Exercise 1

a) Z-score, Burrow's Delta

As a first step, we compute the term frequencies from the given table:

Profile	A1	A2	A3	Q
tf	233	281	139	140
term1	20	21	3	25
term2	36	100	23	40
term3	90	100	12	45
term4	75	3	67	10
term5	12	57	34	20

Table 1: Term frequencies **tf**

From here we can easily get to the *relative frequencies* if we divide by the tf. Based on those we compute the means μ_i and the standard deviations σ_i:

Profile	A1	A2	A3	μ	σ	Q
rel-tf	1	1	1			
term1	0.0858	0.0747	0.0216	0.06072	0.02804	0.17857
term2	0.1545	0.3559	0.1655	0.22528	0.09245	0.28571
term3	0.3863	0.3559	0.0863	0.27616	0.1348	0.32143
term4	0.3219	0.0107	0.48201	0.27153	0.19569	0.07143
term5	0.0515	0.2028	0.2446	0.16632	0.08296	0.14286

We can now compute the Z-scores for a particular term i in the texts of an author j according to the following formula:

$$\text{Z-score}(w_{ij}) = \frac{\text{rel-tf}_{ij} - \mu_i}{\sigma_i}$$

We get the values from the table below:

Z-scores	A1	A2	A3	Q
term1	0.89579609	0.49981606	-1.39561215	4.20285958
term2	-0.76556006	1.4125559	-0.64699584	0.65367998
term3	0.81684351	0.59136572	-1.40820924	0.33584986
term4	0.25735577	-1.3329727	1.07561693	-1.02252125
term5	-1.38403154	0.44033636	0.94369518	-0.2828038

We can now compute the so-called *Burrows' Delta*. This distance function between two sub-corpora A (known) and Q (doubtful) is given by

$$\Delta(Q, A) = \frac{1}{n} \sum_i^n |\text{Z-score}(w_{iQ}) - \text{Z-score}(w_{iA})|$$

We get the following values for Burrow's Delta:

Delta	$\Delta(Q, A_1)$	$\Delta(Q, A_2)$	$\Delta(Q, A_3)$
term1	3.3070635	3.70304352	5.59847173
term2	1.41924005	0.75887592	1.30067582
term3	0.48099366	0.25551587	1.74405909
term4	1.27987703	0.31045145	2.09813818
term5	1.10122774	0.72314016	1.22649898
$\frac{\sum(...)}{n}$	$\frac{7.58840197}{5} \approx 1.51768039$	$\frac{5.75102692}{5} \approx 1.15020538$	$\frac{11.96784381}{5} \approx 2.39356876$

Therefore Q is most likely written by author A2.

b)

Proceeding in the exact same manner as in task a), we get the following values:

Delta	$\Delta(Q, A_1)$	$\Delta(Q, A_2)$	$\Delta(Q, A_3)$
term1	0.54752793	0.07078951	1.81425415
term2	2.09772653	0.46919608	0.30112584
term3	4.08442023	2.58517358	5.01235305
term4	4.07612375	1.78273079	2.18426853
term5	1.09583832	0.61394747	1.75998275
$\frac{\sum(...)}{n}$	$\frac{11.90163675}{5} \approx 2.38032735$	$\frac{5.52183744}{5} \approx 1.10436749$	$\frac{11.07198433}{5} \approx 2.21439687$

As before, the text Q is most likely written by author A2.