**ZPEM2302**

**Mathematical Tools for Science**

**Two Components:**

**Multivariable Calculus**

**Data Analysis**

**Course Information**

- ☐ Two components of course will be taught concurrently.
- ☐ Two lectures per week for Multivariable Calculus.
- ☐ One two-hour Data Analysis lecture/lab per week (held in a computer room).
- ☐ One tutorial per week (alternating between MVC and DA).
- ☐ Will use Matlab throughout course.
- ☐ Matlab's statistics toolbox used in Data Analysis.

**Topics for Data Analysis Component**

- ☐ Chap 1: Data summaries
- ☐ Chap 2: Population distributions and random variables
- ☐ Chap 3: Sampling and sampling distributions
- ☐ Chap 4: Confidence intervals and hypothesis tests for a single sample
- ☐ Chap 5: Confidence intervals and hypothesis tests for two samples
- ☐ Chap 6: Comparing several means
- ☐ Chap 7: Correlation and regression

**Further Details and Some Examples**

- ☐ Chap 1: Data summaries
  - Types of data, eg. numerical, categorical
  - Picturing data, eg. histograms, stem–and–leaf plots, box plots
  - Measures of location, eg. mean, median, mode
  - Measures of spread, eg. range, quartiles, sample standard deviation

- Chap 2: Population distributions and random variables
  - Discrete and continuous distributions
  - Binomial and Normal distributions
  - Means and variances of distributions
  - Quantile-quantile plots

- Chap 3: Sampling and sampling distributions
  - Parameters and statistics
  - Sampling and surveys
    - Simple random sampling
    - Stratified sampling
    - Bias
  - Designed experiments
    - Replication
    - Randomization
    - Lurking variables
    - Blocking
  - Sampling distributions and the Central Limit Theorem

- ☐ Chap 4: Confidence intervals and hypothesis tests for a single sample

**Hypothesis Testing Example**

A well-designed and safe workplace can contribute greatly to increasing productivity. It is especially important that workers not be asked to perform tasks, such as lifting, that exceed their capabilities. The following data on maximum weight of lift (MWOL) in kg for a frequency of 4 lifts per minute was reported in the article "The Effects of Speed, Frequency, and Load on Measured Hand Forces for a Floor-to-Knuckle Lifting Task" (*Ergonomics* [1992]: 833-843):

25.8 36.6 26.3 21.8 27.2

Suppose that it is reasonable to regard the sample as a random sample from the population of healthy males, aged 18 – 30 years. Do the data suggest that the population mean MWOL exceeds 25? Carry out a test of the relevant hypotheses using a 0.05 significance level ($\alpha = 0.05$)?

- Chap 5: Confidence intervals and hypothesis tests for two samples
  - Two-sample t tests
  - Paired samples

- Chap 6: Comparing several means
  - F distribution
  - One-way ANOVA
  - Two-way ANOVA
  - Interactions
  - Checking assumptions

### One-way ANOVA Example

To test washing powders, 18 batches of clothes were washed — 6 in Omo, 6 in Surf, 6 in Drive. The amount of dirt removed in the wash was measured and the results were recorded in the table. (Three of the results were accidentally flushed down the sink.)

| Omo | Surf | Drive |
|---|---|---|
| 35 | 33 | 35 |
| 29 | 30 | 29 |
| 36 | 35 | 32 |
| 35 | 28 | 33 |
| 39 | 32 | |
| | 31 | |

Is there enough evidence here to indicate that the washing powders differ in effectiveness, or are the data consistent with all washing powders being equally effective?

### Two-way ANOVA Example

To decrease experimental variability, we could sort the clothes into two categories (or blocks), natural and man-made fibres. Suppose we designed an experiment in which there were three batches of clothes made from natural fibres and three from man-made fibres. Within each block, each brand was used to wash one batch. The results were

| | | Block | | |
|---|---|---|---|---|
| | | Natural | Man-made | average |
| Treatment | Omo | 38.6 | 32.6 | 35.6 |
| | Surf | 35.7 | 28.9 | 32.3 |
| | Drive | 30.5 | 26.5 | 28.5 |
| | average | 34.9 | 29.3 | 32.1 |

Do the washing powders differ in effectiveness? Does the dirt removal differ depending on fibre type? Is there an interaction between washing powder and fibre?

- Chap 7: Correlation and regression
  - Correlation coefficient
  - Simple linear regression
  - Inference from regression
  - Quadratic regression
  - Multiple regression
  - Checking assumptions

## Regression Example

The figure below shows the ozone concentration in New York City plotted against both wind speed ($x_1$) and ambient temperature ($x_2$). Does a model of the form $\mu_y = a + b_1 x_1 + b_2 x_2$ appear to be appropriate?