Data Analysis Round

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).

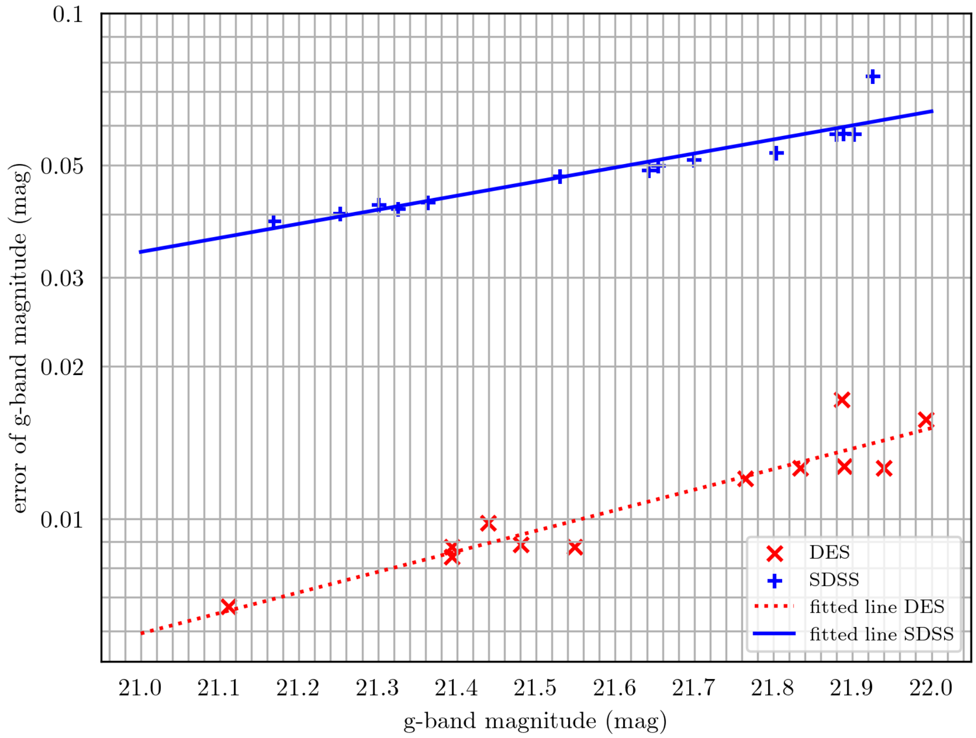
Table 1					**Table 2**				
ID_1	RA (deg)	Dec (deg)	m_g (mag)	err m_g (mag)	ID_2	RA (deg)	Dec (deg)	m_g (mag)	err m_g (mag)
1	0.047255	0.000406	21.7649	0.0120	1	0.006167	0.066874	21.9020	0.0576
2	0.064741	0.021568	21.1111	0.0067	2	0.018660	0.007450	21.8039	0.0529
3	0.064911	0.026395	21.3931	0.0084	3	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

(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.

> **Solution:** DES is Table 1 and SDSS Table 2, since faint stars will have higher uncertainties. SDSS uses a smaller telescope (2.5 m) than DES (4 m) hence, it is typically shallower and has larger errors. Just looking at the error distribution should be enough to tell who has larger errors.

(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.

> **Solution:** The linear regression of each curve:



The reported parameters are:

SDSS:

$$\log_{10}(err\ m_g) = A \cdot m_g + B$$

$A = 0.2780$ (acceptable from 0.2502 to 0.3058 which is within 10% range of the fit)

$B = -7.3108$(acceptable from -8.0419 to -6.5797 which is within 10% range of the fit)

DES:

$$\log_{10}(err\ m_g) = A \cdot m_g + B$$

$A = 0.4063$(acceptable from 0.3657 to 0.4470 which is within 10% range of the fit)

$B = -10.7597$(acceptable from -10.2217 to -11.2977 which is within 10% range of the fit)

(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?

