R = \frac{X}{t}$$

Single Count Rate

Continuing with the previous example, if the count time t were 10 minutes, the count rate (R) would be:

$$R = \frac{209\ cts}{10\ \min} = 20.9\,cpm$$

Single Count Rate

- The standard deviation of the count rate (σ_R) can be calculated two ways:

$$\sigma_R = \frac{\sqrt{X}}{t} \qquad or \quad \sigma_R = \sqrt{\frac{R}{t}}$$

- The next slide shows how these two equations are equivalent.

Single Count Rate

$$R = \frac{X}{t} \text{ then } \mathrm{Rt} = \mathrm{X}$$

$$\text{Substitute Rt for X}: \sigma_{\mathrm{R}} = \frac{\sqrt{\mathrm{X}}}{\mathrm{t}} \text{ then } \sigma_R = \frac{\sqrt{Rt}}{t}$$

$$\text{Squaring both sides}: (\sigma_R)^2 = \left(\frac{\sqrt{Rt}}{t}\right)^2$$

$$\text{Reducing the equation}: (\sigma_R)^2 = \frac{Rt}{t^2} = \frac{R}{t}$$

$$\text{Taking the square root of both sides}: \sigma_R = \sqrt{\frac{R}{t}}$$

Single Count Rate

$$\sigma_R = \frac{\sqrt{X}}{t}$$

$$\sigma_R = \sqrt{\frac{R}{t}}$$

$$\sigma_R = \frac{\sqrt{209\,counts}}{10\ \text{min}}$$

$$\sigma_R = \sqrt{\frac{20.9\,cpm}{10\ \text{min}}}$$

$$\sigma_R = 1.4\,cpm$$

$$\sigma_R = 1.4\,cpm$$

Single Count Rate

- When calculating the standard deviation for a count rate, many people feel like they are making a mistake because they are “dividing by t twice”.

- But this is correct.

$$\sigma_R = \sqrt{\frac{R}{t}} = \sqrt{\frac{(\frac{X}{t})}{t}}$$

Single Count Rate

- So, for the previous example:

$$\sigma_R = \sqrt{\frac{20.9\,cpm}{10\ \text{min}}} = 1.4\,cpm$$

- The result would be expressed:

$$20.9\,cpm \pm 1.4\,cpm$$

$$\text{or } 19.5\,cpm \le R \le 22.3\,cpm$$

Coefficient of Variation

- Sometimes it is convenient to express the uncertainty as a percent of the count.

- This is called the coefficient of variation (COV) or the percent variation (%error).

Coefficient of Variation

For a given count: $\% \sigma_X = \dfrac{100 \ \sigma_X}{X}$

$$\text{Squaring both sides}: (\%\sigma_X)^2 = \frac{(100)^2 X}{X^2} = \frac{(100)^2}{X}$$

$$\text{Taking the square root of both sides}: \%\sigma_X = \frac{100}{\sqrt{X}}$$

Coefficient of Variation

For a given count rate: $\%\sigma_R = \dfrac{100\ \sigma_R}{R}$

$$\text{Since } \sigma_R = \frac{\sqrt{X}}{t}, \text{ then } \%\sigma_R = \frac{100\ \frac{\sqrt{X}}{t}}{\frac{X}{t}}$$

$$\text{Squaring both sides}: (\%\sigma_R)^2 = \frac{(100)^2\ \frac{X}{t^2}}{\frac{X^2}{t^2}} = \frac{(100)^2 X}{X^2} = \frac{(100)^2}{X}$$

$$\text{Taking the square root of both sides}: \%\sigma_R = \frac{100}{\sqrt{X}}$$

Coefficient of Variation

Continuing with the previous example:

$$\%\sigma_R = \frac{100}{\sqrt{209}} = 6.9\%$$

Coefficient of Variation

- Note that the COV decreases with increasing values of the count.

