

# DSC 385 - Project 4 Causal Inference

## Overview

Right heart catheterization (RHC) is a diagnostic procedure for directly measuring cardiac function in critically ill patients. In an influential study, Connors et al. (1996) studied the effectiveness of RHC with an observational study design. The study collected data on 5735 hospitalized adult patients; 2184 of them are assigned to the experimental treatment, receipt of RHC within 24 hours of admission, and the remaining 3551 assigned to the control condition. The outcome was death at 30 days after admission to the hospital. The goal is to assess the causal effect of RHC (the treatment) on the binary outcome, death at 30 days after admission.

The dataset provided here is a cleaned version of the original dataset. The treatment variable in the dataset is **swang1** and the outcome variable is **death**. To simplify the analysis, we have restricted the data to 20 covariates that have been identified as the top confounders in an ad-hoc exploratory analysis.

Below is a table describing each of the variables in the dataset:

Variable Name	Interpretation
age	Age in years
sex	Male / Female
cat1	Primary Diagnosis: COPD, Multiple Organ System Failure (MOSF) w/Sepsis, MOSF w/Malignancy, CHF, Coma, Cirrhosis, Lung Cancer, Colon Cancer
cat2	Secondary Diagnosis: MOSF w/Sepsis, Coma, MOSF w/Malignancy, Lung Cancer, Cirrhosis, Colon Cancer
ca	yes = Cancer localized, no = metastatic
death	Death within 30 days after hospital admission
paf1	PaO2 / F102 ratio
wtkilo1	Weight

Variable Name	Interpretation
surv2mdl	Estimate of probability of surviving 2 months
dementhx	Dementia, stroke or cerebral infarct, Parkinsons's disease
gastr	Gastrointestinal diagnosis
wblcl	White blood count
temp1	Temperature
das2d3pc	DASI-Duke Activity Status Index
chfhx	Congestive heart failure
hema	Hematological diagnosis
chrpulhx	Chronic pulmonary disease, severe pulmonary disease
cardiohx	Cardiovascular symptoms
meta	Metabolic diagnosis

The goal of the project is to try to determine the effect of right heart catheterization on 30-day mortality using data from an observational study. In doing so, you will use propensity score matching to reduce the potential impact from confounding variables.

## Getting the Data

The data can be obtained from the following GitHub repository:

<https://github.com/Principles-of-Data-Science/Project4>

You can pull the GitHub repository into RStudio as follows:

1. Click on File > New Project...
2. Click on "Version Control"
3. Click on "Git"
4. Paste in the URL **<https://github.com/Principles-of-Data-Science/Project4>** into the "Repository URL" field
5. Type in the "Project directory name" if needed
6. Set the directory if you don't want to use the default
7. Click "Create Project"

## The Report

The text of your report will provide a narrative structure around your code and outputs with R Quarto. Answers without supporting code will not receive credit and outputs without comments will not receive credit either. Write full sentences to describe your findings. All code contained in your final project document must work correctly (render early, render often)!

**The report template provided in the GitHub repository contains prompts/questions that you will need to answer. Please follow the prompts in the template and answer all of the questions there.**

## Formatting

- Create the report using R Quarto knitted to a PDF file, with headers for each section and each question answered;
- Include comments describing your R code;
- Include any references (datasets, context), if needed.
- Please do not print out very large objects that require multiple pages; only print out what is needed to explain your reasoning for a question.
- It is extremely important that you **select pages** when submitting on Gradescope (see more below). Points will be taken off if you do not select the appropriate pages for each question in the Gradescope outline.

## Submission of the Report to Gradescope

This project report will be submitted on Gradescope for grading. Gradescope is a tool that enables the teaching staff to efficiently grade assignments like this one according to a defined rubric. **You will not be submitting this project on Canvas.** Anything submitted to Canvas will be ignored.

If you have never submitted anything to the Gradescope web site, please watch this [video demonstration of how to do so](#).

To submit your project report, please follow these steps:

1. First render your project report into a PDF file. This can be done by either rendering directly to PDF in RStudio or by rendering to an HTML file and then “Printing” to a PDF file. Either way, **you must have a PDF file to submit to Gradescope.**
2. Go to the course Canvas page and click on the “Gradescope” link on the navigation bar on the left hand side.

3. When the Gradescope page loads, click on the assignment titled “Project 3: Prediction Modeling”.
4. You should be prompted with a window allowing you to submit a PDF file of your assignment.
5. After uploading your PDF file, you will be prompted to select pages of your PDF file that correspond to questions in the “Question Outline”. Please make sure to do this carefully, as it is essential for allowing us to grade your project efficiently.
6. After selecting the pages, submit the assignment.

### **Late Policy**

As per the Syllabus, projects will not be accepted late. There are no exceptions; please do not contact the instructor or TA to request an exception.