Multiple Testing Problem

- In some situations we conduct multiple tests
- In the previous example we tested for e.g. multiple cities
- Suppose each test is conducted at level α
- For any one test, the chance of falsely rejecting the null is α
- But the chance of at least one false rejection is much higher
- This is the **multiple testing problem**
- The problem comes up in many data mining applications where we
  sometimes test thousands or millions of hypotheses

Definition

- Consider m hypothesis tests: H_{0i} vs. H_{1i}, $i = 1, \ldots, m$
- We denote with $P_1, \ldots, P_m$ the m p-values for these tests:

Bonferroni Correction

Given p-values $P_1, \ldots, P_m$ reject null hypothesis H_{0i} if:

$$P_i < \frac{\alpha}{m}.$$

Using the Bonferroni correction, the probability of falsely rejecting any null
hypothesis is less thanb or equal to α

- proof

Let R be the event that at least one null hypothesis is falsely rejected. Let
R_i be the event that the ith null hypothesis is falsely rejected:

$$P(R) = P\left(\bigcup_{i=1}^{m} R_i\right) \leq \sum_{i=1}^{m} P(R_i) = \sum_{i=1}^{m} \frac{\alpha}{m} = \alpha$$