Solution: Using that $S/N \sim 1/err\ m_g$ and above fits we can arrive at the following:
SDSS: $\log_{10}(err\ m_g) = 0.2780 \cdot 21.5 - 7.3108 \rightarrow err\ m_g = 0.0464$
DES: $\log_{10}(err\ m_g) = 0.4063 \cdot 21.5 - 10.7597 \rightarrow err\ m_g = 0.0095$
Accepted answers follow below.

	err	S/N
SDSS err at 21.5 mag	0.05	**22** (acceptable from 19.8 to 24.2 -> within 10% range of the result)
DES err at 21.5 mag	0.01	**106** (acceptable from 95.4 to 116.6 -> within 10% range of the result)

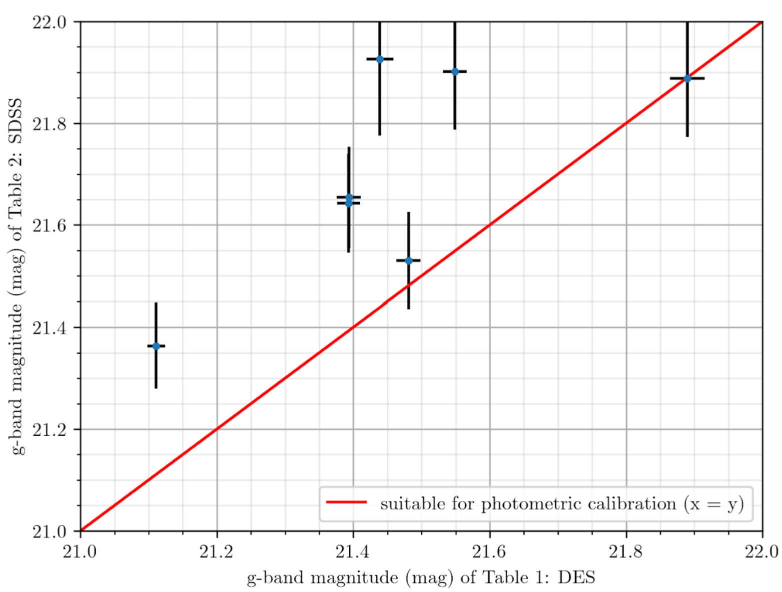
(d) (15 points) An object in Table 1 that is within 1 arcsecond of an object in Table 2 can be considered to be the same object. By looking at the RA and Dec of the objects in both tables, identify the objects in common and write down a new table with the matching IDs, ID_1 and ID_2.

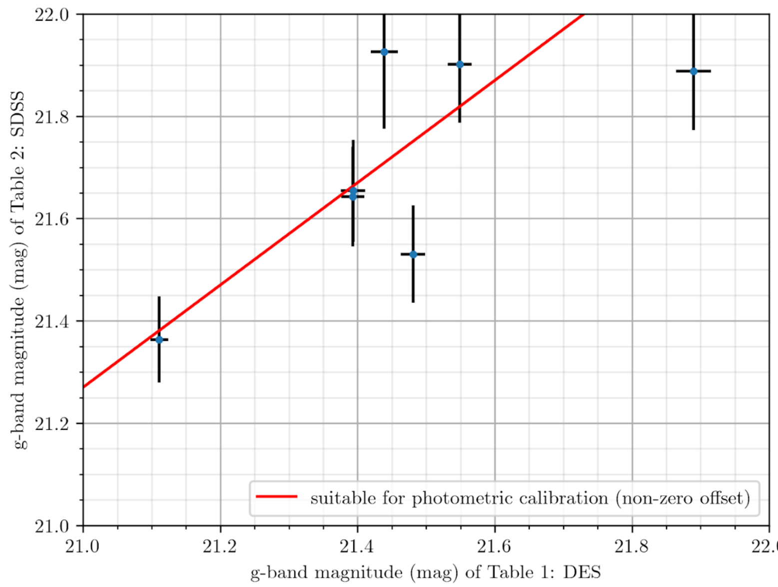
Solution: One tip for this question is that the student should realize that the SDSS RA coordinate is sorted, so it can be used as a reference for the scanning of DES coordinates to perform the match. These are the stars that can be matched between catalogs:

ID_1	RA (deg)	Dec (deg)	m_g (mag)	err m_g (mag)	ID_2	RA (deg)	Dec (deg)	m_g (mag)	err m_g (mag)	Sep (arcsec)
3	0.064911	0.026395	21.3931	0.0084	8	0.064910	0.026419	21.6428	0.0488	0.08475
9	0.057422	0.076006	21.8897	0.0127	6	0.057414	0.075999	21.8884	0.0578	0.03724
4	0.098343	0.054871	21.3934	0.0088	13	0.098343	0.054854	21.6542	0.0499	0.06186
6	0.006188	0.066928	21.5490	0.0088	1	0.006167	0.066874	21.9020	0.0576	0.2076
12	0.071089	0.091053	21.4390	0.0098	9	0.071102	0.091058	21.9259	0.0751	0.05009
2	0.064741	0.021568	21.1111	0.0067	7	0.064745	0.021583	21.3634	0.0422	0.05655
8	0.076715	0.079496	21.4808	0.0089	11	0.076709	0.079747	21.5303	0.0476	0.08276

(e) (15 points) Using the matched table from part (d), plot the g-band magnitude of each survey against the other, Table 1 on the x-axis, and Table 2 on the 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 corresponding IDs from Table 1.

