Statistical Inference with R

Linear & Logistic Regression

•••

GW Libraries Workshop
Dan Kerchner ~ February 19, 2021

go.gwu.edu/rstats

Logistics

- Just speak up OR use the Zoom chat
- Kiri can provide individual help
- Plan for 1 brief ☕ break

Super-Brief Review of Inference for Regression

High-Level Objective

Population	Sample
Parameter	Statistic

Regression Model

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n$$

<u>Interpretation</u>

β_0 = Mean Y when X_i values are 0

β_i = Mean change in Y for a 1-point increase in X_i, adjusting for other X variables

Correlation between dependent & independent variables

Pearson correlation (ϱ) measures strength and direction (+/-) of linear association between X_i and Y. Ranges from -1 to 1.

If relationship looks non-linear (but monotonic) then **Spearman** correlation should be considered.

$H_0: \varrho = 0$
$H_1: \varrho \neq 0$

Valid if joint distribution of X, Y is bivariate normal

Linear Regression Model

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n$$

where Y is a <u>continuous</u> outcome

<u>Interpretation</u>

β_0 = Mean Y when X_i values are 0

β_i = Mean change in Y for a 1-point increase in X_i, adjusting for other X variables

Linear Regression Assumptions

- Observations are independent
- Linearity
- Homoscedasticity
- Normality

GLMs - Generalized Linear Models

$$g(\mu) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n$$

where $\mu = E(Y)$

Common link functions

$$g(\mu) = \mu$$

$$g(\mu) = \log(\mu)$$

$$g(\mu) = \log(\mu \,/\, (1 - \mu))$$

GLMs: Logistic Regression Model

$$\log\Big(\underbrace{\frac{p}{1-p}}_{\text{odds}}\Big) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_n X_n$$

where p = Probability of Y = 1

1-p = Probability of Y = 0

Interpretation

β_i = log OR (Odds Ratio) for having $Y = 1$ for a 1-point increase in X_i, *adjusting for other predictors*

e^{β_i} = OR for having $Y = 1$ for a 1-point increase in X_i

Logistic Regression Assumptions

- Binary outcome (0, 1)
- Each predictor is linearly related to the log odds of the outcome

Inference for Regression Modeling

Confidence Interval

$95\%\text{CI for } \beta = (0.44, 0.49)$

Hypothesis Testing

H_0: $\beta = \beta_0 \leftarrow$ Null Hypothesis
H_A: $\beta \neq \beta_0 \leftarrow$ Alternative Hypothesis

p-value: Chance that we are rejecting H_0 when we should not be

Goals

Today's Goal

- Learn to use R to read in data and conduct regression analysis and associated inference tests
  - Checking assumptions
  - Visualizing
  - Computing p-values, regression coefficients, confidence intervals, and odds ratios (for logistic models)

Today: 2 Scenarios

- Linear Regression (continuous outcome)
  - Single variable (1 continuous)
  - Multivariable (1 continuous, 1 categorical)
- Logistic Regression (categorical outcome)
  - Multivariable (continuous and categorical)

Today's Data Set: Framingham Heart Study

- framinghamheartstudy.org
- Long-term prospective study of the etiology of cardiovascular disease among a population of subjects in Framingham, MA
- Began in 1948 with 5,209 subjects
- Is the source of the term "risk factor"
- Over 3,000 peer-reviewed papers published based on this study
- Participants were each followed for a total of 24 years for cardiovascular events (heart attack, stroke, death, etc.)

