Instructions

1. The Data Analysis competition will be 3 hours in duration and is marked out of a total of 150 points.

2. There are **Detailed Worksheets** for carrying out detailed work / rough work. On each of the **Detailed Worksheets**, please fill in

- Student Code
- Question no.
- Page no. and total number of pages.

3. Start each problem on a new page of the Detailed Worksheets. Please write only on the printed side of the sheet. Do not use the reverse side. If you have written something on any sheet that you do not want to be marked, cross it out.

4. Graph Paper is required for your solutions. On each Graph Paper sheet, please fill in

- Student Code
- Question no.
- Graph no. and total number of graph paper sheets used.

5. There is a summary **Answer Sheet** with your student ID code for your final answers.

6. Please remember that the graders may not understand your language. As far as possible, write your solutions only using mathematical expressions and numbers. If it is necessary to explain something in words, please use short phrases (if possible in English).

7. You are not allowed to leave your exam desk without permission. If you need any assistance (malfunctioning calculator, need to visit a restroom, need more Detailed Worksheets, etc.), please put up your hand to signal the invigilator.

8. The beginning and end of the competition will be indicated by a long sound signal. Additionally, there will be a short sound signal fifteen minutes before the end of the competition (before the final long sound signal).

9. At the end of the competition you must stop writing immediately. Sort and put your Summary Answer Sheets, Graph Papers, and Detailed Worksheets in one stack. Put all other papers in another stack. You are not allowed to take any sheet of paper out of the examination area.

10. Wait at your table until your envelope is collected. Once all envelopes are collected, your student guide will escort you out of the competition room.

D1. Photometric comparison of surveys (75 points)

You are an astronomer working with large photometric surveys, such as the Sloan Digital Sky Survey (SDSS) and the Dark Energy Survey (DES), both of which have your host, Observatório Nacional, as a participant. SDSS used a 2.5 m telescope in Apache Point, USA, during the 2000s, and DES used a 4 m telescope in Cerro Tololo, Chile, from 2013 to 2019. Even though they mostly covered different hemispheres of the sky, they had an equatorial region in common known as Stripe 82 that you can use to compare and calibrate the photometry of different data sets, like SDSS and DES.

The following tables containing object positions and magnitudes from Stripe 82 were downloaded for analysis. However, due to a file system corruption on the computer, the file names were scrambled, and now you cannot tell which table belongs to which survey.

Tables 1 and 2 appear next to each other below, with an identification number for each source, its equatorial coordinates, and its magnitude in the g-band (m_g) with its error (err m_g).

(a) (5 points) From these tables, which survey (SDSS or DES) is Table 1 and which is Table 2? Assume that both surveys are equivalent regarding detector response, exposure times, and site characteristics.

(b) (35 points) Using the data in the table, plot the magnitude (m_g) on the x-axis (linear scale) and the error in magnitude (err m_g) on the y-axis (logarithmic scale) using the semi-log paper marked as Graph 1. Estimate the angular coefficient A (slope) and linear coefficient B (y-axis intercept) for each dataset. There is no need to calculate the associated errors.

(c) (5 points) The Signal to Noise ratio (S/N) is approximately the inverse of the error in the magnitude, $S/N \approx 1/(\text{err } m_g)$. Using the linear fit calculated in the previous part, what is the S/N reached for each survey at a magnitude of $m_g = 21.5$ mag?

(d) (15 points) An object in Table 1 that is within 1 arcsecond of an object in Table 2 can be considered to be the same object. By looking at the RA and Dec of the objects in both tables, identify the objects in common and write down a new table with the matching IDs, ID_1 and ID_2.

(e) (15 points) Using the matched table from part (d), plot the g-band magnitude of each survey against the other, Table 1 on x-axis, and Table 2 on y-axis using the millimetre (linear) paper marked as Graph 2. Draw on error bars for each point in both horizontal and vertical directions, using values **double** err m_g (known as a 2σ uncertainty). From your graph, identify the stars that would be suitable for photometric calibration between the two surveys and write down their correspondings IDs from Table 1.