- For example, if the count was 809 instead of 209:

$$\%\sigma_R = \frac{100}{\sqrt{809}} = 3.5\%$$

Coefficient of Variation

The COV can be reduced to an arbitrary value by increasing the counting time:

$$t = \frac{1}{R}\left(\frac{100}{\%\ \sigma}\right)^2$$

Coefficient of Variation

- For example, suppose a 1% uncertainty (error) was desired for the previous example:

$$t = \left(\frac{1}{20.9\,cpm}\right)\left(\frac{100}{1}\right)^2 = 478 \text{ min or almost } 8 \text{ hours}$$

- Some survey instruments now display %error, so you can count until you reach the desired % error.

Uncertainty of Net Count Rate

- Often in health physics, measurements are expressed as net count rates.

$$R_{net} = R_g - R_b$$

- There is uncertainty associated with both R_g and R_b.

Net Count Rate

- Uncertainties cannot simply be added or subtracted when two numbers are combined.

- The proper way to calculate the combined uncertainty with more than one term is to sum the squares of the uncertainties, then take the square root.

Net Count Rate

For example, the uncertainty of a net count rate is:

$$\sigma_{Rnet} = \sqrt{\sigma_{R_g}{}^2 + \sigma_{R_b}{}^2}$$

Net Count Rate

$$\sigma_{Rnet} = \sqrt{\sigma_{R_g}^2 + \sigma_{R_b}^2} \text{ and } \sigma_{\mathrm{R}} = \sqrt{\frac{R}{t}}$$

$$\sigma_{Rnet} = \sqrt{\left(\sqrt{\frac{R_g}{t_g}}\right)^2 + \left(\sqrt{\frac{R_b}{t_b}}\right)^2}$$

$$\sigma_{Rnet} = \sqrt{\frac{R_g}{t_g} + \frac{R_b}{t_b}}$$

Net Count Rate

- For example, suppose the gross count of a sample was 2,000 counts for a 10 minute count.
- The background was 200 counts for a 5 minute count.
- What is the net count rate and associated uncertainty?

Net Count Rate

$$R_{net} = \left(\frac{2{,}000\, counts}{10\, m} \right) - \left(\frac{200\, counts}{5\, m} \right) = 160\, cpm$$

$$\sigma_{net} = \sqrt{\frac{200\, cpm}{10\, m} + \frac{40\, cpm}{5\, m}} = 5.3\, cpm$$

Net Count Rate

The result would be expressed:

$$160\ cpm \pm 5.3\ cpm$$

$$154.7\ cpm \leq R_{net} \leq 165.3\ cpm$$

Net Count Rate
Special case: Rate-meters

The uncertainty associated with a reading on a count rate meter is:

$$\sigma_{crm} = \sqrt{\frac{\text{count rate}}{2 \text{ x time constant}}}$$

Time constant = response time

Net Count Rate

The optimum distribution of a given total counting time between the gross count and background count can be determined by determining the ratio:

$$\frac{t_g}{t_b} = \sqrt{\frac{R_g}{R_b}}$$

Net Count Rate

- Using the previous example:

$$\frac{t_g}{t_b} = \sqrt{\frac{R_g}{R_b}}$$

$$\frac{t_g}{t_b} = \sqrt{\frac{200\ cpm}{40\ cpm}} = 2.23$$

- Then the time allotted for the gross count should be 2.23 times that allotted for the background.