Solution: The figure of g magnitude of DES vs SDSS should look like this with the best linear fit:





The student should identify the stars that are closest to the one-to-one line (x = y) between surveys. This could be with a zero offset, or a non-zero offset. The **two** stars that would be suitable for photometric calibration with zero offset are the ones with ID's **8** and **9** of Table 1 (DES). The **four** stars that would be suitable for photometric calibration with non-zero offset are the ones with ID's **2, 3, 4 and 6** of Table 1 (DES).

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
VW-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) (20 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.

Solution: The first step is to convert the extinction corrected distance moduli (DM) of all globular clusters to distance (d), in parsecs:

$$DM = (m - M) - Av = 5 \cdot \log(d) - 5 \quad \Rightarrow \quad d = 10^{(DM+5)/5}\ \text{pc}$$

So, it follows to calculate the cartesian coordinates (x, y, z) of the globular clusters with respect to the Sun, using the galactic coordinates (longitude l and latitude b). The x-axis points to the Galactic Centre, the y-axis points to direction of galactic rotation and the z-axis is perpendicular to the galactic disk, and points in the direction antiparallel to the angular momentum. The conversion of coordinates should be done as follows:

$$x = d \cdot \cos(l) \cdot \cos(b) \quad , \quad y = d \cdot \sin(l) \cdot \cos(b) \quad , \quad z = d \cdot \sin(b)$$

The table below shows the calculated values of d, x, y, and z for all globular clusters.

Name	d (pc)	x (pc)	y (pc)	z (pc)
NGC 6522	7244.36	7226.20	129.29	-496.01
NGC 6401	7585.78	7553.77	455.39	526.52
NGC 6342	7943.28	7800.55	668.47	1341.77
NGC 6553	5248.07	5218.73	479.81	-277.32
NGC 6440	8317.64	8223.94	1116.16	551.39
Ter 12	5248.07	5188.85	762.34	-192.40
VVV-CL 160	6918.31	6809.92	1219.29	36.47
2MASS-GC01	3311.31	3256.16	601.79	5.78
NGC 6517	9120.11	8551.60	2982.17	1073.85
NGC 6402	9120.11	8213.73	3206.36	2330.31
NGC 6712	7244.36	6528.00	3093.30	-545.44
NGC 6426	20892.96	17697.53	9444.44	5840.86
NGC 5466	15848.93	3319.16	3004.35	15203.48
NGC 7089	11481.54	5558.08	7476.05	-6711.34
NGC 288	9120.11	-86.55	47.41	-9119.57
NGC 2298	10000.00	-3966.46	-8755.80	-2757.38
NGC 4590	10471.29	4185.03	-7359.22	6162.41
NGC 4372	5754.40	2919.15	-4859.63	-987.77
NGC 362	8709.64	3150.05	-5133.77	-6291.21
BH 140	4786.30	2611.38	-3995.02	-359.45
NGC 5927	8317.64	6919.31	-4561.75	704.68
Patchick 126	7943.28	7465.49	-2661.12	-530.03
NGC 5897	12589.25	10392.20	-3187.93	6350.49
NGC 6380	9549.93	9393.28	-1625.54	-570.03
Djor 1	10000.00	9973.79	-579.45	-433.40

(b) (20 points) From the given data, estimate the distance from the Sun to the centre of the distribution of globular clusters and its uncertainty.

