

Multivariable Analysis with R



Oct. 14, 2026



4:00 – 5:30 p.m.

Online



**Sherman
Centre**
for Digital Scholarship

scds.ca

scds@mcmaster.ca

A joint initiative of McMaster's
Faculty of Humanities and
McMaster University Libraries



Adjusted Odds Ratios

Variable	OR	95% CI
Age	1.05	1.03 – 1.08
HTN	1.80	1.20 – 2.70
BMI	1.02	1.00 – 1.04
Female	0.75	0.50 – 1.10



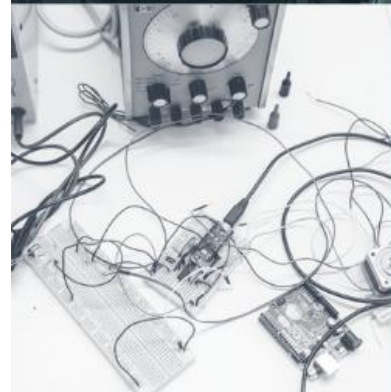
```
model <- glm(  
  died ~ age + bmi + htn + sex,  
  data = clinical,  
  family = binomial)  
summary(model)
```

Multivariable Analysis with R

Presenter:

Shubrandu Sanjoy

PhD Candidate (Health Research Methodology)
Department of Health Research Methods, Evidence,
and Impact, McMaster University





Land Acknowledgement

McMaster University is located on the traditional Territories of the Mississauga and Haudenosaunee Nations, and within the lands protected by the “Dish With One Spoon” wampum agreement.

McMaster's
Communities and
City Libraries

McMaster
University 

 **Sherman
Centre**
for Digital Scholarship

Code of Conduct

The Sherman Centre and the McMaster University Libraries are committed to fostering a supportive and inclusive environment for its presenters and participants.

As a participant in this session, you agree to support and help cultivate an experience that is collaborative, respectful, and inclusive, as well as free of harassment, discrimination, and oppression. We reserve the right to remove participants who exhibit harassing, malicious, or persistently disruptive behaviour.

Please refer to our code of conduct webpage for more information: **scds.ca/code-of-conduct/**

Session Recording and Privacy

This session is being recorded with the intention of being shared publicly via the web for future audiences. In respect of your privacy, participant lists will not be shared outside of this session, nor will question or chat transcripts.

Questions asked via the chat box will be read by the facilitator without identifying you. Note that you may be identifiable when asking a question during the session in an audio or visual format.

Book an Appointment with the DASH Team

Receive help from a member of the DASH team! DASH can assist with the following topics:

- Creating data visualizations, including charts, graphs, and scatter plots
- Figuring out which statistical tests to run (e.g., t-test, chi-square, etc.).
- Analyzing data with software including SPSS, Python, R, SAS, ArcGIS, MATLAB, and Excel
- Choosing which software package to use, including free and open-source software
- Troubleshooting problems related to file formats, data retrieval, and download
- Selecting methodology and type of data analysis to use in a thesis project

Book an appointment: <https://library.mcmaster.ca/services/dash>

Book an Appointment with the RDM Team

McMaster RDM Specialists are available to consult on and assist with a variety of research data management topics including:

- General consulting on data management needs
- Assisting researchers in creating Data Management Plans and in using the Portage Network DMP Assistant tool;
- Advising on issues related to research data storage and backup;
- Facilitating data sharing--whether by consulting on prospective discipline-specific data repositories or assisting with data deposit to general repositories such as McMaster Dataverse and the Federated Research Data Repository.
- Providing guidance on creating data documentation and metadata;

Learn more at: rdm.mcmaster.ca



Certificate Programs

The Sherman Centre for Digital Scholarship Certificate of Attendance

The Sherman Centre's certificate program recognizes attendance at our workshops. It complements degree training, supports the development of critical competencies in data analysis, research data management, and digital scholarship, and formalizes core skills fostered by our workshops.

Participants are invited to attend seven workshops and receive a certificate of attendance. To verify your participation in today's workshop, we will provide a code and additional instructions at the end of the session.