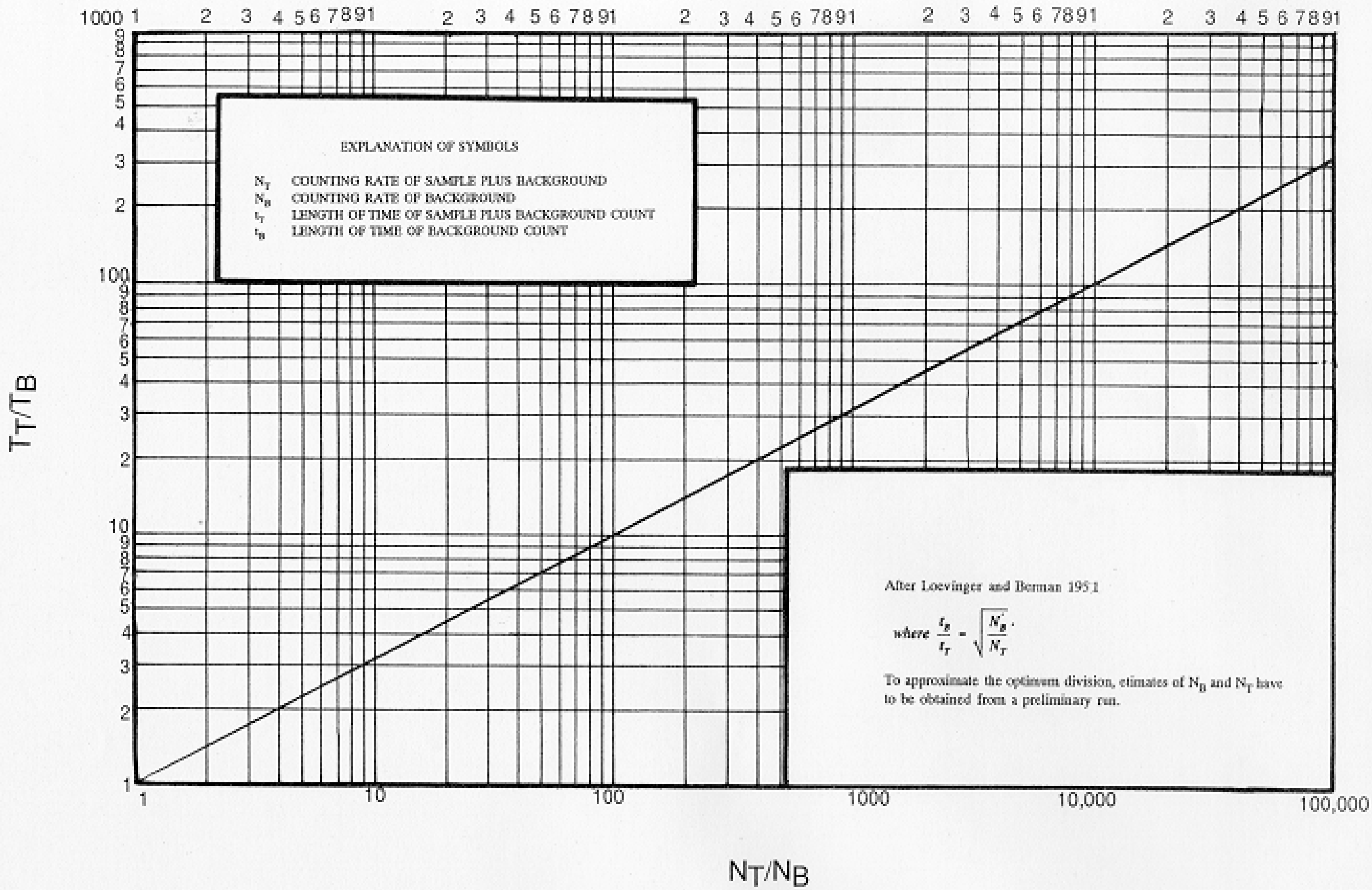


Figure 9.9 Graph Showing the Most Efficient Distribution of Counting Time Between Sample and Background

PART III:

MULTIPLE COUNTS

Counting Statistics for Several Counts

- Normally a radioactive sample would be counted only once, and a Poisson distribution assumed.

- In other words, the standard deviation is calculated from the mean.

Several Counts

However, to determine if uncertainty is due only to the random nature of radioactivity and not other sources, the standard deviation can be determined experimentally from a series of counts.

