Research methods, alternative exercise for re-sit statistics assignment

# Research methods, alternative exercise for re-sit statistics assignment

# Instructions

You have been provided with a data file in CSV format.

Please answer the questions below.

To do your analysis you may use whatever statistical software you prefer including R.

Please turn in a document that shows your answers to the questions, **and how you got them using statistical software**. For example, print out the log of your R session.

Your analysis will be based on analysis of a research question, which is first introduced.

# Scenario

Based on: Altered Microenvironment Promotes Progression of Preinvasive Breast Cancer: Myoepithelial Expression of αvβ6 Integrin in DCIS Identifies High-risk Patients and Predicts Recurrence, Michael D. Allen, Gareth J. Thomas, Sarah Clark, Marwa M. Dawoud, Sabarinath Vallath, Sarah J. Payne, Jennifer J. Gomm, Sally A. Dreger, Sarah Dickinson, Dylan R. Edwards, Caroline J. Pennington, Ivana Sestak, Jack Cuzick, John F. Marshall, Ian R. Hart and J. Louise Jones. Clin Cancer Res January 15 2014 (20) (2) 344-357; DOI: 10.1158/1078-0432.CCR-13-1504

## Research question

A working hypothesis of the authors was that upregulation of αvβ6 may be a key marker of transition of DCIS to invasive disease.

Like most solid tumours, breast cancers require new blood vessel growth in order to grow more than a few millimetres in diameter. The degree of angiogenesis in association with breast cancer has been shown to be an independent predictor of prognosis. There is a correlation between tumour microvessel density (MVD) and the presence of axillary lymph node, and distant metastases. Some evidence from a limited number of studies also suggests that angiogenesis is associated with more aggressive ductal carcinoma in situ, which is a non-invasive pre-cursor to invasive breast cancer.

To help assess the overall research hypothesis, you will investigate whether expression of avb6 in DCIS is associated with angiogenesis.

## Analysis aims

1. Determine the association between avb6 and grade of DCIS
2. Assess the relationship between avb6 on myoepithelial cells in DCIS and angiogenesis

## Study design

140 patients with histologically confirmed DCIS were identified from the hospital database at St. Bartholomew’s Hospital (N=90) or from the International Breast Cancer Intervention Study II (IBIS-II) clinical trial database (N=50).

## Methods

Myoepithelial cell staining was categorised homogenous, heterogenous or negative in relation to the expression of the avb6 integrin. The pathological grade of the DCIS was scored according to nuclear grade, as used for NHS breast screening program: low, intermediate, high grade. Micro vessel density was assessed using the Chalkley grid method.

# Questions

### Q1: Describe the data type of the variables in your data file

1. **B6**: determines *avb6 positivity*
2. **B6t**: determines *avb6 level* (0 = negative, 1 = heterogeneous, 2 = positive)
3. **CHALKLEY** the main variable of interest: *Chalkley count*
4. **GRADE** histological *grade* of the cancer (coded as 0,1,2)

### Q2: Which variables will you use for analysis to look each of the two main aims above?

### Q3: Determine how many and the percentage samples have been excluded from the data file

### Q4: How many are avb6 positive?

### Q5: Of those avb6 positive, how many are heterogeneous?

### Q6: What is the range of Chalkely count, and the mean Chalkely count?

### Q7: How many are in the different tumour grade categories?

### Q8: Show a boxplot of Chalkely count by avb6 positivity

### Q9: Looking at the box plots or otherwise, what are the approximate median and inter-quartile range for the two groups?

### Q10: Apply a t-test for the difference in mean chalkey count between the avbg positive (homogeneous or heterogenous) vs negative

### Q11: Describe how strong the evidence is for a difference between the groups based on your results?

### Q12: Are the t-test results robust to an assumption of constant variance between the two groups? Give reasons for your answer based on the data.