Psi Chi Honor Society & Psychology Club

 / **Workshop**

Correlations, T-tests, Linear Regression, Logistic Regression
April 6th, 2021

https://joshmgarciaa.github.io/PsiChiDASW1/WS2.html

jamovi download link

- https://www.jamovi.org/download.html

R & R Studio download links

1. https://cran.microsoft.com/
2. https://rstudio.com/products/rstudio/download/
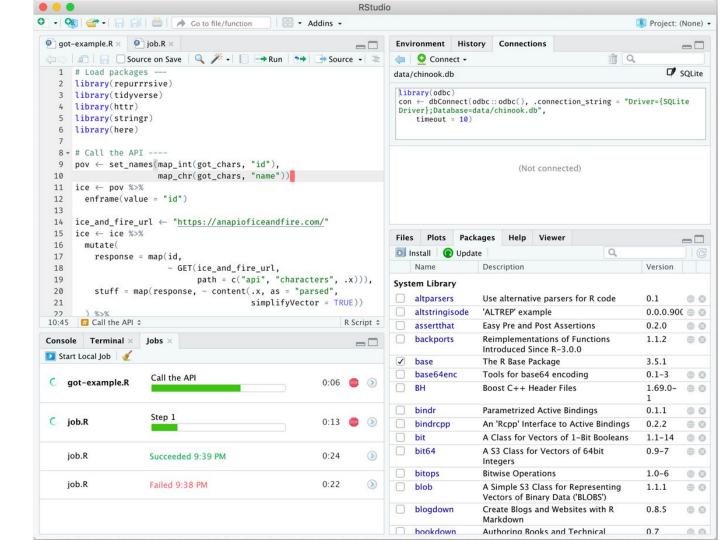
You need both R and R Studio

Agenda

1. Correlations
2. T-tests
3. Linear Regression
4. Logistic Regression

R / R Studio

R is an integrated suite of software facilities for data manipulation, calculation and graphical display.

R packages

"Packages are the fundamental units of reproducible R code. They include reusable R functions, the documentation that describes how to use them, and sample data." (Wickham and Bryan, 2021)

They are like collections of formulas where you can input your data and select specific options to run analyses.

Example code for loading packages:

For more info:
https://r-pkgs.org/index.html

```
library(jtools)

library(jmv)

library(tidyverse)
```

jamovi

ANALYSES

jamovi provides a complete suite of analyses for (not just) the social sciences; t-tests, ANOVAs, correlation and regression, non-parametric tests, contingency tables, reliability and factor analysis. Need more analyses? then see the jamovi library – a library of additional analyses contributed by experts in their field.

STATISTICAL SPREADSHEET

jamovi is a fully functional spreadsheet, immediately familiar to anyone. Enter, copy/paste data, filter rows, compute new values, perform transforms across many columns at once – jamovi provides a streamlined spreadsheet experience, optimised for statistical data.

R SYNTAX

Love R? Check out jamovi's "syntax mode", where the underlying R syntax for each analysis is made available. Just copy and paste this into R for a seamless transition. Alternatively, run R code directly inside jamovi with the Rj Editor.

How are R and Jamovi connected?
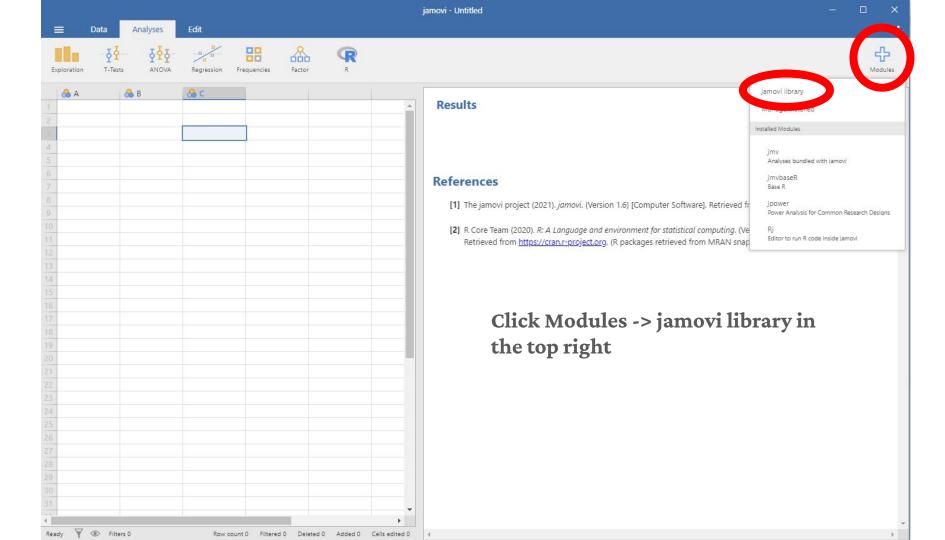
JMV R PACKAGE
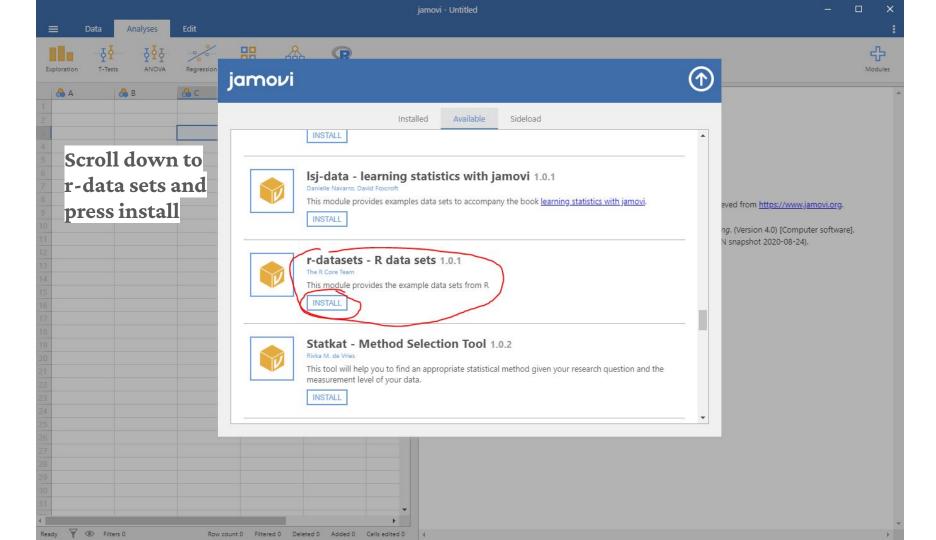
jmv is the jamovi R package. All the analyses included with jamovi are available from within R using this package. For examples on how to use jmv, jamovi can be placed in 'syntax mode' (available from the top right menu). Syntax mode produces the R syntax required to reproduce jamovi analyses in R.

