Biometry 726

Fall 2010

Final Project

Due 8am, Monday 20 December 2010

Human pappilomavirus (HPV) infection is an established causal agent for cervical cancer, and growing evidence shows these same high-risk HPV types, especially HPV16, are causal agents in the etiology of some head and neck cancers, particularly in the oropharynx and tonsillar regions. While the majority of head and neck cancer cases are attributable to alcohol and tobacco abuse, up to 20% of head and neck squamous cell carcinoma (HNSCC) patients present with little to no tobacco and alcohol exposure, suggesting alternative risk factors for the disease. HPV-related oropharyngeal cancer is associated with increased numbers of lifetime sexual partners and increased numbers of oral sex partners; it is therefore hypothesized that oral HPV infection occurs via sexual transmission during orogenital contact. In addition to apparent differences in risk factors, HPV positive (HPV+) HNSCC patients are more responsive to treatment, and generally have better outcomes than their HPV negative (HPV-) counterparts.

Cytokines are small signaling proteins involved in human immune response, and many oncology researchers have investigated their role as cancer biomarkers. A head and neck cancer researcher at MUSC was interested in identifying cytokines associated with HPV-related HNSCC. Using a multiplex immunoassay experimental platform, she measured cytokine concentration levels in the serum of 59 subjects: 17 HPV+ patients, 22 HPV- patients, and 20 control subjects recruited from sleep apnea patients at MUSC. The researcher is interested in knowing the cytokines' associations with disease, whether any are predictive of disease, and quantifying differences in serum concentration levels comparing the different patient subgroups. Additionally, the researcher is interested in knowing if there are any subsets of cytokines that collectively are associated with disease. She considers this a more exploratory aim, but believes underlying data structure may in turn shed light on underlying biologic structure.

The data set `hncadat.csv` is available on the class website. The variable `PID` is a unique patient ID. The variable `DSCAT` is equal to 1 for HPV+ HNSCC patients, 2 for HPV- HNSCC patients, and 3 for control subjects. Additionally, serum concentration levels (measured in picograms per milliliter - pg/ml) of 13 cytokines are provided: interleukin1-β (`IL1b`), IL2, IL4, IL5, IL6, IL7, IL8, IL10, IL12, IL13, IL17, granulocyte colony-stimulating factor (`GCSF`), granulocyte macrophage colony-stimulating factor (`GMCSF`), interferon-γ (`IFNg`), monocyte chemotactic protein-1 (`MCP1`), macrophage inflammatory protein-1β (`MIP1b`), and tumor necrosis factor-α (`TNFa`).

Conduct an appropriate data analysis to address the investigator's aim. You should write a report in the form of a paper with introduction, materials and methods, results and discussion sections. The paper may not exceed 8 pages in length, must be double spaced, and use 11 or 12 point font. An additional document is provided on the class website providing the rubric detailing how your work will be evaluated.