Exercise 2: darts, statistics, and the normal distribution

Purpose: investigate the probability distribution function for a measured quantity resulting from small random errors. Explore the difference between precision and accuracy.

Introduction

The statistical methods presented in earlier exercises are based on measured values of a quantity being symmetrically distributed about the true value. The spread in the values was assumed to be due to small random errors. In this case the limiting distribution describing the results was a bell-shaped curve centered on the true value. If properly normalized (the area beneath the curve equal to unity), the curve describes a probability distribution function for a given result. The most probable result is where the curve has a maximum. Improbable results are where the curve has small values. The probability for measuring a value x of a quantity with a true value of X is given by the *normalized Gauss distribution function* (specific probability distribution also known as the *normal distribution*),

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-X)^2/2\sigma^2} .$$

The factor σ characterizes the width of the curve. A broad curve has large σ corresponding to large uncertainty and poor precision of the measurements. A narrow curve has small σ corresponding to small uncertainty and good precision of the measurements. Verify the given Gauss distribution is correctly normalized by integration over all possible values of x. Present your calculation.

Laboratory assignment

Each group should grab a set of darts. The object is to hit the center target as many times as possible.

1. Take a few practice throws before you begin your turn.
2. Each student should throw darts at the center target (bin number 7). The total number of darts thrown per group should total no less than 200. Record how many darts you throw in each bin.
3. After each student has completed throwing the required number of darts at the target, tabulate the total number of darts thrown in each bin of the paper target for your group. The bins are numbered 1 through 13. Report your group data in the "n_i (# of darts)" column of Table 1.
4. Calculate the average bin hit for the group using the formula given in the appendix.

$$X_{group} = \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.$$

5. Calculate the corrected standard deviation for the group using the formula in the appendix.

$$\sigma_{group} = \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.$$

6. Calculate the standard error of the mean for the group using the formula in the appendix.

$$\text{SEM}_{group} = \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.$$

Table 1: **Group** data

Bin	# of darts n_i
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	

7. Compare your data with other groups.
8. Which group had the greatest accuracy (average closest to the true value of 7)? _____________.
9. Which group had the greatest precision (smallest standard deviation)?

    ____________________.

10. Combine the total darts thrown for the whole class and place the values in the "# of darts n_i" column of Table II.

Table 2: **Class** data

Bin	# of darts n_i	$\langle n_i \rangle$	$(\langle n_i \rangle - n_i)^2 / \langle n_i \rangle$
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			
13			

11. Calculate the average bin hit for the class using the formula given in the appendix.

$$X_{class} = \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.$$

12. Calculate the corrected standard deviation for the class using the formula in the appendix.

$$\sigma_{class} = \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.$$

13. Calculate the standard error of the mean for the class using the formula in the appendix.

$$\text{SEM}_{class} = \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.$$

14. Calculate the expected number of darts for each bin $\langle n_i \rangle$ using the formula given in the appendix and place in Table II.
15. Calculate the value for $(\langle n_i \rangle - n_i)^2 / \langle n_i \rangle$ and place in Table II.
16. Calculate the *chi-squared* value for the observed number of darts compared with the expected number for the entire class's results. Use $Q_i^2 = \langle N_i \rangle_{\text{observed}}$, which is appropriate when comparing an integer number of observed measurements with an expected number. In this case the uncertainty Q_i is given by *Poisson statistics*.

$$\chi^2 = \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.$$

17. Calculate and report the *reduced chi-squared* value.

$$\tilde{\chi}^2 = \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_.$$

18. If a value of $\tilde{\chi}^2$ is much larger than 1, the expected distribution is not supported. State whether or not the normal distribution accurately describes the experimental results and why.

____

____.

Equipment list: Darts, 8.5"x13" paper targets with 13-equal area bins.

Appendix I: formula for analyzing finite binned data representing a Gaussian distribution

For bins of equal width, the mean can be calculated using a simple average,

$$X = \frac{1}{N_{\text{darts}}} \sum_{i=1}^{N_{\text{bin}}} x_i n_i \,,$$

where x_i is the ith bin, n_i is the number of darts thrown in that bin, N_{bin} is the total number of bins, and N_{darts} is the total number of darts.

The standard deviation of binned data can be calculated by taking the square root of the variance. The standard deviation of the weighted sum follows as

$$\sigma = \frac{1}{\sqrt{N_{\text{darts}}}} \sqrt{\sum_{i=1}^{N_{\text{bin}}} (x_i - X)^2 n_i} \,.$$

Another important quantity is the standard error of the mean, which is an estimate of how far the sample mean is likely to be from the population mean. The standard error of the mean for a finite number of samples may be taken as

$$\text{SEM} = \frac{\sigma}{\sqrt{N_{\text{darts}}}} \,.$$

The expected number of darts at each bin location (estimated from the center of each bin) is found by simply multiplying the total number of darts thrown by the probability distribution defined as the variable N_{darts} multiplied by the probability distribution $P(x)$. For the normal distribution,

$$\langle n_i \rangle = \int_{x_{i-1/2}}^{x_{i+1/2}} N_{\text{darts}} P(x)\, dx$$

$$= \int_{x_{i-1/2}}^{x_{i+1/2}} \frac{N_{\text{darts}}}{\sigma\sqrt{2\pi}}\, e^{-(x-X)^2/2\sigma^2}\, dx$$

$$\approx \frac{N_{\text{darts}}}{\sigma\sqrt{2\pi}}\, e^{-(x_i-X)^2/2\sigma^2}\, \Delta x$$

where you should use $\Delta x = 1$ bin. Note that it is much simpler to use this approximation when tabulating your results in procedure 9.

The *chi-squared* value calculated from our tabulated data for the measured value of the number of darts thrown in each bin is given by

$$\chi^2 = \sum_{i=1}^{N_{\text{bin}}} \frac{(\langle n_i \rangle - n_i)^2}{Q_i^2} \,.$$

For Poisson statistics, the denominator Q_i^2 is replaced via the substitution $Q_i = \sqrt{\langle n_i \rangle}$. This substitution is relevant for this laboratory and should be used.

The *reduced chi-squared* value is given by

$$\tilde{\chi}^2 = \frac{\chi^2}{d},$$

where d is the number of degrees of freedom. The number of degrees of freedom d is the number of bins minus the number of constraints. Since the expected number of measurements is derived from three values characterizing the class data (total number of “measurements,” average measured value, and standard deviation), the number of constraints is three. Thus, the number of degrees of freedom is 13 – 3 = 10.