jmv is available from CRAN here, and can be installed in R with:

```
install.packages('jmv')
```

Results

References

[1] The jamovi project (2021). *jamovi*. (Version 1.6) [Computer Software]. Retrieved fr

[2] R Core Team (2020). *R: A Language and environment for statistical computing*. (Ve
Retrieved from https://cran.r-project.org. (R packages retrieved from MRAN snap

Click Modules -> jamovi library in the top right

jamovi - Untitled

Data | Analyses | Edit

Exploration | T-Tests | ANOVA | Regression | | | | Modules

| | A | | B | | C | |

Scroll down to r-data sets and press install

jamovi

Installed | Available | Sideload

INSTALL

lsj-data - learning statistics with jamovi 1.0.1
Danielle Navarro, David Foxcroft
This module provides examples data sets to accompany the book learning statistics with jamovi.

INSTALL

r-datasets - R data sets 1.0.1
The R Core Team
This module provides the example data sets from R

INSTALL

Statkat - Method Selection Tool 1.0.2
Rivka M. de Vries
This tool will help you to find an appropriate statistical method given your research question and the measurement level of your data.

INSTALL

eved from https://www.jamovi.org.

ng. (Version 4.0) [Computer software].
N snapshot 2020-08-24).

Ready | Filters 0 | Row count 0 | Filtered 0 | Deleted 0 | Added 0 | Cells edited 0

jamovi

jmv - Analyses bundled with jamovi 1.6.14

Jonathon Love, Damian Dropmann, Ravi Selker, Victor Moreno

This module represents the analyses included with jamovi. It contains many common analyses (such as t-tests, ANOVAs, regression, correlation matrices, proportion tests, contingency tables, factor analysis, etc).

HIDE

gamlj - General Analyses for Linear Models in jamovi 2.3.0

Marcello Gallucci

A suite for estimation of linear models, such as the general linear model, linear mixed model, generalized linear models and generalized mixed models. For ech family, models can be estimated with categorical and/or continuous variables, with options to facilitate estimation of interactions, simple slopes, simple effects, post-hoc tests, contrast analysis and visualization of the results.

HIDE   REMOVE
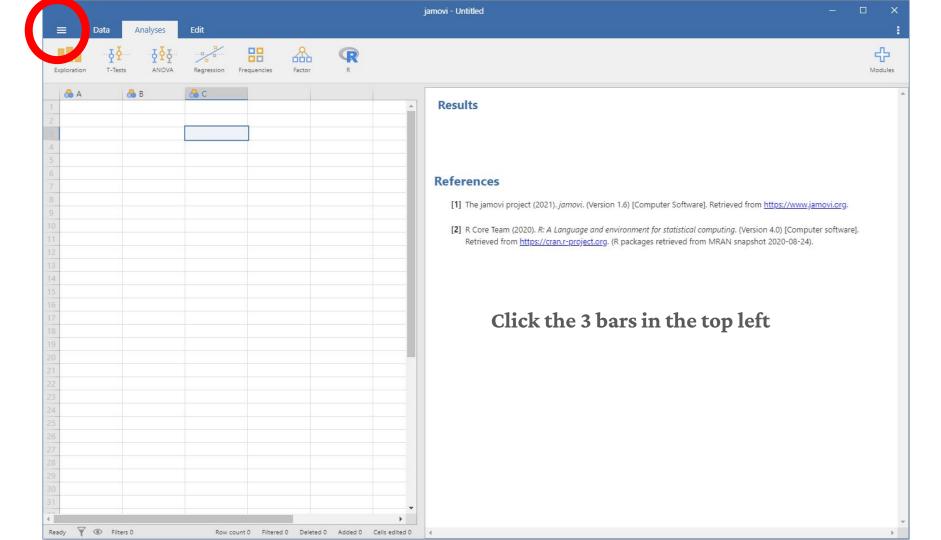
jmvbaseR - Base R 1.1.2
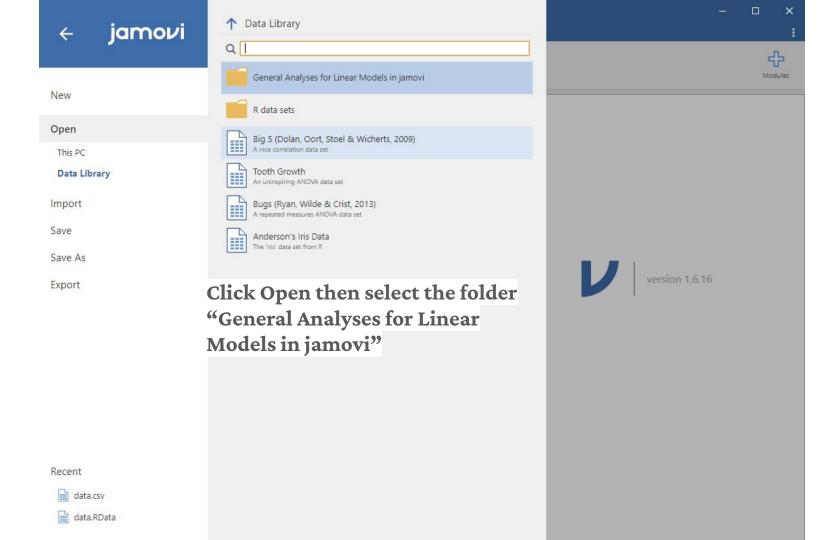
Jonathon Love, Damian Dropmann, Ravi Selker
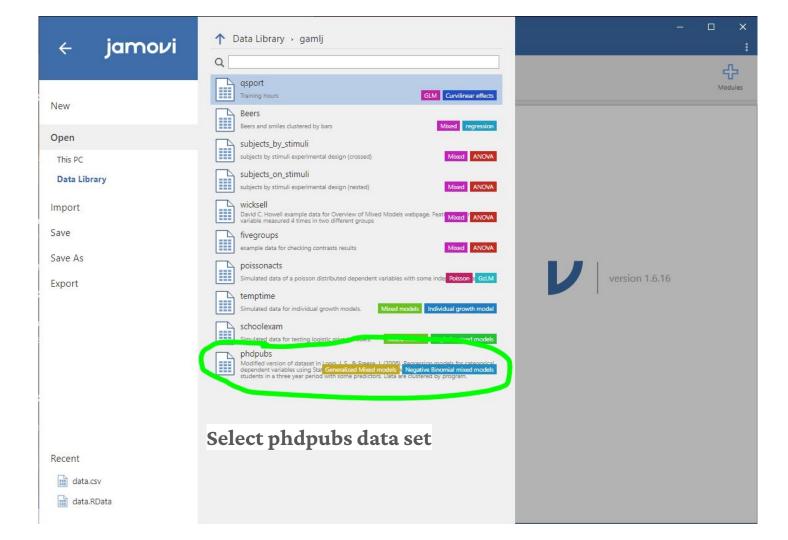
A simple module which makes the analyses from the stats package (included with R) useable from jamovi. This module is ideal for teaching and learning R (when run in 'syntax mode') as it demonstrates the syntax for some basic but common R analyses.

