EPID 822 Intermediate Biostatistics Winter 2015

Coordinator: Dr. Paul Peng, Department of Public Health Sciences

Instructors: Course instructors:

<table>
<thead>
<tr>
<th>Name</th>
<th>Office</th>
<th>Phone</th>
<th>Email</th>
</tr>
</thead>
<tbody>
<tr>
<td>Keyue Ding</td>
<td>CTG</td>
<td>77705</td>
<td><a href="mailto:kding@ctg.queensu.ca">kding@ctg.queensu.ca</a></td>
</tr>
<tr>
<td>Michael McIsaac</td>
<td>PHS</td>
<td>77460</td>
<td><a href="mailto:mcisaacm@queensu.ca">mcisaacm@queensu.ca</a></td>
</tr>
<tr>
<td>Paul Peng</td>
<td>CCE</td>
<td>78525</td>
<td><a href="mailto:pengp@queensu.ca">pengp@queensu.ca</a></td>
</tr>
</tbody>
</table>

SAS lab instructors:

<table>
<thead>
<tr>
<th>Name</th>
<th>Office</th>
<th>Phone</th>
<th>Email</th>
</tr>
</thead>
<tbody>
<tr>
<td>Andrew Day</td>
<td>KGH CRC 834998 <a href="mailto:daya@kgh.kari.net">daya@kgh.kari.net</a></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jina Zhang-Salomons</td>
<td>CCE 78562 <a href="mailto:jina.zhang-salomons@krcc.on.ca">jina.zhang-salomons@krcc.on.ca</a></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Teaching assistant:

<table>
<thead>
<tr>
<th>Name</th>
<th>Office</th>
<th>Email</th>
</tr>
</thead>
<tbody>
<tr>
<td>Colleen Webber</td>
<td>PHS</td>
<td><a href="mailto:webberc@queensu.ca">webberc@queensu.ca</a></td>
</tr>
</tbody>
</table>

Lecture Time: Mondays, 11:30am-1:00pm, and Thursdays, 9:30am - 11:00am

Lecture Room: ONT 209

SAS Lab Time: Thursdays, 11:00am - 1:00pm

SAS Lab Room: Jeff 157

Tutorial Time: TBA

Tutorial Room: TBA

Prerequisites: EPID-821

Course Web Site: A course web page at Moodle (https://moodle.queensu.ca/) is available and used throughout the teaching. All course materials and announcements for this course will be posted there.

Objectives: At the end of the course, students should be able to

- understand the use of some complex statistics in the medical literature.
- have a basic understanding of the concepts, logic, and numerical steps involved in the development of some commonly used statistical models in epidemiology.
- apply these statistical models to analyze data for research projects using computer packages and interpret results.

• Cantor A. SAS Survival Analysis Techniques for Medical Research, SAS Institute, NC, 2003.

**Evaluation:** Students will be evaluated in the following aspects:

- Homework. 25%
- SAS Sessions. 15%
- Midterm exam (Feb 24). 26%
- Final examination. 34%

**Please note:**

- The minimum passing grade in Graduate School is 70% for this course.
- The midterm and final exams are closed-book exams. Students are allowed to have one sheet (letter size, double sided) of formulas prepared by themselves and a non-programmable and non-graphical calculator in the exams.
- Late homework and SAS assignments without valid reasons will only receive 75% of marks if handed in before solutions are posted.

**Session Information:** 1. Two way analysis of variance — (Jan 6, 8, McIsaac)

This session extends the ANOVA technique to analyze continuous data with two classifying variables. Parametric methods will be described in detail. The concept of interaction will be presented.

- analysis of matched-pairs
- randomized blocks design (simple repeated measurement design)
- two-way ANOVA
- interactions
- assumptions

Reading (Lecture notes Chapter 1; Rosner 12.6; Fisher 10.3)
2-4. Multiple regression — (Jan 13, 15, 20, 22, 27, 29, TBA)

Multiple regression extends the simple linear regression to allow us to predict the value of a continuous variable from the value of several other variables. Variable selection procedure to choose regression model will be introduced. Analysis of residuals and regression diagnostics will also be emphasized.

- review of simple linear regression and correlation
- linear regression with two independent variables
- multiple and partial correlations
- multiple regression
- regression, ANOVA and dummy variables
- testing hypotheses in multiple regression
- variable selection
- goodness of fit
- analysis of residuals and regression diagnostics
- confounding and interaction in regression
- dummy variables in regression

Reading (Rosner 11.9 - 11.11; Selvin (95), Chapter 4; Kleinbaum et al Chapters 8-12, 13-14)

5. Analysis of Covariance (ANCOVA) — (Feb 3, 5, Ding)

This session will introduce the analysis of covariance (ANCOVA) which combines ANOVA and regression, and incorporates quantitative and categorical predictor (explanatory) variables to form models for a continuous dependent (outcome) variable.

- the ANCOVA model
- test of parallelism
- using dummy variables
- interpretation of the model

Reading (Rosner, p537, 589 - 591; Kleinbaum et al Chapter 15; and Selvin (95), Chapter 5)

6. Analysis of categorical data using logistic model — (Feb 10, 12, Ding)

- simple logistic function and logistic regression
- logistic model and its relation with odds ratio
- fitting logistic models to contingency tables
- estimating and interpreting parameters
- confounding and interaction

Reading (Salvin(96) Chapter 7; Rosner Chapter 10, 13.8)
7. Multiple logistic regression — (Feb 26, Mar 3, Ding)

These sessions emphasize the use of logistic regression to analyze dichotomous response data in relation to multiple independent variables. Selecting and testing appropriate logistic models will be discussed.

- multiple logistic regression
- variable selection
- regression diagnostics and goodness of fit

Reading (Kleinbaum et al Chapter 23; Rosner 13.8; Salvin (96) Chapter 8-9; Selvin (95), Chapter 10)

8. Conditional logistic regression — (Mar 5, 10, Peng)

This session introduces data from matched pair case-control studies and models to analyze such data. We will focus on conditional logistic regression and its computation methods.

- Matched case-control studies
- Conditional logistic regression for matched data

Reading (Kleinbaum/Klein Chapter 11; Rosner 13.9)

9. Poisson regression — (Mar 12, 17, 19, Peng)

This session presents the regression approach to analyze the outcome data being frequency counts (such as disease incidence) or rates (incidence rates) in relation to risk factors and covariates (possible confounding variables).

- Person-years
- Analysis of incidence rates
- Poisson distribution
- Adjusted rates and SMR
- Poisson regression
- Poisson regression and logistic regression

Reading (Selvin (95), Chapter 12)

10-11. Analysis of time to event data — (Mar 24, 26, 31, Apr 2, Peng)

These sessions present methods for the analysis of time to event (censored) data. Life table and Kaplan-Meier estimates and logrank test in comparing survival curves will be introduced. Cox regression model will also be introduced if there is time.

- censored data
- Kaplan-Meier survival curve
- life table analysis
- comparing survival curve using logrank test
• Cox’s regression
• presentation of results

Reading (Rosner Chapter 14; Allison; Cantor)