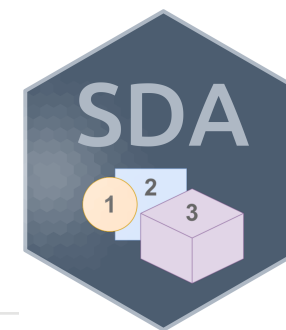


Statistik och dataanalys • sda123 • cheat sheet



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 SDAILL.

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

| | |
|-----------------------|--|
| anova | om TRUE, ANOVA table med F-test |
| fit_measures | om TRUE, anpassningsmått (R ²) |
| param | om TRUE, parameterskattningar |
| conf_intervals | om TRUE, konfidensintervall params |
| vif_factors | 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)

| | |
|------------------|--|
| n | antal observationer |
| sigma_eps | standardavvikelse för feltermen epsilon |
| betavect | vektor med populationsparametrar |
| intercept | om TRUE, intercept i modellen |
| rho_x | 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)
```

Hjälp

Hjälpsidan för varje funktion nås genom frågetecken

```
?reg_summary
```

Prediktionsintervall

Plotta prediktionsintervall för enkel regression

obs! **enkel** regression

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

PARAMETRAR

| | |
|---------------------------|--|
| conf_interval_line | om TRUE, konfidensintervall regr.linje |
| pred_interval | om TRUE, prediktionsintervall |
| level | konfidensnivå |

Korsvalidering

K-fold korsvalidering av regressionsmodeller

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

PARAMETRAR

| | |
|------------------|---|
| nfolds | antalet folds i korsvalideringen |
| obs_order | ordningen på observationerna före splits obs_order = "random" för slumpmässig ordning. |

Residualanalys

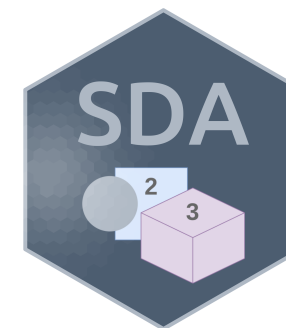
Plotta residualer för att undersöka antaganden

```
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

$$\Pr(y = 1|x) = \frac{\exp(\beta_0 + \beta_1x_1 + \dots + \beta_px_p)}{1 + \exp(\beta_0 + \beta_1x_1 + \dots + \beta_px_p)}$$

```
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 |

