

Package ‘moderndive’

January 22, 2018

Type Package

Title Accompaniment Package to ModernDive: An Introduction to
Statistical and Data Sciences via R

Version 0.1.1

Maintainer Albert Y. Kim <albert.y.s.kim@gmail.com>

Description An accompaniment R package to ModernDive: An Introduction to
Statistical and Data Sciences via R available at <<http://moderndive.com/>>,
in particular wrapper functions targeted at novices to easily generate tidy
linear regression output.

Depends R (>= 3.2.4)

License MIT + file LICENSE

Encoding UTF-8

LazyData true

URL https://github.com/ModernDive/moderndive_package

BugReports https://github.com/ModernDive/moderndive_package/issues

Imports magrittr, dplyr, rlang, tibble, janitor, broom (>= 0.4.3),
formula.tools, stringr, knitr, assertive

RoxygenNote 6.0.1

Suggests testthat, covr, tidyverse, ggplot2

NeedsCompilation no

Author Albert Y. Kim [cre],
Chester Ismay [aut],
Andrew Bray [ctb]

Repository CRAN

Date/Publication 2018-01-22 16:38:26 UTC

R topics documented:

bowl	2
bowl_samples	3
get_regression_points	4
get_regression_summaries	4
get_regression_table	5
moderndive	6
pennies	6
rep_sample_n	7

Index	9
--------------	----------

bowl	<i>A sampling bowl of red and white balls</i>
------	-----------------------------------------------

Description

A sampling bowl used as the population in a simulated sampling exercise. Also known as the urn sampling framework https://en.wikipedia.org/wiki/Urn_problem.

Usage

```
bowl
```

Format

A data frame 2400 rows representing different balls in the bowl, of which 900 are red and 1500 are white.

ball_ID ID variable used to denote all balls. Note this value is not marked on the balls themselves
color color of ball: red or white

Examples

```
library(dplyr)
library(ggplot2)

# Take 10 different samples of size n = 50 balls from bowl
bowl_samples_simulated <- bowl %>%
  rep_sample_n(50, reps = 10)

# Compute 10 different p_hats (prop red) based on 10 different samples of
# size n = 50
p_hats <- bowl_samples_simulated %>%
  group_by(replicate, color) %>%
  summarize(count = n()) %>%
  mutate(proportion = count/50) %>%
  filter(color == "red")
```

```
# Plot sampling distribution
ggplot(p_hats, aes(x = proportion)) +
  geom_histogram(binwidth = 0.05) +
  labs(x = expression(hat(p)), y = "Number of samples",
       title = "Sampling distribution of p_hat based 10 samples of size n = 50")
```

bowl_samples

Sampling from a tub of balls

Description

Counting the number of red balls in 10 samples of size $n = 50$ balls from https://github.com/moderndive/moderndive/blob/master/data-raw/sampling_bowl.jpeg

Usage

```
bowl_samples
```

Format

A data frame 10 rows representing different groups of students' samples of size $n = 50$ and 5 variables

group Group name

red Number of red balls sampled

white Number of white balls sampled

green Number of green balls sampled

n Total number of balls samples

Examples

```
library(dplyr)
library(ggplot2)

# Compute proportion red
bowl_samples <- bowl_samples %>%
  mutate(prop_red = red / n)

# Plot sampling distributions
ggplot(bowl_samples, aes(x = prop_red)) +
  geom_histogram(binwidth = 0.05) +
  labs(x = expression(hat(p)), y = "Number of samples",
       title = "Sampling distribution of p_hat based 10 samples of size n = 50")
```

get_regression_points *Get regression points*

Description

Wrapper for tidy `lm()` regression function fit output

Usage

```
get_regression_points(model, digits = 3, print = FALSE, ...)
```

Arguments

<code>model</code>	a model object (Currently only <code>lm</code> is supported)
<code>digits</code>	number of digits precision in output table
<code>print</code>	If TRUE, return in print format suitable for R Markdown
<code>...</code>	other arguments passed to <code>lm()</code>

Value

A tibble or nicely formatted table

Examples

```
mpg_model <- lm(mpg ~ cyl, data = mtcars)
get_regression_points(mpg_model)
```

get_regression_summaries
Get regression summary values

Description

Wrapper for `lm()` regression function fit summary data

Usage

```
get_regression_summaries(model, digits = 3, print = FALSE, ...)
```

Arguments

<code>model</code>	a model object (Currently only <code>lm</code> is supported)
<code>digits</code>	number of digits precision in output table
<code>print</code>	If TRUE, return in print format suitable for R Markdown
<code>...</code>	other arguments passed to <code>lm()</code>