Several Counts

- This is the approach taken in quality control testing of equipment.
- The distribution of counts from a sample repeatedly counted is best described by a Gaussian (or normal) distribution.

Several Counts

- The Gaussian distribution is familiar to many people as the “bell-shaped” curve used in grading.

- The percentage of counts that will fall within a given range around the mean are:
  - **$\pm 1\sigma$ 68.3%**
  - **$\pm 2\sigma$ 95.5%**
  - **$\pm 3\sigma$ 99.7%**

Several Counts

The Gaussian distribution can be described by the:

- Mean.

$$\overline{X} = \frac{\sum X_i}{n}$$

- Standard deviation.

$$S = \sqrt{\frac{\sum \left(X_i - \overline{X}\right)^2}{n-1}}$$

Several Counts

- For a sufficiently large sample (mean > 30), the Poisson distribution approximates the Gaussian distribution.

- Therefore, if the only uncertainty is from radioactivity, the standard deviation of a series of counts of the same sample determined assuming a Gaussian distribution should nearly equal the standard deviation determined assuming a Poisson distribution.

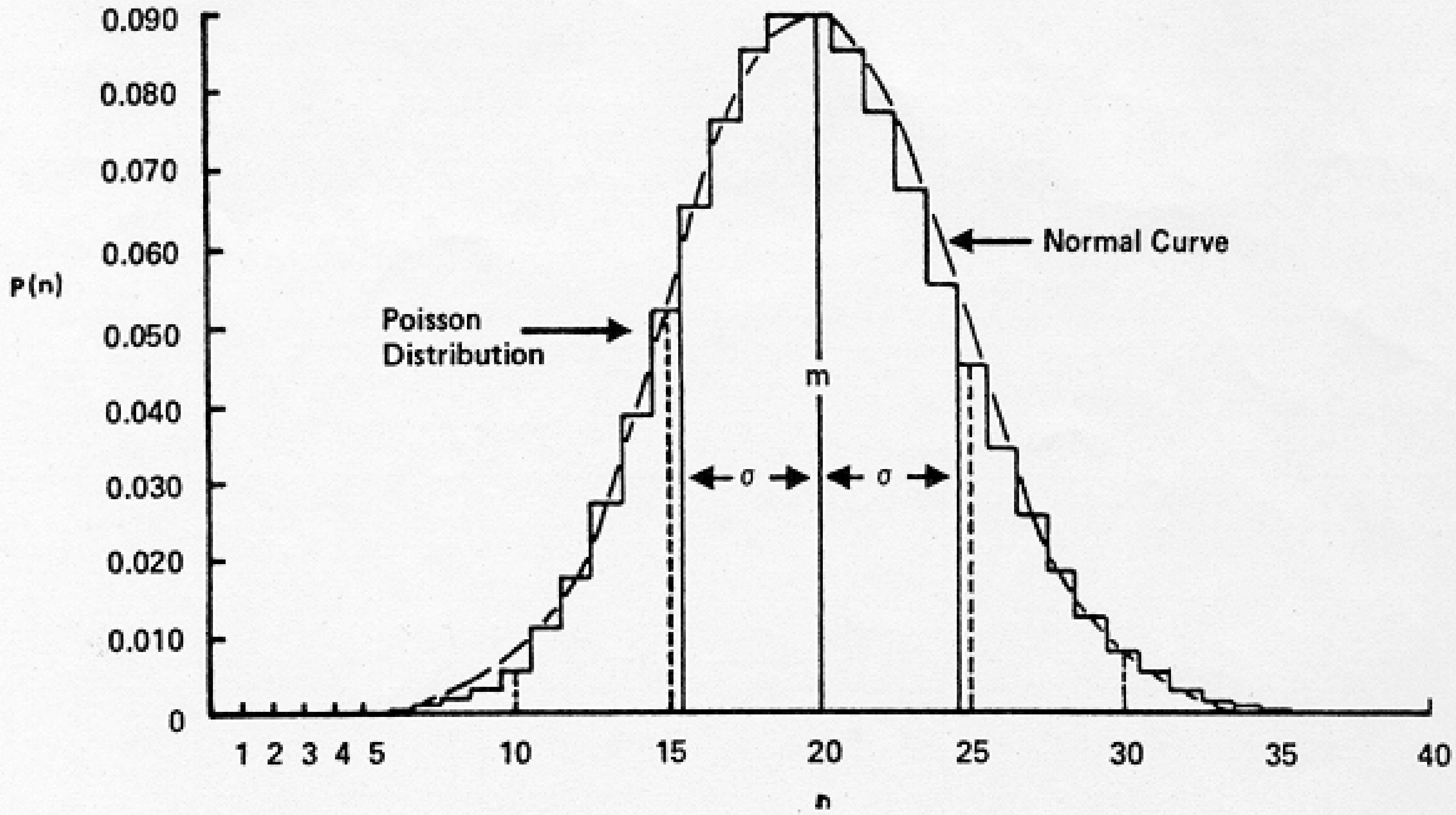


Figure 1. Comparison of a Poisson Distribution and a Normal Curve for $m = 20$ and $\sigma = \sqrt{20} = 4.47$.

Several Counts

If the two standard deviations are not nearly equal, then other factors (equipment problems) are contributing to uncertainty, and should be eliminated.

Several Counts

- There are several ways to determine “nearly equal”.
- The Chi-Square Test is one method.
- The other method is the Reliability Factor, and we’ll use it in the statistics lab exercise.

Several Counts

The Reliability Factor is defined as the standard deviation determined by Gaussian statistics divided by the Poisson standard deviation.

$$R.F. = \frac{S}{\sqrt{\overline{\overline{X}}}}$$

Several Counts

- The RF versus the number of repetitions is then plotted on a graph such as the one in your Radiological Health Handbook.

- One of three conclusions can be drawn from the position of the RF value on the graph.

