

第四題：資料匿名化 (Anonymize)

問題敘述

在現今數位化的時代，有許許多多的資料，這些資料因為研究或其他需求常常需要公開，但是資料可能涉及個人隱私，因此，資料的公開與個人隱私保護這兩者經常是相互衝突的。資料在公開之前，會先刪去某些內容屬性，以便保護個人隱私，但是這樣做還是不夠的。資料如果太獨特，他人就比較容易就可以從資料來推斷資料屬於何人，因此，發布資料前會先進行匿名化的工作。

為方便起見，本題的每一份資料是 N 個非負整數所形成的序列 $\{A[i]: 0 \leq i \leq N-1\}$ 。對於一個正整數 K ，一份資料中，如果每個出現的數字都至少出現 K 次，則該資料就被稱為滿足「 K -匿名性」。例如， $\{1,5,1,3,3,3,5\}$ 這個序列滿足「2-匿名性」，但不滿足「3-匿名性」。所謂「 K -匿名化」是將一個序列 $\{A[i]: 0 \leq i \leq N-1\}$ 修改成為一個滿足「 K -匿名性」的序列 $\{B[i]: 0 \leq i \leq N-1\}$ ，而「 K -匿名成本」則定義為每一項變動量的總和，也就是 $\sum_{i=0}^{N-1} |A[i] - B[i]|$ 。由於我們希望資料變動越小越好的情況下將資料匿名化，因此，給定序列 A 與整數 K ，我們的目標是計算出最小的 K -匿名成本。

舉例來說， A 序列是 $\{1,4,2,3,3,3,5\}$ 而 $K=2$ ，我們可以將資料修改為 $\{1,5,1,3,3,3,5\}$ ，此時的匿名成本為 2；我們可以很容易看出來，如果修改成 $\{2,5,2,3,3,3,5\}$ ，此時的匿名成本也是 2，但我們找不到成本小於 2 的方法，因此，最低的匿名成本就是 2。對於相同的序列但是 $K=3$ ，將資料 A 修改成 $\{3,4,3,3,3,4,4\}$ 的匿名成本是 5；修改成 $\{2,4,2,2,4,4,4\}$ 或 $\{2,3,2,2,3,3,3\}$ 的匿名成本也都是 5，同時，也找不到比 5 更小成本的修改方式，所以最小匿名成本是 5。

輸入格式

輸入第一行為兩個正整數 N 與 K ，其意義如上所述。第二行有 N 個不超過 10^7 的非負整數，這些整數為資料內容，同一行的數值間以空白隔開。

輸出格式

請輸出最小的 K -匿名成本。請留意，本題的答案可能超過 2^{31} 但不會超過 2^{60} 。

輸入範例 1 7 3 1 4 2 3 3 3 5	輸出範例 1 5
---------------------------------------	--------------------

評分說明

本題共有 4 組測試題組，條件限制如下所示。每一組可有一或多筆測試資料，該組所有測試資料皆需答對才會獲得該組分數。

子任務	分數	額外輸入限制
1	9	$N \leq 100$ ， $K > N/2$ ，也就是將所有資料修改成相同值。
2	12	$K \leq N \leq 1,000$ 。
3	23	$N \leq 50,000$ ， $K \leq 100$ 。
4	56	$K \leq N \leq 300,000$ 。