Table 1					Table 2				
ID_1	RA	Dec	m_g	err m_g	ID_2	RA	Dec	m_g	err m_g
	(deg)	(deg)	(mag)	(mag)		(deg)	(deg)	(mag)	(mag)
1	0.047255	0.000406	21.7649	0.0120	1	0.006167	0.066874	21.9020	0.0576
2	0.064741	0.021568	21.1111	0.0067	2	0.018660	0.007450	21.8039	0.0529
3	0.064911	0.026395	21.3931	0.0084	3	0.047853	0.061487	21.3007	0.0418
4	0.098343	0.054871	21.3934	0.0088	4	0.050870	0.015659	21.1678	0.0388
5	0.022256	0.039129	21.9933	0.0157	5	0.051270	0.020812	21.2524	0.0401
6	0.006188	0.066928	21.5490	0.0088	6	0.057414	0.075999	21.8884	0.0578
7	0.083945	0.074259	21.9395	0.0126	7	0.064745	0.021583	21.3634	0.0422
8	0.076715	0.079496	21.4808	0.0089	8	0.064910	0.026419	21.6428	0.0488
9	0.057422	0.076006	21.8897	0.0127	9	0.071102	0.091058	21.9259	0.0751
10	0.024412	0.087688	21.8341	0.0126	10	0.074946	0.002792	21.3258	0.0410
11	0.044723	0.091782	21.8868	0.0172	11	0.076709	0.079474	21.5303	0.0476
12	0.071089	0.091053	21.4390	0.0098	12	0.092635	0.077395	21.6995	0.0513
					13	0.098343	0.054854	21.6542	0.0499
					14	0.099332	0.093711	21.8802	0.0577

D2. Shapley Hypothesis (75 points)

Globular clusters are one of the oldest components of galaxies. About a century ago, Harlow Shapley studied the distribution of globular clusters in the Milky Way in order to determine the distance from the Sun to the Galactic Centre, with the hypothesis that globular clusters were symmetrically distributed around the Galactic Centre. The table below shows the positions and distance moduli of a few known globular clusters in the Milky Way. The first three columns in the table show the cluster name, galactic longitude (*l*), and galactic latitude (*b*). The fourth column shows the distance modulus (i.e. the difference between the apparent and absolute magnitude), for which the values are extinction-corrected. Based on the data in the table:

Name	*l* (degrees)	*b* (degrees)	Distance modulus (mag)
NGC 6522	1.025	-3.926	14.3
NGC 6401	3.450	3.980	14.4
NGC 6342	4.898	9.725	14.5
NGC 6553	5.253	-3.029	13.6
NGC 6440	7.729	3.801	14.6
Ter 12	8.358	-2.101	13.6
VVV-CL160	10.151	0.302	14.2
2MASS-GC01	10.471	0.100	12.6
NGC 6517	19.225	6.762	14.8
NGC 6402	21.324	14.804	14.8
NGC 6712	25.354	-4.318	14.3
NGC 6426	28.087	16.234	16.6
NGC 5466	42.150	73.592	16.0
NGC 7089	53.371	-35.770	15.3
NGC 288	151.285	-89.380	14.8
NGC 2298	245.629	-16.006	15.0
NGC 4590	299.626	36.051	15.1
NGC 4372	300.993	-9.884	13.8
NGC 362	301.533	-46.247	14.7
BH 140	303.171	-4.307	13.4
NGC 5927	326.604	4.860	14.6
Patchick 126	340.381	-3.826	14.5
NGC 5897	342.946	30.294	15.5
NGC 6380	350.182	-3.422	14.9
Djor 1	356.675	-2.484	15.0

(a) (25 points) Calculate the distance (in parsecs) of each globular cluster from the Sun as well as their Cartesian coordinates (x, y, z). The x-axis points to the Galactic Centre and the y-axis points in the direction of galactic rotation. The system is right-handed.

(b) (15 points) From the given data, estimate the distance from the Sun to the centre of the distribution of globular clusters and its uncertainty.

(c) (30 points) To test the validity of Shapley's hypothesis that globular clusters are symmetrically distributed around the Galactic Centre, make histograms with five bins (i.e. sort the data and divide them into five equally-sized intervals) for each of the distributions in the x, y, and z directions. Mark the value of the quartiles (Q_1, Q_2, Q_3) of the three distributions with solid lines on the histograms.

Hint: The three quartiles divide the sorted sample into four sections, each containing 25% of the data, with the second and third sections representing the interquartile range.

(d) (5 points) Using the quartiles, calculate the symmetry factor value for the three distributions as given by:

$$\Phi_x = \frac{|Q_{1,x}+Q_{3,x}-2Q_{2,x}|}{Q_{3,x}-Q_{1,x}} \;,\; \Phi_y = \frac{|Q_{1,y}+Q_{3,y}-2Q_{2,y}|}{Q_{3,y}-Q_{1,y}} \;,\; \Phi_z = \frac{|Q_{1,z}+Q_{3,z}-2Q_{2,z}|}{Q_{3,z}-Q_{1,z}}$$

Classify the three distributions in the x, y, and z directions based on their calculated symmetry factor values, according to the table shown below. Hence, on the answer sheet, write True (**T**) if the analysed sample follows Shapley's hypothesis or False (**F**) otherwise.

Symmetry factor value	Symmetry type
$0.0 \leq \Phi \leq 0.1$	symmetrical
$0.1 < \Phi \leq 0.2$	quasi-symmetrical
$\Phi > 0.2$	asymmetrical