You can learn more about the certificate program at [**scds.ca/certificate-program**](https://scds.ca/certificate-program)

Upcoming Workshops

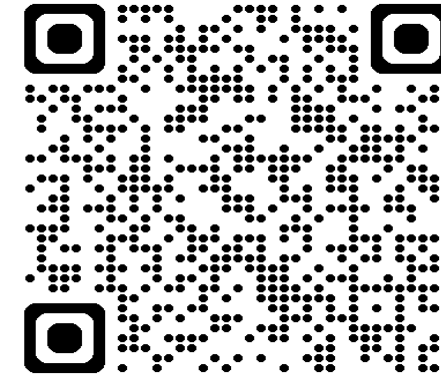
Data Analysis Support Hub

Oct 15: Introduction to Python Programming

Oct 22: Exploring Statistics Canada Data Resources

Nov 4: Introduction to R Programming

Nov 18: Prediction Model Performance for Binary Outcomes in R



Research Data Management

Oct 21 : Best Practices for Managing Data

Nov 17: Setting Up a Research Project with Data from Humans and Preparing for Ethics

Do More with Digital Scholarship

Oct 23: Social Media Scraping and Analysis with Zeeschuimer and 4Cat

scds.ca

scds@mcmaster.ca

A joint initiative of McMaster's
Faculty of Humanities and
McMaster University Libraries



**Sherman
Centre**
for Digital Scholarship

Register for Upcoming Workshops: <https://u.mcmaster.ca/scds-workshops>

SCDS Links

Send SCDS an Email:

scds@mcmaster.ca

Subscribe to our Newsletter:

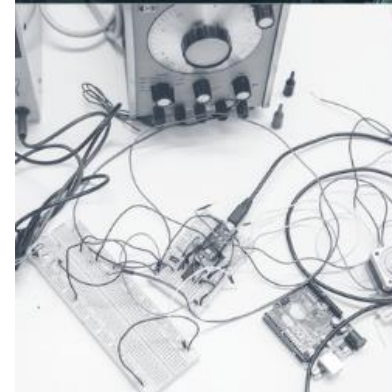
<https://u.mcmaster.ca/sign-up>

Register for a Workshop:

<https://u.mcmaster.ca/scds-workshops>

Schedule a Consultation:

<https://libcal.mcmaster.ca/appointments>



Learning Objectives

By the end of this workshop, you will be able to:

- Prepare and clean research datasets for multivariable analysis in R.
- Fit and interpret multivariable linear and logistic regression models.
- Select appropriate variables and develop multivariable models using practical modeling strategies.

What Is Multivariable Analysis?

Multivariable analysis examines the relationship between one outcome and two or more predictor variables.

- Estimates the association between each predictor and the outcome.
- Accounts for other variables included in the model.
- Helps address potential confounding.
- Is widely used in clinical and epidemiological research.

Clinical example: Is hypertension associated with mortality after adjusting for age, sex, and BMI?

Why use multivariable regression?

- Explore how several predictors relate to an outcome
- Estimate associations while accounting for selected confounders
- Understand how estimates change after adjustment
- Support transparent, reproducible research

Common Multivariable Regression Models

- Linear regression
- Binary logistic regression
- Ordinal logistic regression
- Multinomial logistic regression
- Poisson regression
- Negative binomial regression
- Cox proportional hazards regression

Why Do We Need Different Regression Models?

- **Type of outcome variable** – Continuous, binary, count, or time-to-event.
- **Statistical assumptions** – Different models have different assumptions.
- **Data structure** – Independent, repeated, or clustered observations.
- **Research objectives** – Estimating associations, explaining outcomes, or predicting outcomes.
- **Model performance** – Different models may fit the data differently.
- **Model complexity** – Some models require more parameters and larger sample sizes.

Multicollinearity

What is it?

- Occurs when predictor variables are highly correlated.
- Can lead to unstable coefficients and larger standard errors.
- Makes individual predictor effects difficult to interpret.

