# Statistical methods – an introduction (SS 2022)

# Problem set 8

1. In this problem set, the (measurement) error $\Delta x$ of a quantity $x$ is, as usual, an abbreviation for $\sigma_x$, i.e., if the actual measurement is $x = x_m + \delta x$, then $E(x) = x_m$, and $\Delta x = \sqrt{Var(x)} = \sqrt{E\{(\delta x)^2\}}$.

2. For calculating variances and covariances, use the calculation rules provided in Chap. 6 of the script, resulting from a linearization. In particular, remember the last equation on page 134.

**Problem 1** [2 points] *Error propagation*

Suppose a function

$$f(x, y) = x + y,$$

where $x$ and $y$ are *not* independent. What is the corresponding squared error, $(\Delta f)^2$? Now, let $y = g(x)$ and re-calculate $(\Delta f)^2$ as a function of $(\Delta x)^2$ by determining $cov(x, g(x))$.

Show that the result is identical with the squared error $(\Delta f)^2$ when *directly* calculating this quantity for the expression

$$f(x) = x + g(x).$$

**Problem 2** [5 points] *Wind-momentum and luminosity*

[Note: In the following, 'logarithmic' refers to the logarithm with base 10.]

The theory of radiation driven winds predicts a linear relation between the logarithmic luminosity, $\log L'$, and the logarithmic 'modified' wind-momentum rate, $\log D'$, of a hot star ($T_{\rm eff} >$ 10,000 K), where

$$\begin{aligned} L' &= T_{\rm eff}'^4 R_*'^2 \\ D' &= \dot{M}' v_\infty' R_*'^{0.5}, \end{aligned}$$

and the primes indicate convenient normalizations (such that all quantities are dimensionless). $T_{\rm eff}$ and $R_*$ are stellar effective temperature and radius, and $\dot{M}$ and $v_\infty$ are the mass-loss rate and the terminal velocity of the corresponding wind, respectively.

Unfortunately, typical diagnostics does not directly yield the mass-loss rate, but a related quantity,

$$Q' = \frac{\dot{M}'}{R_*'^{3/2}}.$$

Note: $Q', T_{\rm eff}', R_*', v_\infty'$ are independent quantities!

a) Which normalization quantities would be useful for the first of the above equations ($L' = \ldots$)?

b) In order to check the theoretical predictions or to use the observed relation, one has to plot $\log D'$ vs. $\log L'$, and to perform a linear regression. To obtain meaningful results and errors, the covariance between both quantities needs to be accounted for! Calculate the covariance matrix for ($\log L'$, $\log D'$) in terms of $Var(\log T'_{\text{eff}})$, $Var(\log R'_*)$, $Var(\log Q')$ and $Var(\log v'_\infty)$.

c) In the above considerations, $Var(\log x)$ corresponds to

$$Var(\log x) = \sigma^2(\log x) \approx \big(\Delta \log(x)\big)^2 = \big(\log(x+\Delta x) - \log(x)\big)^2.$$

Show that for $\Delta x/x \ll 1$, $Var(\log x)$ can be approximated in terms of $\frac{\Delta x}{x}$. Derive the corresponding relation, and convince yourself about the validity of this approximation, by using appropriate examples (numbers).

d) As outlined in b), a typical plot displays $\log D'$ vs. $\log L'$. From theory, this should result in a linear relation (at least for objects of not too different spectral type), and one of the interesting questions relates to the slope of this relation, $m$. To "measure" this slope, one performs a linear regression, i.e., minimizes the corresponding $\chi^2$ from a series of measurements for different objects, $i$,

$$\chi^2 = \sum_i^N \Big(\frac{y_i - (mx_i + b)}{\sigma_i}\Big)^2.$$

$y_i$ corresponds to $\log D'_i$, $x_i$ to $\log L'_i$, $b$ is the other regression parameter, the intercept, and $\sigma_i$ the error. Contrasted to the standard approach, however, in our case one has also to consider the errors on the x-axis, and the correlation between $\log D'$ and $\log L'$. This can be achieved by using the same algorithm, however interpreting now $\sigma_i$ as the *total* error of the devation between $y_i$ and the theoretically expected relation:

$$\sigma_{\text{tot},i} =: \Delta l_i, \quad \text{with} \quad l_i = \log D'_i - (m \log L'_i + b).$$

Derive $\Delta l_i$ using your results from problem b). Note that $m$ and $b$ have to be considered as constants. Both values can be derived by an iterative approach, but here we assume that they are already known.

When comparing with the "standard approach" ($\Delta x_i = 0$), there is an additional term in the expression for $\sigma_{\text{tot},i}$. When is this additional term larger than zero, i.e., under which conditions is the total error larger than the error for $\log D'_i$ alone? Quantify your result, by adopting $(\Delta\ T_{\text{eff}}/T_{\text{eff}}) = 0.1$, and $(\Delta\ R_*/R_*) = 0.3$. (If you would need other errors, you have done something wrong.)

**Problem 3** [5 points] *Transformed covariance matrix*

Assume that you can independently measure the mass, $m$, and velocity, $v$, of a particle, and that there are no systematic errors. The relative measurement errors shall be known and constant, i.e., loosely speaking, $\Delta m/m = a$ and $\Delta v/v = b$. More precisely, we assume $\sigma_m/m_m = a$ and $\sigma_v/v_m = b$ (see top of this problem set).

a) Write down the covariance matrix for $m$ and $v$.

b) Calculate the corresponding (transformed) covariance matrix for the momentum $p$ and kinetic energy $E$ of the particle, and express the matrix elements in terms of $p_m, E_m, a$ and $b$, where $p_m$ and $E_m$ are the 'true' values of $p$ and $E$

c) Determine the (linear) correlation coefficient, $\rho(p, E)$, in dependence of $a, b$, and derive its value for the three cases $a = b$, $a = 0$, and $b = 0$, respectively. What do you conclude, particularly from the results for $a = 0$ and $b = 0$, respectively?

Have fun, and much success!