MTH142 Intro to Stats:: **CHEAT SHEET**

Getting started

Signing into RStudio

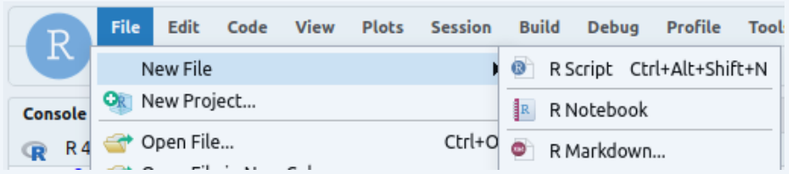
http://hcc-rstudiosrv1.hcc.edu:8787/

Username: Your HCC email with @hcc.edu
Password: Your HCC password

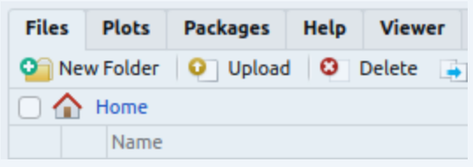
Downloading MTH142Rtutorials:
```
install.packages("remotes")
remotes::install_git("https://github.com/Nics-Github/MTH142Rtutorials")
```
Select 3 to install no updates.

.Rmd files

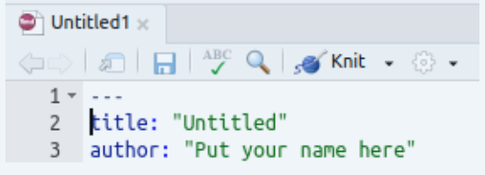
New .rmd File>New File>R Markdown...



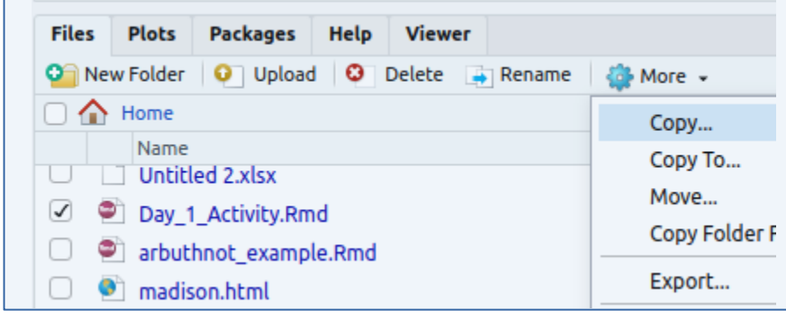
Uploading a file Lower Right quadrant click upload



Knitting a file click the knit button and save.



Exporting a file select file in lower right quadrant click **more** then **Export...**



Screen Shots

Mac: Shift + Command + 3

Windows: + Shift + S

Chromebook: Shift + Ctrl +

Libraries

```
library(MTH142Rtutorials)
library(mosaic)
library(openintro)
library(tidyverse)
```

View Data Sets

cars93 is the data set
glimpse(cars93)
head(cars93)
First six rows **(**tail() works too)
view(cars93)
makes window for all data
names(cars93)
show variable names only

Summary Statistics

favstats(~weight,data=cars93)
outputs mean,sd, and five number summary
fivenum(~weight,data= cars93)
Just the five number summary
summary(cars93)
summarizes all the variables
tally(~cars93)
counts the observations
tally(~type,data=cars93)
counts the observation by subgroup

Variables

a_variable <- c(1,2,3)
takes the vector 1,2,3 and saves it as a variable with the assignment operator

Math Type

Use **$$** to enclose math type.

μ	σ
$\bar{x}$	$\hat{p}$
H_o	α
$\ne$	$\approx.$
$\sim$	

Graphs

box plots
```
gf_boxplot(~weight,data=cars93)
gf_boxplot(type~weight,data=cars93)
```

histograms
```
gf_histogram(~weight,data=cars93)
```

bar
```
gf_bar(~type,data=cars93)
```

scatter plots
```
gf_point(weight~mpg_city,data=cars93)
```

lines
```
gf_point(hp~mpg) %>%
  gf_lm(hp~mpg)
```

titles and labels
```
gf_histogram(~weight,data=cars93,xab="weig
ht",ylab="count",title="weight in pounds")
```

Confidence Intervals

From Data

prop.test(~type,success="small",data=cars93,conf.level=0.95,correct=FALSE)
produces a 95% confidence interval for the proportion of cars that were small in 1993.

t.test(~weight,data=cars93)
produces a 95% confidence interval for the mean weights of cars sold in 1993.