SHOW   REMOVE

jamovi - Untitled

Data | Analyses | Edit

Exploration | T-Tests | ANOVA | Regression | Frequencies | Factor | R | Modules

	A	B	C		
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					
11					
12					
13					
14					
15					
16					
17					
18					
19					
20					
21					
22					
23					
24					
25					
26					
27					
28					
29					
30					
31					

Results

References

[1] The jamovi project (2021). *jamovi*. (Version 1.6) [Computer Software]. Retrieved from https://www.jamovi.org.

[2] R Core Team (2020). *R: A Language and environment for statistical computing*. (Version 4.0) [Computer software]. Retrieved from https://cran.r-project.org. (R packages retrieved from MRAN snapshot 2020-08-24).

Click the 3 bars in the top left

Ready | Filters 0 | Row count 0 | Filtered 0 | Deleted 0 | Added 0 | Cells edited 0

Data Library

New

Open

This PC

Data Library

Import

Save

Save As

Export

📁 General Analyses for Linear Models in jamovi

📁 R data sets

📊 Big 5 (Dolan, Oort, Stoel & Wicherts, 2009)
A nice correlation data set

📊 Tooth Growth
An uninspiring ANOVA data set

📊 Bugs (Ryan, Wilde & Crist, 2013)
A repeated measures ANOVA data set

📊 Anderson's Iris Data
The 'iris' data set from R

Click Open then select the folder "General Analyses for Linear Models in jamovi"