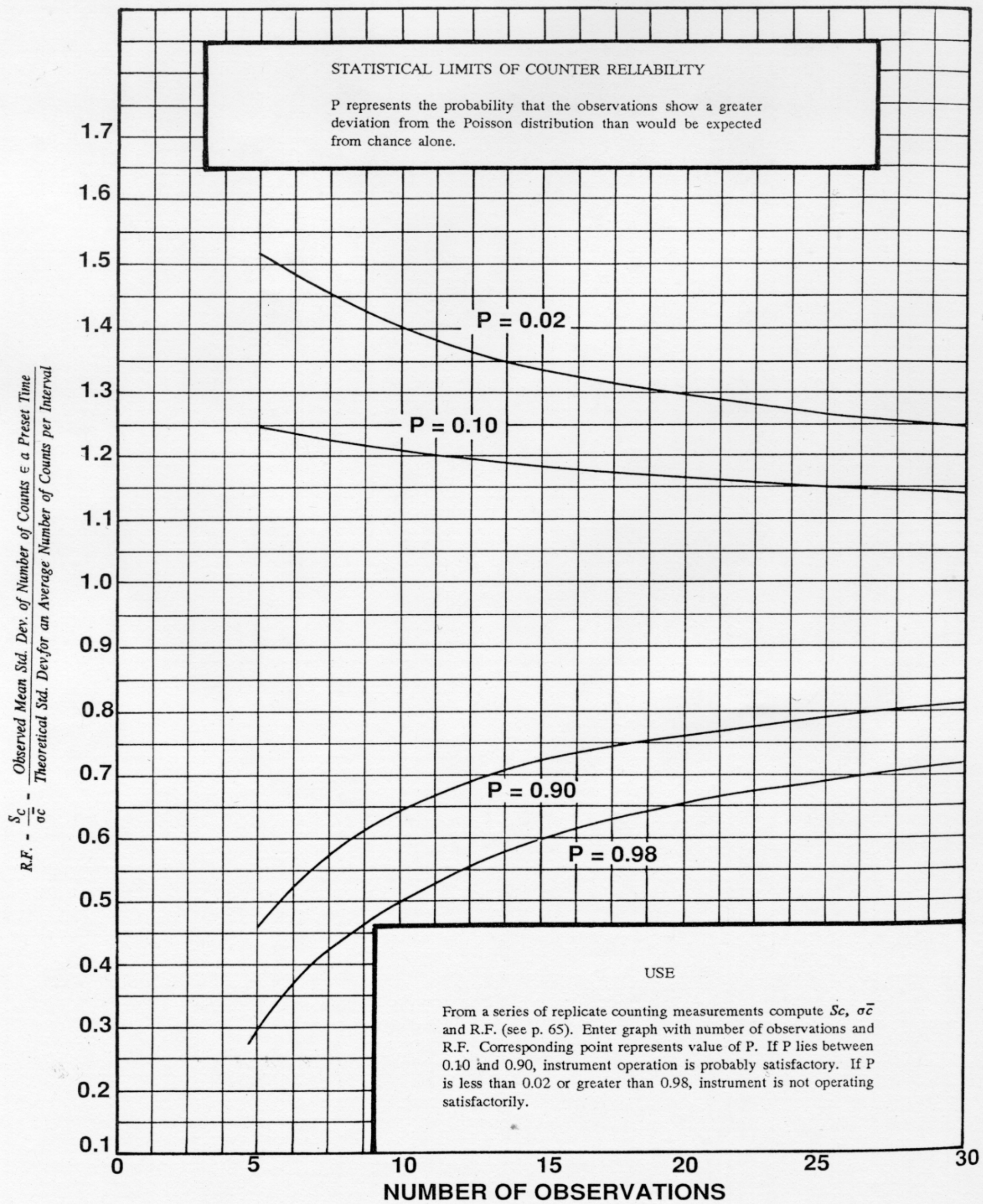


Figure 9.11 Statistical Limits of Counter Reliability (After Radiological Health Handbook, 1970)

Several Counts

- If the point lies within the inner set of lines, the uncertainty is due only to the random nature of radioactivity.
- If the point lies outside the outer set of lines, there is a strong indication that other uncertainty is involved (instrument calibration, etc.).
- If the point lies between the lines, the result is inconclusive and needs to be repeated.

Repetition	Counts	$\left(X_i - \overline{X}\right)^2$
1	209	449
2	217	174
3	248	317
4	235	23
5	224	38
6	223	52
7	233	8
8	250	392
9	228	5
10	235	23
Total	2302	1481
Mean	230.2	

Several Counts

- Gaussian:

$$\overline{X} = \frac{2302}{10} = 230.2$$

$$S = \sqrt{\frac{1481}{9}} = 12.8$$

- Poisson:

$$\overline{X} = \frac{2302}{10} = 230.2$$

$$\sigma = \sqrt{\overline{X}} = \sqrt{230.2} = 15.2$$

Several Counts

- The RF is:

$$RF = \frac{12.8}{15.2} = 0.84$$

- From the graph, this is an acceptable RF for 10 repetitive counts.

