# **Principles of Biostatistics**

# **Section II: Statistical Modelling with Data**

# **PROGRAM OVERVIEW**

|  |  |
| --- | --- |
| **Date** | May 23 - June 2, 2023 (excluding weekends) |
| **Time** | 1:00-4:00 pm (Mountain Time) |
| **Hours** | 27 hours (non-credit, 13.5 hours lecture + 13.5 hours lab) |
| **Max No. of attendees** | 25 |
| **Location** | Zoom and Slack |
| **Format** | Hybrid of lecture, and hands-on component (lab) |
| **Audience** | All kinds of learners from domestic to international students (undergrad, grad, and post-doc), as well as professionals and adult learners |
| **Certificate** | Students must attend 70% of the sessions (7/10 sessions) in order to receive an electronic certificate of participation. |
| **Course fee** | $300 CAD (per section) |

# **PROGRAM DETAILS**

* **Description:**

**The Principles of Biostatistics, Section II: Statistical Modelling with Data** is offered by [One Health at UCalgary](https://research.ucalgary.ca/one-health) in collaboration with the [University of Calgary Biostatistics Center](https://obrieniph.ucalgary.ca/groups/university-calgary-biostatistics-centre)**.**

This course will serve as an advanced foundation for participants to develop sophisticated models of multiple variables, conduct model comparisons and selections for statistical research problems.

* **Learning Outcomes: At the end of the course, participants will be able to**

1. Model the multiple linear relationships between a response variable (Y) and all explanatory variables (both categorical and numerical variables) with interaction terms. Interpret model parameter estimates, construct confidence intervals for regression coefficients, evaluate model fits, and visualize correlations between a response variable (Y) and all explanatory variables (X) by graphs (scatter plot, residual plot) to assess model validity.
2. Predict the response variable at a certain level of the explanatory variables once the fit model exists.
3. Implement R-software and analyze statistical results for biomedical and veterinary data.

* **Instructors:**
* **Qing (Leah) Li** ([qing.li2@uclagary.ca](mailto:qing.li2@uclagary.ca))
* **Communication**
* You must have an audio and video-enabled laptop, desktop, or mobile device with reliable internet access.
* [Slack](https://slack.com/) is a communication platform, and we will be using it as the main tool for course communications, discussions, announcement, and sharing documents.
* In order to access Slack, you must click[**here**](https://join.slack.com/share/enQtMzI1MTg1MjQ3NzAyNy01NDlmMjFjZWU5ZTA4NDc5NzkxYmExODk3ZGY0ZmM1OWQ0NDU0MjQyODc1YTU0ZTBjN2VjMzVhOTJjNDY2ZWVk)NOW, and join the Biostatistics Course Channel (by providing your email address and full name). This link will expire in 14 days.
* You can use Slack as a desktop app, phone app, or web browser. However, the desktop app is more recommended.
* To join the sessions, please click on the ZOOM LINK below, and enter the passcode. The link is the same for all 10 sessions, and please do NOT share this link with anyone else.

[**ZOOM LINK**](https://ualberta-ca.zoom.us/j/94318536877?pwd=TkFUYjdmOUJtdnU5TmJQSWVldnZ3Zz09) **| Passcode: biostat23**

* **Assessment:**

Assignments will be posted on Slack (our communication tool with students).

Students must attend 70% of the sessions in order to receive the certificate and are encouraged to work on the assignments progressively throughout the course as the relevant material is covered.

* **Lecture Delivery and Hands-on Component:** Course materials will be delivered synchronously through online lectures via Zoom. There will also be a hands-on component in which students will work in **R studio** to gain practical skills in implementing the concepts covered during the lecture. The instructor will spend a portion of the class demonstrating the necessary skills and then will be available to assist students in solving a variety of problems with the software. The hands-on component will primarily cover the application of the material outlined in the tentative lecture schedule below.
* **Required Software:** R and RStudio software (free to download).
* **Topics for Biostatistics:** Material will be catered to a general audience with a large degree of focus on concepts, implementation in R and interpretation of results.
* **Module 4: Applied Statistical Modeling with R**
  + Multiple Linear Regression Model Building (First order model/Model with interaction/Higher order model) for quantitative and qualitative independent variables
  + Estimation and Interpretation of the model parameters
  + Significance Testing (Full model test /Partial test)
  + Model Selections (Stepwise/Backward/Forward Procedure)
  + Model Diagnostics (Linearity/Independence/Normality/Homoscedasticity/Outliers)
  + Model Transformation (Box-Cox Transformation)
  + Model Prediction
  + Transfer-learning (Bonus): extend public models to in-house problems (google Colab, python)
* **Schedule** *(This schedule is subject to change upon the instructors’ assessment of the sessions):*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **#** | **DAY** | **DATE** | **TOPIC** | **Time** |
| 1 | Tuesday | May 23 | First-order models with quantitative independent variables | 1:00-4:00 pm |
| 2 | Wednesday | May 24 | Interaction effects, quantitative and qualitative variables | 1:00-4:00 pm |
| 3 | Thursday | May 25 | Interaction effects and second-order models | 1:00-4:00 pm |
| 4 | Friday | May 26 | Model selection: stepwise regression procedures | 1:00-4:00 pm |
| 5 | Monday | May 29 | Model selection: forward and backward selection procedures | 1:00-4:00 pm |
| 6 | Tuesday | May 30 | Multiple regression diagnostics: verify linearity, independence, and equal variance assumptions. | 1:00-4:00 pm |
| 7 | Wednesday | May 31 | Multiple regression diagnostics: verify normality assumptions and identify multicollinearity and outliers. | 1:00-4:00 pm |
| 8 | Thursday | June 1 | Multiple regression diagnostics: data transformation | 1:00-4:00 pm |
| 9 | Friday | June 2 | Transfer-learning (Bonus): standing on the shoulders of giants. | 1:00-4:00 pm |