Modules

ν | version 1.6.16

Recent

📄 data.csv

📄 data.RData

Select phdpubs data set
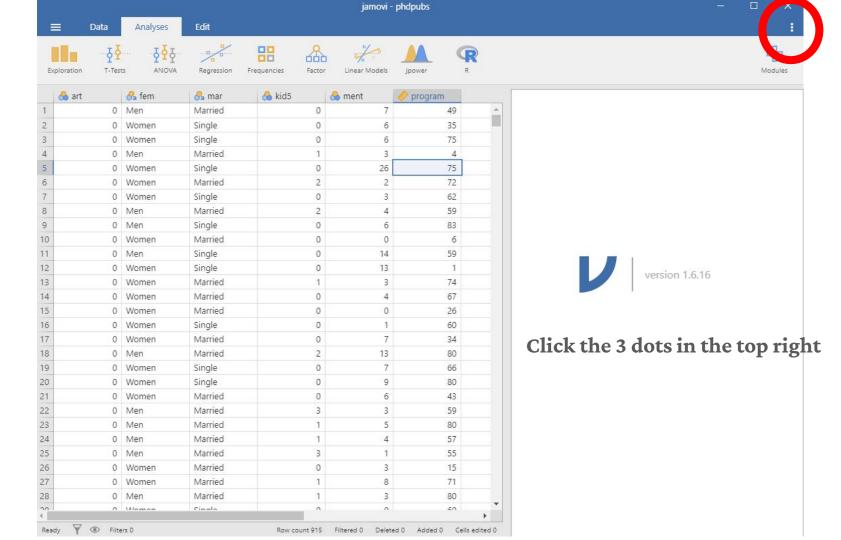
Now, we should be here with the data opened.
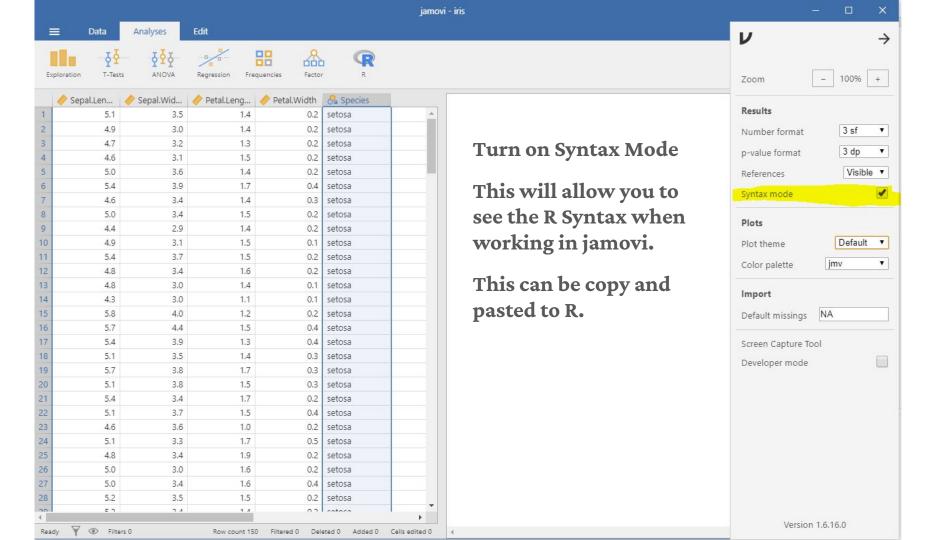
There are 3 continuous variables:
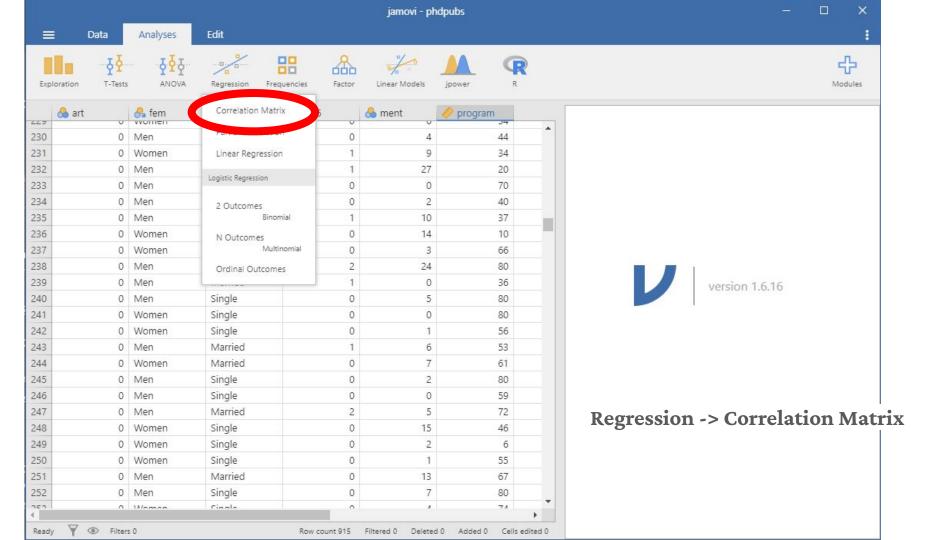
- **Number of articles published in the final three years of PhD studies**
- **Number of young children, age 5 and under**
- **Prestige of the PhD department**
- **Number of publications by the mentor in the preceding three years**
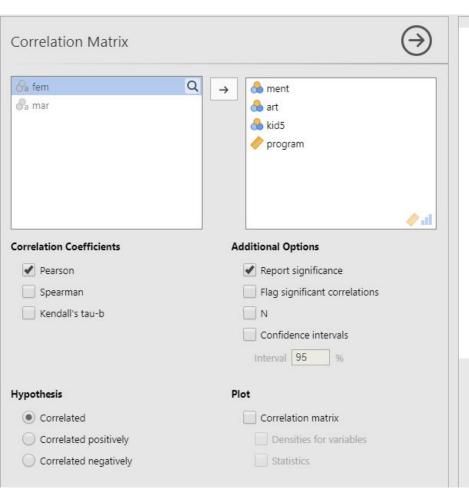
There is 2 dichotomous, nominal variable:

- **Gender**
- **Marital status**

Click the 3 dots in the top right

Turn on Syntax Mode

This will allow you to see the R Syntax when working in jamovi.

This can be copy and pasted to R.

Regression -> Correlation Matrix

Correlation Matrix

Correlation Matrix

Panel options:

Correlation Coefficients
- ☑ Pearson
- ☐ Spearman
- ☐ Kendall's tau-b

Additional Options
- ☑ Report significance
- ☐ Flag significant correlations
- ☐ N
- ☐ Confidence intervals
- Interval 95 %

Hypothesis
- ⦿ Correlated
- ○ Correlated positively
- ○ Correlated negatively

Plot
- ☐ Correlation matrix
- ☐ Densities for variables
- ☐ Statistics

Variable list: fem, mar → ment, art, kid5, program

```
jmv::corrMatrix(
    data = data,
    vars = vars(ment, art, kid5, program))
```

```
#
#  Correlation Matrix
#  ─────────────────────────────────────────────────────────────────────────
#                                    ment        art        kid5     program
#  ─────────────────────────────────────────────────────────────────────────
#    ment      Pearson's r            —
#              p-value                —
#
#    art       Pearson's r          0.3059        —
#              p-value            < .0001         —
#
#    kid5      Pearson's r          0.0370    -0.0452        —
#              p-value              0.2633     0.1718        —
#
#    program   Pearson's r          0.2393     0.0702     -0.0172        —
#              p-value            < .0001      0.0337      0.6037        —
#  ─────────────────────────────────────────────────────────────────────────
#
#
```

DATA VARIABLE

fem

Description

Measure type ⚫ Nominal ▾

Data type Text ▾

Missing values

Levels
Men
Women

Retain unused levels ◯

art | fem | mar | kid5 | ment | program

DATA VARIABLE

mar

Description

Measure type ⚫ Nominal ▾

Data type Text ▾

Missing values

Levels
Married
Single

Retain unused levels ◯

art | fem | mar | kid5 | ment | program

jamovi - phdpubs

Data | Analyses | Edit

Exploration | T-Tests | ANOVA | Regression | Frequencies | Factor | Linear Models | jpower | R | Modules

	art	fem	mar
1	0	Men	Married
2	0	Women	Single
3	0	Women	Single
4	0	Men	Married
5	0	Women	Single
6	0	Women	Married
7	0	Women	Single
8	0	Men	Married
9	0	Men	Single
10	0	Women	Married
11	0	Men	Single
12	0	Women	Single
13	0	Women	Married
14	0	Women	Married
15	0	Women	Married
16	0	Women	Single
17	0	Women	Married
18	0	Men	Married
19	0	Women	Single
20	0	Women	Single
21	0	Women	Married
22	0	Men	Married
23	0	Men	Married
24	0	Men	Married
25	0	Men	Married
26	0	Women	Married
27	0	Women	Married
28	0	Men	Married
29	0	Women	Single
30	0	Men	Married
31	0	Women	Single
32	0	Men	Married
33	0	Women	Single
34	0	Men	Single

