MATH 204: Principles of Statistics 2

Dr David A. Stephens

Department of Mathematics & Statistics
Room 1235, Burnside Hall

d.stephens@math.mcgill.ca
www.math.mcgill.ca/∼dstephens/204/

January 5, 2007

Textbook: McClave and Sincich (2006), *Statistics* (10th Edition), Chapters 10-14.

Prerequisites: MATH 203 (or equivalent)

Some statistical computing knowledge useful.

Method of Assessment:

- Assignments
- Mid-Term
- Final

Precise breakdown to be confirmed.

Course Objectives

- Extensions of MATH 203 topics to other practical experimental contexts
- Introduction to statistical computation using standard software (SPSS)
- Practice in the use of statistical methods, in particular, hypothesis testing and linear modelling.

Three main sections

I. THE ANALYSIS OF VARIANCE AND DESIGNED EXPERIMENTS

II. LINEAR REGRESSION MODELLING

III. NON-PARAMETRIC TESTING

Typical experimental scenario

- two different groups of subjects
- single observation/measurement made on each subject
- scientific question of interest

  ARE THE TWO GROUPS OF SUBJECTS SIGNIFICANTLY DIFFERENT IN TERMS OF THEIR MEASURED VALUES ?

Example: Pre-Natal Care

Objective: To compare the birthweights of babies in two groups of mothers.

- GROUP A: Received five or fewer pre-natal visits
- GROUP B: Received more than five pre-natal visits

Do the GROUP A babies have significantly different birthweights from those from GROUP B ?

Data: Birthweights (grammes)

- GROUP A: 10 subjects

2164	2600	2184	2080	1820
2496	2184	2080	2184	2576

- GROUP B: 7 subjects

3224	2704	2912	2444	3120
		2912	3848	

First step in analysis: statistical summary

- GROUP A:
  - Sample size: $n_A = 10$
  - Sample mean: $\overline{x}_A = 2236.8$
  - Sample variance: $s_A^2 = 61190.4$
- GROUP B: 7 subjects
  - Sample size: $n_B = 7$
  - Sample mean: $\overline{x}_B = 3023.429$
  - Sample variance: $s_B^2 = 198679.6$

Recall, for data $x_1, \ldots, x_n$

$$\begin{aligned} \overline{x} &= \frac{x_1 + \cdots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i \\ s^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 \end{aligned}$$

$\overline{x}$ measures the "average" of sample
s^2 measures the amount of variability around the average.

In the birthweight example

$$\overline{x}_A = 2236.8 \qquad \overline{x}_B = 3023.429$$

so it appears that Group B birthweights are higher....

... BUT ARE THEY SIGNIFICANTLY HIGHER ?

i.e. is the difference due to chance alone

- sample sizes quite small
- birthweights quite variable

Statistical Testing

We adopt the following procedure to assess the "significance" of the difference between $\overline{x}_A$ and $\overline{x}_B$.

1. Define a *test statistic*, T, that permits comparison of the two groups
2. *Predict* how T will behave assuming that the two groups are **not** significantly different.
3. Compare the *prediction* with what was actually *observed*.

Formally, we

- assume a **Normal distribution** for the data in the two groups

  i.e. $x_{A1}, \ldots, x_{An_A}$ are drawn from a population of birthweights that is well-modelled by a

  $$Normal(\mu_A, \sigma_A^2)$$

  distribution.

  Similarly

  $$x_{B1}, \ldots, x_{Bn_B} \sim Normal(\mu_B, \sigma_B^2)$$

  We might initially assume that

  $$\sigma_A^2 = \sigma_B^2$$