Solution: In order to estimate the distance from the Sun to the centre of the distribution and its uncertainty, we first need the mean values of each coordinate $\bar{x}$, $\bar{y}$ and $\bar{z}$, as well as the standard deviations in each axis, σ_x, σ_y and σ_z. The calculations of the mean and standard deviation for the x, y, and z coordinates are as follows:

Mean in the three axes:

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i, \quad \bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i, \quad \bar{z} = \frac{1}{N}\sum_{i=1}^{N} z_i$$

$$\bar{x} = 6164.1 \text{ pc}, \quad \bar{y} = -321.3 \text{ pc}, \quad \bar{z} = 434.3 \text{ pc}$$

Standard deviations in the three axes:

$$\sigma_x = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \bar{x})^2}{N-1}}, \quad \sigma_y = \sqrt{\frac{\sum_{i=1}^{N}(y_i - \bar{y})^2}{N-1}}, \quad \sigma_z = \sqrt{\frac{\sum_{i=1}^{N}(z_i - \bar{z})^2}{N-1}}$$

$$\sigma_x = 4057.7\,\text{pc} \quad , \quad \sigma_y = 4196.4\,\text{pc} \quad , \quad \sigma_z = 4682.6\,\text{pc}$$

Consequently, we obtain the uncertainties associated to the mean in each axis:

Uncertainties of means in the three axes:

$$\delta\bar{x} = \frac{\sigma_x}{\sqrt{N}}, \quad \delta\bar{y} = \frac{\sigma_y}{\sqrt{N}}, \quad \delta\bar{z} = \frac{\sigma_z}{\sqrt{N}}$$

$$\delta\bar{x} = 811.54\text{ pc}, \quad \delta\bar{y} = 839.28\text{ pc}, \quad \delta\bar{z} = 936.52\text{ pc}$$

Finally, the distance D, in parsecs, from the Sun to the centre of the distribution of globular clusters is estimated as:

$$D = \sqrt{\bar{x}^2 + \bar{y}^2 + \bar{z}^2}$$

$$D = \sqrt{(6164.1)^2 + (-321.3)^2 + (434.3)^2} = 6187.73\text{ pc}$$

And, to estimate its uncertainty δD, we should perform an error propagation calculation:

$$(\delta D)^2 = \left(\frac{\partial D}{\partial \bar{x}}\right)^2 \cdot (\delta\bar{x})^2 + \left(\frac{\partial D}{\partial \bar{y}}\right)^2 \cdot (\delta\bar{y})^2 + \left(\frac{\partial D}{\partial \bar{z}}\right)^2 \cdot (\delta\bar{z})^2 \Rightarrow$$

$$(\delta D)^2 = \left[\left(\frac{\bar{x}}{D} \cdot \delta\bar{x}\right)^2 + \left(\frac{\bar{y}}{D} \cdot \delta\bar{y}\right)^2 + \left(\frac{\bar{z}}{D} \cdot \delta\bar{z}\right)^2\right] \Rightarrow$$

$$\delta D = \frac{1}{6187.73} \cdot \left[(6164.1 \cdot 811.54)^2 + (-321.3 \cdot 839.28)^2 + (434.3 \cdot 936.52)^2\right]^{1/2} \Rightarrow$$

$$\delta D = 812.28\,\text{pc}$$

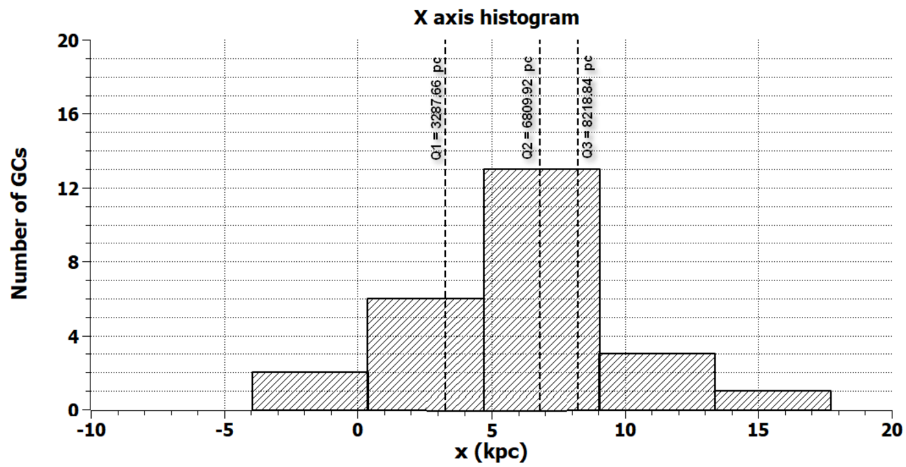
(c) (30 points) To test the validity of Shapley's hypothesis that globular clusters are symmetrically distributed around the Galactic Centre, make histograms with five bins (i.e. sort the data and divide them into five equally-sized intervals) for each of the distributions in the x, y, and z directions.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