Independent Samples T-Test

mar
kid5
ment
program

Dependent Variables
art

Grouping Variable
fem

Tests
- ☑ Student's
  - ☐ Bayes factor
  - Prior 0.707
- ☑ Welch's
- ☑ Mann-Whitney U

Hypothesis
- ◉ Group 1 ≠ Group 2
- ○ Group 1 > Group 2
- ○ Group 1 < Group 2

Missing values
- ◉ Exclude cases analysis by analysis
- ○ Exclude cases listwise

Additional Statistics
- ☐ Mean difference
  - ☑ Confidence interval 95 %
- ☑ Effect size
  - ☐ Confidence interval 95 %
- ☐ Descriptives
- ☐ Descriptives plots

Assumption Checks
- ☐ Homogeneity test
- ☐ Normality test
- ☑ Q-Q plot

```
jmv::ttestIS(
    formula = art ~ fem,
    data = data,
    vars = art,
    welchs = TRUE,
    mann = TRUE,
    qq = TRUE,
    ci = TRUE,
    effectSize = TRUE)
```

```
#
#  Independent Samples T-Test
#
#                        Statistic   df      p        Effect Size
#
#  art   Student's t     3.244       913.0   0.0012   Cohen's d                  0.2152
#        Welch's t       3.330       885.9   0.0009   Cohen's d                  0.2180
#        Mann-Whitney U  94481               0.0141   Rank biserial correlation  0.09142
#
#
#
```

Plots

art

Ready | Filters 0 | Row count 915 | Filtered 0

Descriptives

Variables
- art

Split by
- fem

☐ Frequency tables

Statistics

Sample Size
- ☑ N
- ☐ Missing

Percentile Values
- ☐ Cut points for [4] equal groups
- ☐ Percentiles [25,50,75]

Dispersion
- ☑ Std. deviation
- ☑ Minimum
- ☐ Variance
- ☑ Maximum
- ☐ Range
- ☐ S. E. Mean
- ☐ IQR

Central Tendency
- ☑ Mean
- ☑ Median
- ☐ Mode
- ☐ Sum

Distribution
- ☐ Skewness
- ☐ Kurtosis

Normality
- ☐ Shapiro-Wilk

Available variables:
- mar
- kid5
- ment
- program

(Theoretical Quantiles / Standardized plot)

Descriptives

```
jmv::descriptives(
    formula = art ~ fem,
    data = data,
    missing = FALSE)
```

```
#
#  Descriptives
#  ──────────────────────────────────────────
#                             fem      art
#  ──────────────────────────────────────────
#   N                  Men            494
#                      Women          421
#   Mean               Men          1.883
#                      Women        1.470
#   Median             Men          1.000
#                      Women            1
#   Standard deviation Men          2.179
#                      Women        1.551
#   Minimum            Men              0
#                      Women            0
#   Maximum            Men             19
#                      Women           10
#  ──────────────────────────────────────────
#
#
```

Data   Analyses   Edit

Exploration   T-Tests   ANOVA   Regression   Frequencies   Factor   Linear Models   jpower   R   Modules

	art	fem	mar
1	0	Men	Married
2	0	Women	Single
3	0	Women	Single
4	0	Men	Married
5	0	Women	Single
6	0	Women	Married
7	0	Women	Single
8	0	Men	Married
9	0	Men	Single
10	0	Women	Married
11	0	Men	Single
12	0	Women	Single
13	0	Women	Married
14	0	Women	Married
15	0	Women	Married
16	0	Women	Single
17	0	Women	Married
18	0	Men	Married
19	0	Women	Single
20	0	Women	Single
21	0	Women	Married
22	0	Men	Married
23	0	Men	Married
24	0	Men	Married
25	0	Men	Married
26	0	Women	Married
27	0	Women	Married
28	0	Men	Married
29	0	Women	Single
30	0	Men	Married
31	0	Women	Single
32	0	Men	Married
33	0	Women	Single
34	0	Men	Single

Independent Samples T-Test

fem
kid5
ment
program

Dependent Variables
art

Grouping Variable
mar

Tests
- ☑ Student's
  - ☐ Bayes factor
  - Prior 0.707
- ☑ Welch's
- ☑ Mann-Whitney U

Hypothesis
- ◉ Group 1 ≠ Group 2
- ○ Group 1 > Group 2
- ○ Group 1 < Group 2

Missing values
- ◉ Exclude cases analysis by analysis
- ○ Exclude cases listwise

Additional Statistics
- ☐ Mean difference
  - ☑ Confidence interval 95 %
- ☑ Effect size
  - ☐ Confidence interval 95 %
- ☐ Descriptives
- ☐ Descriptives plots

Assumption Checks
- ☐ Homogeneity test
- ☐ Normality test
- ☑ Q-Q plot

```
jmv::ttestIS(
    formula = art ~ mar,
    data = data,
    vars = art,
    welchs = TRUE,
    mann = TRUE,
    qq = TRUE,
    ci = TRUE,
    effectSize = TRUE)


#
#  Independent Samples T-Test
#
#                          Statistic   df       p                                        Effect Size
#
#   art   Student's t       1.129      913.0    0.2592   Cohen's d                          0.07892
#         Welch's t         1.187      710.3    0.2356   Cohen's d                          0.08087
#         Mann-Whitney U    89634               0.2773   Rank biserial correlation          0.04265
#
#
#
```

