

## sda123

är ett **R-paket** med hjälpfunktioner för kurserna Statistik och dataanalys I, II och III utvecklat vid Stockholms universitet.

De funktioner som används på **alla** tre av SDAI, II och III kurserna ligger först.

På senare sidor presenteras funktioner som enbart används på SDAII respektive SDAIII.

Kurssiffrorna på SDA-ikonen i övre hörnet av varje sida talar om vilka kurser sidan gäller för.

## Regression - utskrift

Skriva ut regressionsresultat med ANOVA-tabell

Valfritt variabelnamn

```
fit <- lm(mpg ~ hp, data = mtcars)
reg_summary(fit)
```

### PARAMETRAR

<b>anova</b>	om TRUE, ANOVA table med F-test
<b>fit_measures</b>	om TRUE, anpassningsmått (R <sup>2</sup> )
<b>param</b>	om TRUE, parameterskattningar
<b>conf_intervals</b>	om TRUE, konfidensintervall params
<b>vif_factors</b>	om TRUE, variance inflation factors

## Regression - simulering

Simulera data från linjär multipel regression

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p + \varepsilon$$

```
reg_simulate(n = 200, sigma_eps = 1,
             betavect = (1, -1, 0.2))
```

### PARAMETRAR (fler finns, se hjälpen)

<b>n</b>	antal observationer
<b>sigma_eps</b>	standardavvikelse för feltermen epsilon
<b>betavect</b>	vektor med populationsparametrar
<b>intercept</b>	om TRUE, intercept i modellen
<b>rho_x</b>	korrelationen mellan alla par av x-variabler

## Installera

Installera paketet (en gång per dator)

```
install.packages("remotes")
library(remotes)
install_github("StatisticsSU/sda123")
```

Läsa in paketet (en gång per session)

```
library(sda123)
```

## Prediktionsintervall

Plotta prediktionsintervall för enkel regression

obs! **enkel** regression

```
reg_predict(mpg ~ hp, data = mtcars)
```

### PARAMETRAR

<b>conf_interval_line</b>	om TRUE, konfidensintervall regr.linje
<b>pred_interval</b>	om TRUE, prediktionsintervall
<b>level</b>	konfidensnivå

## Korsvalidering

K-fold korsvalidering av regressionsmodeller

```
reg_crossval(mpg ~ hp, data = mtcars,
             nfolds = 4)
```

### PARAMETRAR

<b>nfolds</b>	antalet folds i korsvalideringen
<b>obs_order</b>	ordningen på observationerna före splits obs_order = "random" för slumpmässig ordning.

## Hjälp

Hjälpsidan för varje funktion nås genom frågetecken

```
?reg_summary
```

## Residualanalys

Plotta residualer för att undersöka antaganden

```
fit <- reg_residuals(fit)
```

## Korrelationsanalys

Parvisa korrelationer och test för korrelation

```
corr_matrix(data)
```

## Logistisk - utskrift



Skriva ut resultat från logistisk regression

```
fit <- glm(survived ~ age + sex,
  data = titanic, family = binomial)
logisticreg_summary(fit)
```

Ger logistisk regression

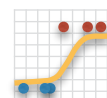
### PARAMETRAR

**odds\_ratio** Om TRUE, inferens för oddskvoter

**param** Om TRUE, inferens för beta parametrar

**conf\_intervals** om TRUE, konfidensintervall för parametrar

## Logistisk - simulering



Simulera data från logistisk regression

$$\Pr(y = 1|x) = \frac{\exp(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)}{1 + \exp(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)}$$

```
logisticreg_simulate(n = 500,
  betavect = (1, -2, 1))
```

### PARAMETRAR

**n** antal observationer

**betavect** vektor med populationsparametrar

**intercept** om TRUE, intercept i modellen

**rho\_x** korrelationen mellan alla par av x-variabler

## AR(1) - simulering



Simulera tidsserie från autoregressiv modell

```
x = simAR1(n = 100, phi = 0.7)
```

### PARAMETRAR

**n** antal simulerade observationer i tidsserien

**phi** autoregressiv parameter - lag 1

**sigma\_eps** standardavvikelsen för feltermen

