Statistical methods – an introduction (SS 2022)

Problem set 9

Problem 1 [3 points] *Systematic errors: impact on sample mean*

a) We consider a similar situation as on page 138 of the script, but now for N measurements of the variable $x_i = x_i^R + x_i^S$, $i = 1, N$, where x_i^R is an iid-variate. Each measurement suffers from an independent, but *common* systematic contribution $x_i^S \equiv x^S$, and

$$\forall\, i: \quad E(x_i^R) = \mu_x, \quad Var(x_i^R) = \sigma^2, \quad \text{and} \quad E(x^S) = \mu_s, \quad Var(x^S) = S^2$$

Determine $E(\bar{x})$ and $Var(\bar{x})$, when $\bar{x}$ is the arithmetic (=sample) mean of the N measurements x_i.

b) What would be the result for the expectation value and the variance of the arithmetic mean, if the systematic errors would be different (though following the same distribution) for each measurement i? In other words, here, we assume that $x_i^S \neq x_j^S$ for $i \neq j$ but still $E(x_i^S) = \mu_s$ and $Var(x_i^S) = S^2$.

Problem 2 [6 points] *Systematic errors*

a) Assume that you can measure a certain length with a yardstick, with random reading error $\pm\Delta r$, and that there is a systematic shift of the zero-point, $z = z_0 \pm \Delta z$, where Δz is the uncertainty of this shift.

   Thus, if you would try to measure a certain (actual) length a', your measurement would need to be corrected by $-z_0$, i.e.,

$$a' = (a_{\text{obs}} - z_0) \pm \Delta r \pm \Delta z.$$

   Now assume that you measure two lengths, a_{obs} and b_{obs}, with the same stick, and that you are interested in the ratio of the *actual* lengths, a'/b'. Calculate the *expected* value for this ratio, and the corresponding total error, as a function of the above quantities.

b) Adopt the following values: $a_{\text{obs}} = 40$ cm, $b_{\text{obs}} = 20$ cm, $\Delta r = 0.5$ cm, $z_0 = 6$ cm and $\Delta z = 1$ cm. Provide your result for the (expected) ratio of the actual lengths, and its error.

c) Check your results numerically, by a suitable simulation using normally distributed errors. Then vary Δz from zero to 3 cm, and investigate the reaction.

   Check also what happens if you would have used two different yardsticks with the same z_0 and corresponding uncertainty (but a different realization of the actual z).

Problem 3 [3 points] *Efficiency of estimators, and weighted mean*

x_1, x_2 and x_3 are the elements of a sample drawn independently from a continuous population of unknown expectation value $\hat{x}$ but known (population) variance σ^2.

a) Show that

$$\begin{aligned} S_1 &= \frac{1}{2}x_1 + \frac{1}{5}x_2 + \frac{3}{10}x_3 \\ S_2 &= \frac{1}{5}x_1 + \frac{2}{3}x_2 + \frac{2}{15}x_3 \\ S_3 &= \frac{1}{5}x_1 + \frac{9}{20}x_2 + \frac{7}{20}x_3 \end{aligned}$$

are unbiased estimators of $\hat{x}$.

b) Calculate the variances $\sigma^2(S_1), \sigma^2(S_2), \sigma^2(S_3)$.

c) Show that the arithmetic mean $\bar{x} = 1/3 \sum_{i=1}^{3} x_i$ has the smallest variance of all estimators of the type $S = \sum_{i=1}^{3} a_i x_i$ with the constraint $\sum_{i=1}^{3} a_i = 1$. Compute this variance and compare with the variances obtained in b).

d) Now, x_1, x_2 and x_3 shall be three different measurements of the same quantity. The measurement error for x_1 shall be σ_1, the error for x_2 shall be a factor n_2 larger, and the error for x_3 shall be a factor n_3 larger. Calculate the weights a_i for an optimum estimator S_4 of the expectation value $\hat{x}$ (constructed in analogy to S_1 to S_3). Express the error of S_4 in units of σ_1.

Much success!