Plots

art

Ready   Filters 0   Row count 915   Filtered 0

Data   Analyses   Edit

Exploration   T-Tests   ANOVA   Regression   Frequencies   Factor   Linear Models   jpower   R   Modules

	art	fem	mar
1	0	Men	Married
2	0	Women	Single
3	0	Women	Single
4	0	Men	Married
5	0	Women	Single
6	0	Women	Married
7	0	Women	Single
8	0	Men	Married
9	0	Men	Single
10	0	Women	Married
11	0	Men	Single
12	0	Women	Single
13	0	Women	Married
14	0	Women	Married
15	0	Women	Married
16	0	Women	Single
17	0	Women	Married
18	0	Men	Married
19	0	Women	Single
20	0	Women	Single
21	0	Women	Married
22	0	Men	Married
23	0	Men	Married
24	0	Men	Married
25	0	Men	Married
26	0	Women	Married
27	0	Women	Married
28	0	Men	Married
29	0	Women	Single
30	0	Men	Married
31	0	Women	Single
32	0	Men	Married
33	0	Women	Single
34	0	Men	Single

Descriptives

Search box: fem

- fem
- kid5
- ment
- program

Variables
- art

Split by
- mar

☐ Frequency tables

▼ Statistics

Sample Size
☑ N   ☐ Missing

Percentile Values
☐ Cut points for [4] equal groups
☐ Percentiles [25,50,75]

Dispersion
☑ Std. deviation   ☑ Minimum
☐ Variance   ☑ Maximum
☐ Range   ☐ S. E. Mean
☐ IQR

Central Tendency
☑ Mean
☑ Median
☐ Mode
☐ Sum

Distribution
☐ Skewness
☐ Kurtosis

Normality
☐ Shapiro-Wilk

▶ Plots

Ready   Filters 0   Row count 915   Filtered 0

Descriptives

```
jmv::descriptives(
    formula = art ~ mar,
    data = data,
    missing = FALSE)
```

```
#
# Descriptives
#
#                              mar        art
#
#   N                      Married        606
#                          Single         309
#   Mean                   Married      1.744
#                          Single       1.592
#   Median                 Married      1.000
#                          Single           1
#   Standard deviation     Married      2.019
#                          Single       1.729
#   Minimum                Married          0
#                          Single           0
#   Maximum                Married         19
#                          Single           7
#
#
```

Exploration   T-Tests   ANOVA   Regression   Frequencies   Factor   Linear Models   jpower   R   Modules

Linear Regression

Dependent Variable
- art

Covariates
- program
- ment
- kid5

Factors
- mar
- fem

Model Builder

Predictors
- program
- ment
- kid5
- mar
- fem

Blocks

Block 1
- program
- ment
- kid5
- mar
- fem

+ Add New Block

Reference Levels

Assumption Checks

Assumption Checks
- ☐ Autocorrelation test
- ☑ Collinearity statistics
- ☐ Normality test
- ☑ Q-Q plot of residuals
- ☐ Residual plots

Data Summary
- ☐ Cook's distance

Model Fit

Linear Regression

```
jmv::linReg(
    data = data,
    dep = art,
    covs = vars(program, ment, kid5),
    factors = vars(mar, fem),
    blocks = list(
        list(
            "program",
            "ment",
            "kid5",
            "mar",
            "fem")),
    refLevels = list(
        list(
            var="mar",
            ref="Married"),
        list(
            var="fem",
            ref="Men")),
    r2Adj = TRUE,
    modelTest = TRUE,
    ci = TRUE,
    ciStdEst = TRUE,
    qqPlot = TRUE,
    collin = TRUE,
    emMeans = ~ fem + ment + kid5)
```

```
#
# Model Fit Measures
#
#  Model    R       R²       Adjusted R²     F       df1    df2     p
#
#    1     0.3322   0.1103       0.1054     22.55     5     909    < .0001
#
#
#
#
#
```

Model Specific Results

Model 1

```
#
# Model Coefficients - art
#
#  Predictor        Estimate    SE         Lower       Upper       t
#
#  Intercept    ᵃ   1.56475     0.197847   1.176457    1.953036    7.908886
#  program          1.014e-5    0.002882   -0.005646   0.005666    0.003518
#  ment             0.06118     0.006569   0.048291    0.074077    9.313580
#  kid5             -0.29140    0.090927   -0.469855   -0.112951   -3.204785
#  mar:
#    Single - Married  -0.26533  0.145492  -0.550872   0.020207   -1.823691
#  fem:
#    Women - Men   -0.38029    0.128111   -0.631720   -0.128864   -2.968458
#
#    ᵃ Represents reference level
#
```

Assumption Checks

- ☐ Autocorrelation test
- ☑ Collinearity statistics
- ☐ Normality test
- ☑ Q-Q plot of residuals
- ☐ Residual plots

Data Summary

- ☐ Cook's distance

∨ | Model Fit

Fit Measures

- ☑ R
- ☑ R²
- ☑ Adjusted R²
- ☐ AIC
- ☐ BIC
- ☐ RMSE

Overall Model Test

- ☑ F test

∨ | Estimated Marginal Means

| program |
| ment |
| **kid5** |
| mar |
| fem |

→

Marginal Means

Term 1 ✕
👥 fem

Term 2 ✕
👥 ment

Term 3 ✕
👥 kid5

➕ Add New Term

General Options

- ☑ Equal cell weights
- ☑ Confidence interval
  Interval [95] %

Output

- ☑ Marginal means plots
- ☐ Marginal means tables

**Open R Studio
and Press File
then New Project**

Select New Directory

Select New Project

Name the Directory (the new project folder) and choose the location.