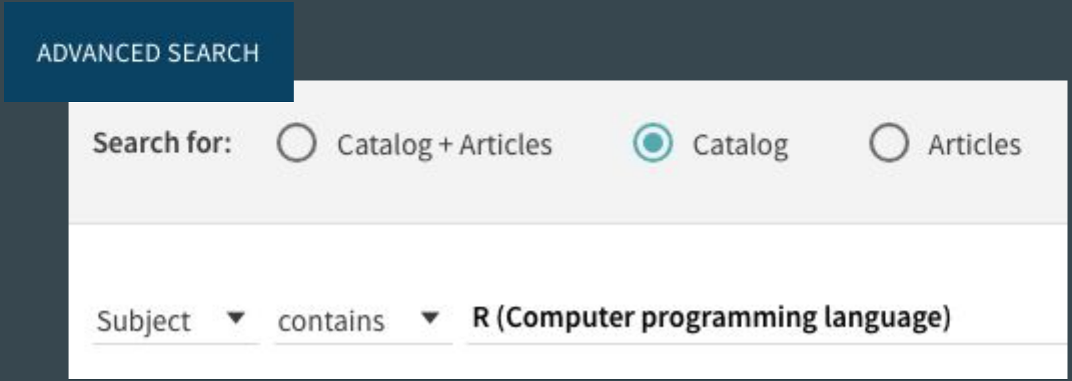
Some Handy R Links

Tutorials

- RStudio R paths: education.rstudio.com/learn/
- Data Carpentry & Software Carpentry:
  - datacarpentry.org
  - software-carpentry.org
- Linkedin Learning @ GW: go.gwu.edu/linkedinlearning
- r-tutor.com/r-introduction & r-tutor.com/elementary-statistics
- UCLA Data Analysis Examples: stats.idre.ucla.edu/other/dae/
- R Graph Gallery (w/code): r-graph-gallery.com

Books you can access for free

- Free books online - Hadley Wickham:
  - R for Data Science r4ds.had.co.nz
  - Advanced R adv-r.hadley.nz/
- Through your GW library privileges:



Reference Links

- R language (CRAN): r-project.org
- R search engine: rseek.org
- rstudio.com
  - Cheat Sheets! rstudio.com/resources/cheatsheets
- stackoverflow.com

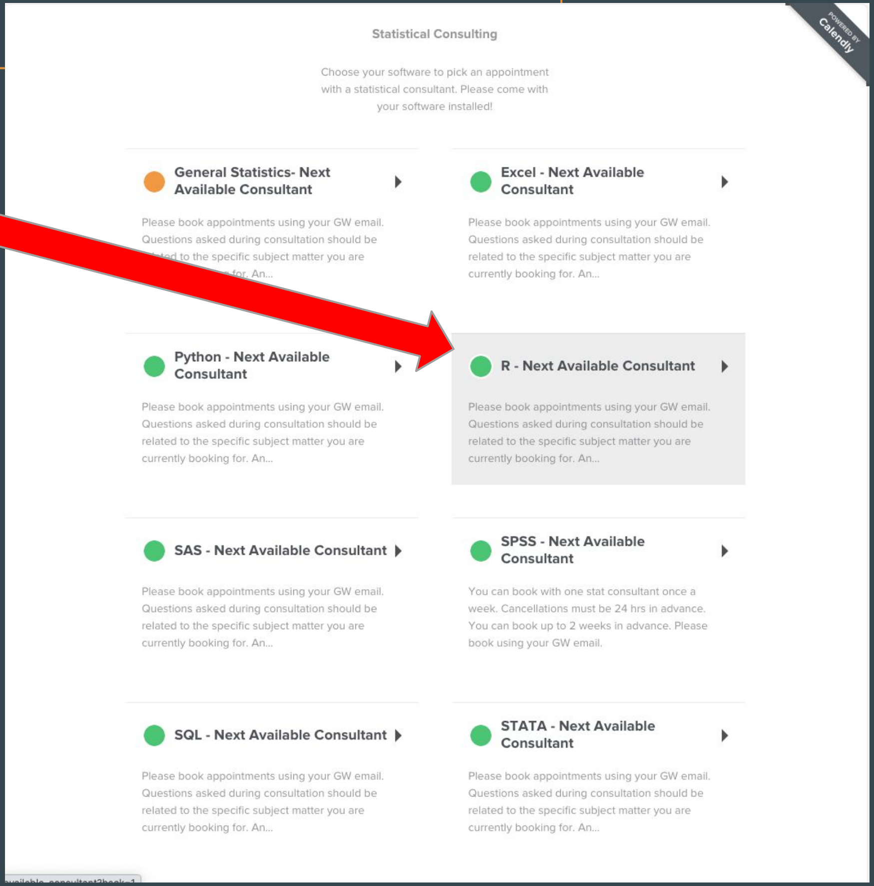
Statistics+R help @ GW

R-Statistics Appointments:
calendly.com/statistical-consulting-gw

Also...

Appointments with me:
calendly.com/kerchner

Coding consultations (Python, git, etc.):
calendly.com/gwul-coding/



Thanks!

Dan Kerchner kerchner@gwu.edu

Disovankiri Boung dboung@gwu.edu