

Using uwIntroStats
Authors: Brian D. Williamson and Scott S. Emerson, M.D., Ph.D.
University of Washington Department of Biostatistics

Contents

1	Introduction	2
2	Preparing uwIntroStats	2
3	Descriptive Statistics	2
4	Analyses	2
4.1	Analysis 1	3

1 Introduction

2 Preparing uwIntroStats

Before we can dive in and run any analyses, we first need to install the package. This is done via

```
> install.packages("uwIntroStats")
```

Regardless of the graphical user interface (GUI) that you are using, R will prompt you to select a CRAN mirror. It is essentially asking you where you want to download the package files from. Select the mirror closest to you - for us at the University of Washington it is **WA(1)** or the Fred Hutchinson Cancer Research Center (FHCRC) - and the package will download and say that it has installed. Now each time we open a new R session (whether that is at the command line, a new RGui window, or a new RStudio window) we need to load the package for use.

The **uwIntroStats** package relies on five other packages. These other packages provide key functions that the **uwIntroStats** package uses or adds functionality to. We must install these packages like we did above if we have not installed them previously, and then load **uwIntroStats**. While these packages do not need to be loaded every time (in fact, some are only used for specific functions) it is good practice to load them for the R session where you need to use **uwIntroStats**.

```
> library(Exact)
> library(geepack)
> library(plyr)
> library(sandwich)
> library(survival)
> library(uwIntroStats)
```

Last, we load the data, **mri** that we will be using throughout this document. Information about the dataset can be found at <http://www.emersonstatistics.com/datasets/mri.pdf>. Since the data is part of the package, we can load it via

```
> data(mri)
```

The **uwIntroStats** package should be used for descriptive statistics, basic plotting (like scatterplots and boxplots), and regression analyses. The following sections will go through examples of these tasks, in addition to pointing out how our package differs from base R and other existing packages. We will assume familiarity with basic data manipulation and statistical tasks (for a refresher, see <http://www.emersonstatistics.com/GeneralMaterials/R/IntroToR.pdf>).

3 Descriptive Statistics

4 Analyses

First, a disclaimer. All of the following analyses are for teaching. This is not the way that the authors recommend performing an analysis. We do emphasize heavily the need to prespecify all analyses in any setting, in order to have correct reproducibility and error rates. The two functions we will use in these analyses are **regress** and **lincom**. To learn more about either of these functions, type

```
> ?regress
> ?lincom
```

From these help files, we learn that the minimum we need to enter into **regress** is a functional, a formula, and a dataset. A functional takes an object and returns a value; for instance, the mean is a functional because it takes a distribution and returns the mean. The allowed functionals for **regress** are:

Functional	Type of Regression
"mean"	Linear Regression
"geometric mean"	Linear Regression on logarithmically transformed Y
"odds"	Logistic Regression
"rate"	Poisson Regression
"hazard"	Proportional Hazards Regression

4.1 Analysis 1

A natural question is, “Do males have lower cerebral atrophy score than females”? To answer this question, we run a linear regression of atrophy on the variable male:

```
> regress("mean", atrophy~age, data=mri)
Call:
regress(fnctl = "mean", formula = atrophy ~ age, data = mri)

Residuals:
    Min       1Q   Median       3Q      Max
-36.870  -8.589  -0.870   7.666  51.203

Coefficients:
              Estimate Naive SE Robust SE    95%L    95%H      F stat    df Pr(>F)
[1] Intercept      -16.06    6.256    6.701   -29.22   -2.907      5.75  1    0.0168
[2] age              0.6980  0.08368  0.09002    0.5213    0.8747     60.12  1 < 0.00005

Residual standard error: 12.36 on 733 degrees of freedom
Multiple R-squared:  0.08669,    Adjusted R-squared:  0.08545
F-statistic: 60.12 on 1 and 733 DF,  p-value: 2.988e-14
```