It's typically good practice to make folders for:

- **data**
- **docs**
- **tables**
- **figures**

Select File →
New File →
R Markdown

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function   Addins ▾   🄰 PsiChiDemo ▾

Console   Terminal ✕   Jobs ✕   Environment   History   Connections   Tutorial

~/R Practice/PsiChiDemo/ →   📂 💾 🄳 Import Dataset ▾ 🧹   ≡ List ▾ ⟳ ▾

R version 4.0.3 (202
ak Out"
Copyright (C) 2020 T
l Computing
Platform: x86_64-w64

R is free software a
RRANTY.
You are welcome to r
onditions.
Type 'license()' or
etails.

  Natural language s
sh locale

R is a collaborative
s.
Type 'contributors()
'citation()' on how
lications.

Type 'demo()' for so
 help, or
'help.start()' for a
lp.
Type 'q()' to quit R

>

New R Markdown

📄 Document	**Title:**	PsiChiDemo
📽 Presentation	**Author:**	Joshua M. Garcia
🄱 Shiny	
🗔 From Template	

Default Output Format:

◉ HTML
Recommended format for authoring (you can switch to PDF or Word output anytime).

○ PDF
PDF output requires TeX (MiKTeX on Windows, MacTeX 2013+ on OS X, TeX Live 2013+ on Linux).

○ Word
Previewing Word documents requires an installation of MS Word (or Libre/Open Office on Linux).

Create Empty Document   OK   Cancel

R Studio presents with a description of the R Markdown document. Read this for clarification.

Save the file to your docs folder.

File   Edit   Code   View   Plots   Session   Build   Debug   Profile   Tools   Help

Go to file/function   Addins ▾   Ⓡ PsiChiDemo ▾

Untitled1 ✕

Knit ▾   Insert ▾   ▲ ▼   Run ▾

```
  1  ---
  2  title: "PsiChiDemo"
  3  author: "Joshua M. Garcia"
  4  date: "3/2/2021"
  5  output: html_document
  6  ---
  7
  8  ```{r setup, include=FALSE}
  9  knitr::opts_chunk$set(echo = TRUE)
 10  ```
 11
 12  ## R Markdown
 13
 14  This is an R Markdown document. Markdown is a simple formatting syntax for
     authoring HTML, PDF, and MS Word documents. For more details on using R
     Markdown see <http://rmarkdown.rstudio.com>.
 15
 16  When you click the **Knit** button a document will be generated that
```

2:1   # PsiChiDemo   R Markdown

Console   Terminal   Jobs

~/R Practice/PsiChiDemo/

```
Type 'license()' or 'licence()' for distribution details.

  Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> |
```

Environment   History

Import Dataset

Global Environment ▾

Environment is empty

Files   Plots   Packages

New Folder   Delete

PsiChiDemo

▲ Name
  ..
  data
  docs
  figures
  PsiChiDemo.R...
  tables

1. **Select Tools**
2. **Global Options**
3. **R Markdown**
4. **Set your R Markdown document to evaluate chunks in directory to "Project"**

Note that Global Settings are saved between sessions and R projects.

Brief introduction to understanding regression:
https://joshualoftus.com/posts/2020-11-23-least-squares-as-springs/

Common statistical tests are linear models

Last updated: 28 June, 2019. Also check out the Python version!

	Common name	Built-in function in R	Equivalent linear model in R	Exact?	The linear model in words	Icon
Simple regression: lm(y ~ 1 + x)	**y is independent of x** P: One-sample t-test N: Wilcoxon signed-rank	t.test(y) wilcox.test(y)	lm(y ~ 1) lm(signed_rank(y) ~ 1)	✓ for N >14	One number (intercept, i.e., the mean) predicts **y**. - (Same, but it predicts the *signed rank* of **y**.)	
	P: Paired-sample t-test N: Wilcoxon matched pairs	t.test(y₁, y₂, paired=TRUE) wilcox.test(y₁, y₂, paired=TRUE)	lm(y₂ - y₁ ~ 1) lm(signed_rank(y₂ - y₁) ~ 1)	✓ for N >14	One intercept predicts the pairwise y_2-y_1 differences. - (Same, but it predicts the *signed rank* of y_2-y_1.)	
	y ~ continuous x P: Pearson correlation N: Spearman correlation	cor.test(x, y, method='Pearson') cor.test(x, y, method='Spearman')	lm(y ~ 1 + x) lm(rank(y) ~ 1 + rank(x))	✓ for N >10	One intercept plus **x** multiplied by a number (slope) predicts **y**. - (Same, but with *ranked* **x** and **y**)	
	y ~ discrete x P: Two-sample t-test P: Welch's t-test N: Mann-Whitney U	t.test(y₁, y₂, var.equal=TRUE) t.test(y₁, y₂, var.equal=FALSE) wilcox.test(y₁, y₂)	lm(y ~ 1 + G₂)[A] gls(y ~ 1 + G₂, weights=...[B])[A] lm(signed_rank(y) ~ 1 + G₂)[A]	✓ ✓ for N >11	An intercept for **group 1** (plus a difference if **group 2**) predicts **y**. - (Same, but with one variance *per group* instead of one common.) - (Same, but it predicts the *signed rank* of **y**.)	
Multiple regression: lm(y ~ 1 + x₁ + x₂ + ...)	**P: One-way ANOVA** N: Kruskal-Wallis	aov(y ~ group) kruskal.test(y ~ group)	lm(y ~ 1 + G₂ + G₃ +...+ Gₙ)[A] lm(rank(y) ~ 1 + G₂ + G₃ +...+ Gₙ)[A]	✓ for N >11	An intercept for **group 1** (plus a difference if group ≠ 1) predicts **y**. - (Same, but it predicts the *rank* of **y**.)	
	P: One-way ANCOVA	aov(y ~ group + x)	lm(y ~ 1 + G₂ + G₃ +...+ Gₙ + x)[A]	✓	- (Same, but plus a slope on **x**.) *Note: this is discrete AND continuous. ANCOVAs are ANOVAs with a continuous x.*	
	P: Two-way ANOVA	aov(y ~ group * sex)	lm(y ~ 1 + G₂ + G₃ +...+ Gₙ + S₂ + S₃ +...+ Sₖ + G₂*S₂ + G₃*S₃ + ... + Gₙ*Sₖ)	✓	**Interaction term: changing sex changes the y ~ group parameters.** *Note: G₂ to N is an indicator (0 or 1) for each non-intercept levels of the **group** variable. Similarly for S₂ to K for sex. The first line (with G₁) is main effect of group, the second (with Sᵢ) for sex and the third is the **group × sex** interaction. For two levels (e.g. male/female), line 2 would just be "S₂" and line 3 would be S₂ multiplied with each Gᵢ.*	[Coming]
	Counts ~ discrete x N: Chi-square test	chisq.test(groupXsex_table)	**Equivalent log-linear model** glm(y ~ 1 + G₂ + G₃ + ... + Gₙ + S₂ + S₃+ ... + Sₖ + G₂*S₂ + G₃*S₃ +...+ Gₙ*Sₖ, family=...)[A]	✓	**Interaction term: (Same as Two-way ANOVA.)** *Note: Run glm using the following arguments: glm(model, family=poisson())* *As linear-model, the Chi-square test is $log(y_i) = log(N) + log(\alpha_i) + log(\beta_i) + log(\alpha_i\beta_i)$ where α_i and β_i are proportions. See more info in the accompanying notebook.*	Same as Two-way ANOVA
	N: Goodness of fit	chisq.test(y)	glm(y ~ 1 + G₂ + G₃ +...+ Gₙ, family=...)[A]	✓	(Same as One-way ANOVA and see Chi-Square note.)	1W-ANOVA