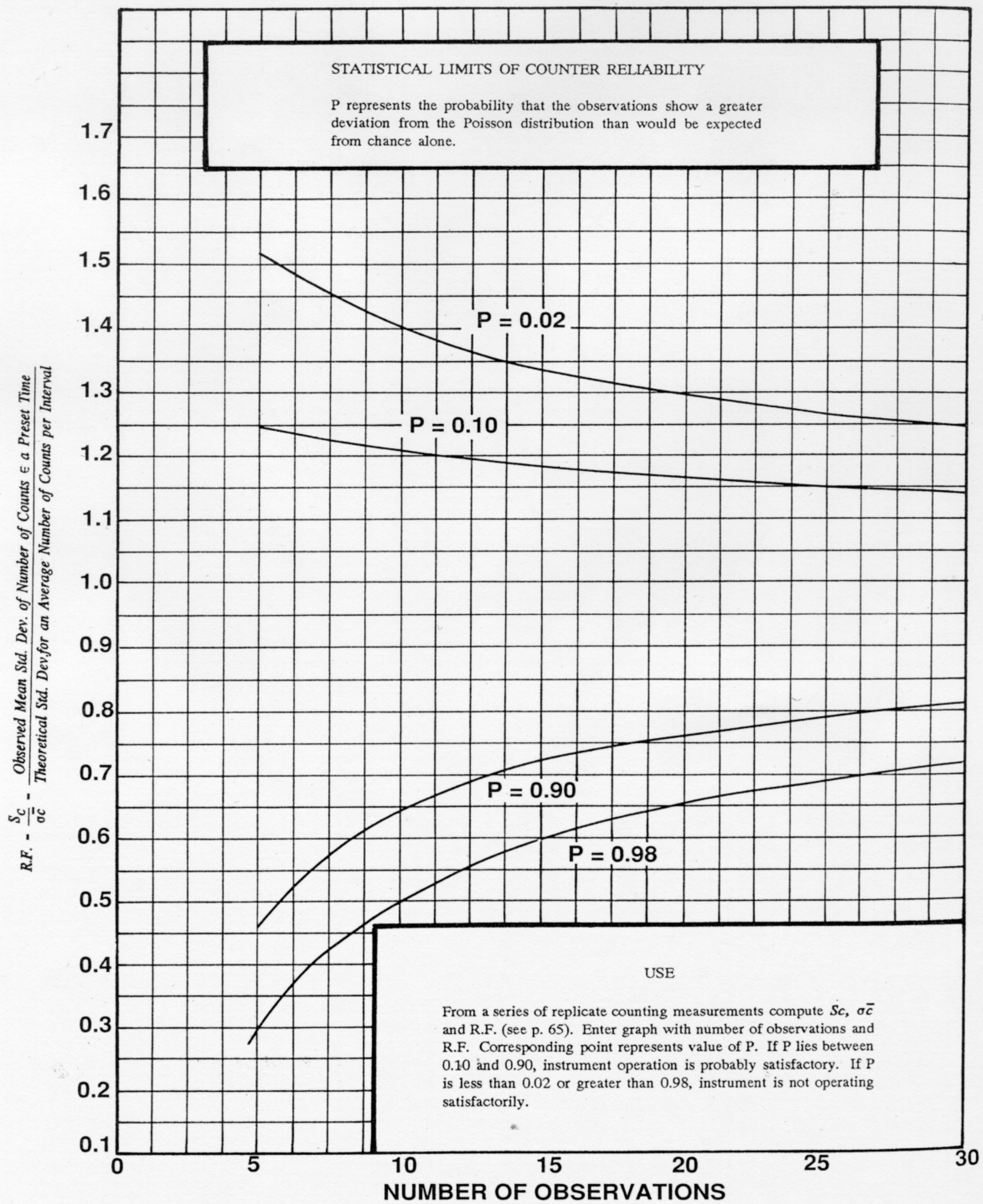


Figure 9.11 Statistical Limits of Counter Reliability (After Radiological Health Handbook, 1970)

References

Cember H. Introduction to Health Physics(1996).

Gollnick, DA. Basic Radiation Protection Technology, 4th edition (2000).

Knoll, GF. Radiation Detection and Measurement, 2nd edition (1989).

National Council on Radiation Protection and Measurements. NCRP Report #112 Calibration of Survey Instruments Used in Radiation Protection for the Assessment of Ionizing Radiation Fields and Radioactive Surface Contamination (1991).

Shleien, BS; Slaback Jr., LA; Birky, BK. Handbook of Health Physics and Radiological Health, 3rd edition (1998).