Value

A tibble or nicely formatted table

Examples

```
mpg_model <- lm(mpg ~ cyl, data = mtcars)
get_regression_summaries(mpg_model)
```

get_regression_table *Get regression table*

Description

Wrapper for tidy `lm()` regression function output

Usage

```
get_regression_table(model, digits = 3, print = FALSE, ...)
```

Arguments

- `model` a model object (Currently only `lm` is supported)
- `digits` number of digits precision in output table
- `print` If TRUE, return in print format suitable for R Markdown
- `...` other arguments passed to `lm()`

Value

A tibble or nicely formatted table

Examples

```
mpg_model <- lm(mpg ~ cyl, data = mtcars)
get_regression_table(mpg_model)
```

moderndive

moderndive - Accompaniment Package to ModernDive: An Introduction to Statistical and Data Sciences via R

Description

An accompaniment R package to ModernDive: An Introduction to Statistical and Data Sciences via R available at <http://moderndive.com/>, in particular wrapper functions targeted at novices to easily generate tidy linear regression output in Chapter 6: Data Modeling with Regression.

Examples

```
# Example usage:
library(moderndive)
library(dplyr)

# Regression models
mpg_model <- lm(mpg ~ hp, data = mtcars)
mpg_mlr_model <- lm(mpg ~ hp + wt, data = mtcars)
mpg_mlr_model2 <- lm(mpg ~ hp + cyl, data = mtcars)

# Regression tables
get_regression_table(model = mpg_model)
get_regression_table(mpg_mlr_model, digits = 4, print = TRUE)

# Regression points. For residual analysis for example
get_regression_points(mpg_mlr_model2)

# Regression summaries
get_regression_summaries(mpg_model)

# Can also use `>%`
mpg_model %>% get_regression_summaries(digits = 5, print = TRUE)
```

pennies

A population of 800 pennies sampled in 2011

Description

A dataset of 800 pennies to be treated as a sampling population. Data on these pennies were recorded in 2011.

Usage

pennies

Format

A data frame 800 rows representing different pennies and 2 variables

year Year of minting

age_in_2011 Age in 2011

Source

StatCrunch <https://www.statcrunch.com/app/index.php?dataid=301596>

Examples

```
library(dplyr)
library(ggplot2)

# Take 10 different samples of size n = 50 pennies from population
many_samples <- pennies %>%
  rep_sample_n(size = 50, reps = 25)
many_samples

# Compute mean year of minting for each sample
sample_means <- many_samples %>%
  group_by(replicate) %>%
  summarize(mean_year = mean(year))

# Plot sampling distribution
ggplot(sample_means, aes(x = mean_year)) +
  geom_histogram(binwidth = 1) +
  labs(x = expression(bar(x)), y = "Number of samples",
       title = "Sampling distribution of  $\bar{x}$  based 25 samples of size n = 50")
```

rep_sample_n

Repeating sampling.

Description

Repeating sampling.

Usage

```
rep_sample_n(tbl, size, replace = FALSE, reps = 1)
```

Arguments

tbl	tbl of data.
size	The number of rows to select.
replace	Sample with or without replacement?
reps	The number of samples to collect.

Value

A `tbl_df` that aggregates all created samples, with the addition of a `replicate` column that the `tbl_df` is also grouped by

Source

https://github.com/OpenIntroOrg/oilabs-r-package/blob/master/R/rep_sample_n.R

Examples

```
library(dplyr)
library(ggplot2)
N <- 2400
tub <- data_frame(
  ball_ID = 1:N,
  color = c(rep("red", 900), rep("white", N-900))
)

p_hats <- tub %>%
  rep_sample_n(size=50, reps=10) %>%
  group_by(replicate) %>%
  summarize(prop_red = mean(color == "red"))

ggplot(p_hats, aes(x = prop_red)) +
  geom_histogram(binwidth = 0.05)
```


Index

*Topic **datasets**

bowl, [2](#)

bowl_samples, [3](#)

pennies, [6](#)

bowl, [2](#)

bowl_samples, [3](#)

get_regression_points, [4](#)

get_regression_summaries, [4](#)

get_regression_table, [5](#)

moderndive, [6](#)

moderndive-package (moderndive), [6](#)

pennies, [6](#)

rep_sample_n, [7](#)