List of common parametric (P) non-parametric (N) tests and equivalent linear models. The notation y ~ 1 + x is R shorthand for y = 1·b + a·x which most of us learned in school. Models in similar colors are highly similar, but really, notice how similar they *all* are across colors! For non-parametric models, the linear models are reasonable approximations for non-small sample sizes (see "Exact" column and click links to see simulations). Other less accurate approximations exist, e.g., Wilcoxon for the sign test and Goodness-of-fit for the binomial test. The signed rank function is signed_rank = function(x) sign(x) * rank(abs(x)). The variables Gᵢ and Sᵢ are "dummy coded" indicator variables (either 0 or 1) exploiting the fact that when Δx = 1 between categories the difference equals the slope. Subscripts (e.g., G₂ or y₁) indicate different columns in data. lm requires long-format data for all non-continuous models. All of this is exposed in greater detail and worked examples at https://lindeloev.github.io/tests-as-linear.

[A] See the note to the two-way ANOVA for explanation of the notation.
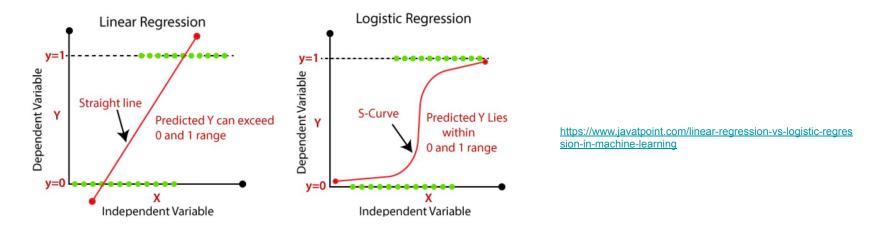
[B] Same model, but with one variance per group: gls(value ~ 1 + G₂, weights = varIdent(form = ~1|group), method="ML").

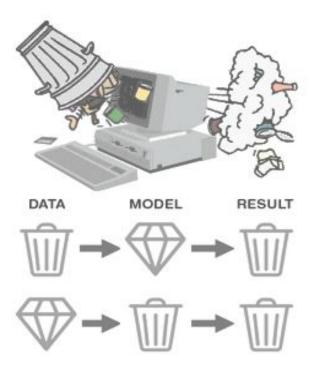**The code to reproduce today's demonstration are available here:**
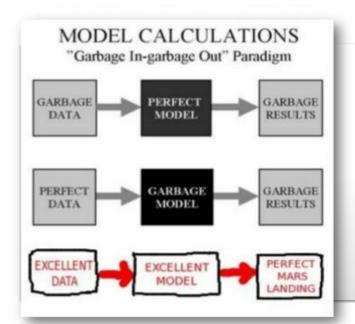
https://joshmgarciaa.github.io/PsiChiDASW1/WS2.html

The linear model assumes that the probability p is a linear function of the regressors, while the logistic model assumes that the natural log of the odds $p/(1-p)$ is a linear function of the regressors.

"The **logistic model is unavoidable if it fits the data much better** than the linear model. And sometimes it does. But in **many situations the linear model fits just as well, or almost as well, as the logistic model**. In fact, in many situations, the l**inear and logistic model give results that are practically indistinguishable except that the logistic estimates are harder to interpret** (Hellevik 2007)."

DATA → MODEL → RESULT

MODEL CALCULATIONS
"Garbage In-garbage Out" Paradigm

GARBAGE DATA → PERFECT MODEL → GARBAGE RESULTS

PERFECT DATA → GARBAGE MODEL → GARBAGE RESULTS

EXCELLENT DATA → EXCELLENT MODEL → PERFECT MARS LANDING

"The old concept of **Garbage in, Garbage out (GIGO)** *still reigns"*

R Resources

- [New to R? Kickstart your learning and career with these 6 steps!](#)
- [R resources (free courses, books, tutorials, & cheat sheets)](#)
- [swirl teaches you R programming and data science interactively, at your own pace, and right in the R console!](#)
- [R for Psychological Science by Danielle Navarro](#)
- [R for Data Science](#)
- [Andy Field - Youtube](#)
- [Statistics of DOOM - Youtube](#)
- [StatQuest with Josh Starmer - Youtube](#)
- [3Blue1Brown - Youtube](#)
- [An Introduction to Psychological Statistics - Textbook](#)
- [Get R Done | R Tutorials - Youtube](#)
- [R for the Rest of Us](#)