Choosing the correct test

Mark Dunning

4 November 2015

Dataset 1 `data1.csv`

Survival times of patients with Ovarian or Breast Cancer.

Is there a difference in survival time between the two diseases?

Dataset 2 `data2.csv`

In the history of data visualization, Florence Nightingale is best remembered for her role as a social activist and her view that statistical data, presented in charts and diagrams, could be used as powerful arguments for medical reform.

After witnessing deplorable sanitary conditions in the Crimea, she wrote several influential texts (Nightingale, 1858, 1859), including polar-area graphs (sometimes called "Coxcombs" or rose diagrams), showing the number of deaths in the Crimean from battle compared to disease or preventable causes that could be reduced by better battlefield nursing care.

Her Diagram of the Causes of Mortality in the Army in the East showed that most of the British soldiers who died during the Crimean War died of sickness rather than of wounds or other causes. It also showed that the death rate was higher in the first year of the war, before a Sanitary Commissioners arrived in March 1855 to improve hygiene in the camps and hospitals.

Do the data support the claim that deaths due to avoidable causes decreased after a change in regime?

Dataset 3: Gene expression `data3.csv`

The expression level of a gene was measured in a breast cancer cohort of ER negative and positive patients.

Is the gene differentially-expressed between ER positive and negative patients?

Dataset4: Sleep Data `data4.csv`

Data which show the effect of two soporific drugs (increase in hours of sleep compared to control) on 10 patients.

Does the drug have an effect on the amount of sleep?

Dataset5: CD4 `data5.csv`

CD4 cells are carried in the blood as part of the human immune system. One of the effects of the HIV virus is that these cells die. The count of CD4 cells is used in determining the onset of full-blown AIDS in a patient. In this study of the effectiveness of a new anti-viral drug on HIV, 20 HIV-positive patients had their CD4 counts recorded and then were put on a course of treatment with this drug. After using the drug for one year, their CD4 counts were again recorded. The aim of the experiment was to show that patients taking the drug had increased CD4 counts which is not generally seen in HIV-positive patients.

Dataset6: Birth Weight `data6.csv`

Risk Factors Associated with Low Infant Birth Weight

Dataset7: Weight gain in Rats `data7.csv`

The data arise from an experiment to study the gain in weight of rats fed on four different diets, distinguished by amount of protein (low and high) and by source of protein (beef and cereal).

Dataset8: Colon cancer

The tumor and the normal counter-part samples were prospectively collected from 9 patients who underwent surgical resection at the INT-MI. Neoplastic samples were obtained from the central area of the neoplasia, avoiding to select necrotic material or transition zones with healthy mucosa. Samples of colonic healthy mucosa were resected at least 20 centimeters far from the neoplasia and distant from the surgical resection margins. Tissue samples were stored in liquid nitrogen until RNA extraction. Total RNA was extracted from 10–20 mg of tumor samples and from 30–40 mg of normal samples.

The gene expression measurements for a paticular gene were extracted from the data.

Is this gene differentially-expressed between tumours and normals?