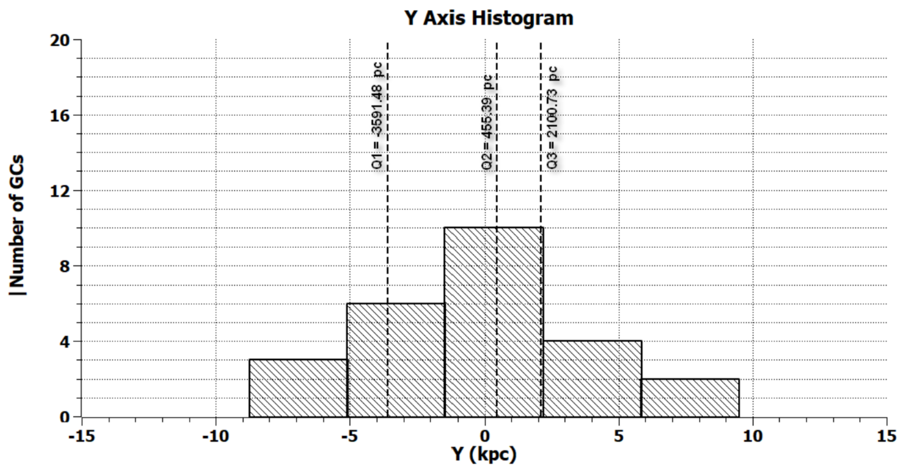
Hint: The three quartiles divide the sorted sample into four sections, each containing 25% of the data, with the second and third sections representing the interquartile range.

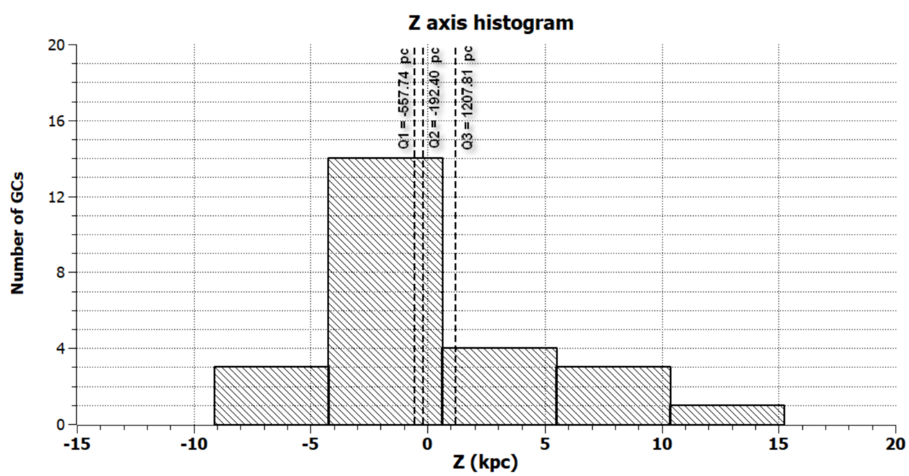
Solution:
The table below shows the calculated values of the quartiles for each distribution.

Axis	Q1 (pc)	Q2 (pc)	Q3 (pc)
x	3287.66	6809.92	8218.84
y	-3591.48	455.39	2100.73
z	-557.74	-192.40	1207.81

The expected histograms for each axis are as follows:


X axis histogram
Number of GCs
x (kpc)
Q1 = 3287.66 pc
Q2 = 6809.92 pc
Q3 = 8218.84 pc
20
16
12
8
4
0
-10
-5
0
5
10
15
20


Y Axis Histogram
Number of GCs
Y (kpc)
Q1 = -3591.48 pc
Q2= 455.39 pc
Q3= 2100.73 pc
20
16
12
8
4
0
-15
-10
-5
0
5
10
15


Z axis histogram
Number of GCs
Z (kpc)
Q1 = -557.74 pc
Q2 = -192.40 pc
Q3 = 1207.81 pc
20
16
12
8
4
0
-15
-10
-5
0
5
10
15
20

(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**
$0.0 \leqslant \Phi \leqslant 0.1$	symmetrical
$0.1 < \Phi \leqslant 0.2$	quasi-symmetrical
$\Phi > 0.2$	asymmetrical

Solution:
F (False). The analysed sample does not follow Shapley's hypothesis, because all distributions are asymmetrical. See table below for the values of calculated symmetry factors for each distribution.

Axis	Φ	**Symmetry Type**
x	0,429	asymmetrical
y	0,422	asymmetrical
z	0,588	asymmetrical