Before fitting the model

- Check correlations between predictors.
- Identify highly correlated variables.

After fitting the model

- Assess the Variance Inflation Factor (VIF).
- $VIF = 1$: No variance inflation
- VIF 1–5: Generally acceptable
- $VIF > 5$: Potential concern

Choosing the Appropriate Regression Model

Feature	Logistic regression	Linear regression
Outcome	Binary	Continuous
Clinical example	Death: Yes/No	Healthcare cost
Effect measure	Odds ratio	Regression coefficient
R function	glm()	lm()

What Is Logistic Regression?

Logistic regression examines the association between predictor variables and a binary outcome. Examples of binary outcomes:

- Mortality: Yes/No
- Hospital readmission: Yes/No
- Disease occurrence: Yes/No
- It estimates associations on the log-odds scale, which can be expressed as odds ratios.

Understanding Odds Ratios

An odds ratio (OR) describes how the odds of an outcome differ between groups or with a change in a predictor.

- $OR = 1$: No difference in odds.
- $OR > 1$: Higher odds.
- $OR < 1$: Lower odds.
- Example: An OR of 1.50 means 50% higher odds of the outcome

Unadjusted vs Adjusted Logistic Regression

- Adjusted estimates may differ from unadjusted estimates because other patient characteristics are included.
- Unadjusted model: Mortality ~ Hypertension. One predictor; no additional adjustment.
- Adjusted model: Mortality ~ Hypertension + Age + Sex + BMI. Estimates the association after accounting for other selected predictors.

Live R Demonstration: Logistic Regression

- Research question: Is hypertension associated with mortality after adjusting for age, sex, and BMI?
- Fit an unadjusted logistic regression model.
- Fit a multivariable logistic regression model.
- Calculate adjusted ORs and 95% CIs.
- Compare results and interpret the findings.
- Backward and Forward for variables selection

What Is Linear Regression?

Linear regression examines the relationship between one or more predictors and a continuous outcome. Examples of continuous outcomes:

- Healthcare cost
- Systolic blood pressure
- Length of hospital stay
- Quality-of-life score

The regression coefficient represents the expected difference in the mean outcome associated with a predictor, holding other included variables constant.

Simple vs Multiple Linear Regression

Multiple linear regression estimates the association of each predictor while accounting for the other variables in the model.

- Simple linear regression: Healthcare Cost $\sim$ Age (one predictor).
- Multiple linear regression: Healthcare Cost $\sim$ Age + Sex + BMI + HTN (two or more predictors).

Live R Demonstration: Linear Regression

- Research question: Are age, sex, BMI, and hypertension associated with healthcare cost?
- Fit a simple linear regression model.
- Fit a multiple linear regression model.
- Interpret coefficients and 95% CIs.
- Predict an outcome for an example patient.

Reporting Multivariable Regression Results

Component	Logistic regression	Linear regression
Main estimate	Adjusted OR	Adjusted coefficient
Uncertainty	95% CI	95% CI
Significance	p-value	p-value
Variables	Predictors and reference groups	Predictors and units
Model information	Sample size, model checks	Sample size, model checks

Confounding in Multivariable Analysis

Confounding occurs when another variable is associated with both the predictor and the outcome and distorts the association of interest. Older patients may be more likely to have hypertension and may also have higher mortality.

- Predictor: Hypertension
- Outcome: Mortality
- Potential confounder: Age
- Key message: Include relevant confounders based on clinical knowledge and the research question, rather than relying only on p-values.

Key Takeaways and Hands-on Practice

- Next: Let's move to R!
- Choose regression methods based on the outcome type.
- Use clinical knowledge to guide predictor selection.
- Interpret adjusted estimates with their 95% confidence intervals.
- Report results clearly and reproducibly.
- Participants will fit logistic and linear regression models using the example dataset, examine results, and practice interpreting adjusted associations.

Tell us what you think | Verify your attendance

Provide feedback on this session and verify your participation at
<https://u.mcmaster.ca/verification>

This session's verification code:
MULTI-R-26