From Statistics

prop.test(20,60,conf.level=0.90,correct=FALSE)
produces a 90% confidence interval for the population proportion given 20 successes in 60 trials.

For mean use formula below
For the confidence interval of the mean from statistics we use the following formula:

$$\bar{x} \pm t^* SE$$

List of Standard Error formulas

$$SE = \sqrt{\frac{p(1-p)}{n}}$$ SE for a single proportion

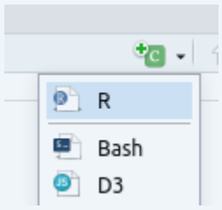
$$SE = \frac{\sigma}{\sqrt{n}} \approx \frac{s}{\sqrt{n}}$$ SE for a single mean or paired means.

$$SE = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$ SE for two means (not paired)

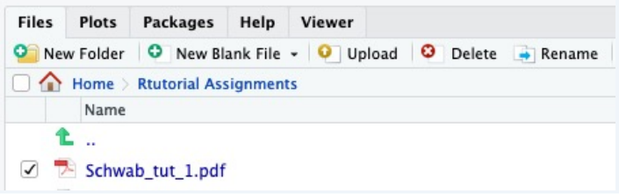
MTH142 Intro to Stats:: **CHEAT SHEET**

R Markdown cont.

New chunk click the green c+ and select r



Rename a file
Click the box next to the file and chose rename.



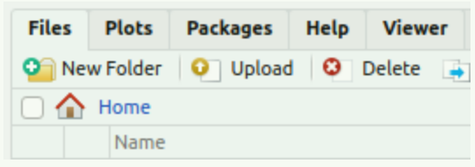
Uploading Data

1. Copy and paste data into a spreadsheet.

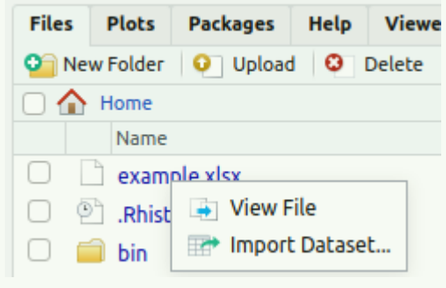
2. Variables are column headers (no spaces in variable name)

3. Save as an .xls file to your computer (no spaces in file names)

4. Upload .xls file



5. Click on the file you uploaded > Import Dataset



normal distribution

xpnorm(1,mean=0,sd=1)
Outputs the probability if z=1 from a standard normal dist.
xqnorm(0.01,mean=0, sd=1)
Outputs the z-score if probability=0.01

$$z^* = \frac{\hat{p} - p}{(SE)}$$

This is the formula for a z*-score for a single proportion.

binomial distribution

dbinom(x=4, size=5,prob=0.25)
Outputs the probability of 4 successes in a trial of size 5 with a probability of success of 0.25

pbinom(x=2,size=5,prob=0.25)
Outputs the probability of 0,1,or 2 successes in a trial of size 5 with a probability of success of 0.25

student t distribution

xpt(1,df=11)
outputs the p-value if t=1 and 11 degrees of freedom
xqt(0.01,df=11)
outputs the t-score if probability=0.01

$$t^* = \frac{\bar{x} - \mu}{SE}$$

This is the formula for a t*-score for a single mean.

help

To get more info on a dataset, load a library then use ?
library(MTH142Rtutorials)
?MTH142Rtutorials

Hypothesis Tests

Proportions
prop.test(~type,success="small",p=0.5 , data=cars93,alternative="t", correct=FALSE)
Outputs a two tailed hypothesis test with parameter p=0.5.
prop.test(x=20, n=60, p=0.5, correct=FALSE)
Outputs the results of hypothesis test with 20 successes in 60 trials
binom.test(~type,success="small" , data=cars93, conf.level=0.90, ci.method="Wald")
Outputs a 90% confidence interval. Can use prop.test if appropriate.

Single Mean
t.test(~weight,data=cars93, mu=2000, alternative="g")
Outputs of a right tailed hypothesis test with parameter mu=2000.

Difference of Means
t.test(time_hrs~division,data=nyc_marathon, alternative="t")
Computes the difference of men and women winning times in NYC marathon

ANOVA
cars_anova<-aov(weight~type,data=cars93)
performs the analysis of variance and saves it as a variable "cars_anova"
anova(cars_anova)
outputs the summary of the analysis

Linear Regression
cars_linear<-lm(price~weight,data=cars93)
performs the linear regression and saves it as a variable "cars_linear"
msummary(cars_linear)
outputs the linear summary of the regression

 Nic Schwab • nschwab@hcc.edu • Updated: Dec 29, 2021