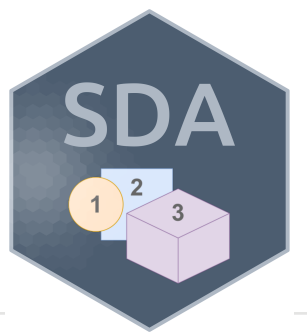


Statistik och dataanalys • sda123 • cheat sheet



sda123

är ett **R-paket** för grundkurserna Statistik och dataanalys I, II och III utvecklat vid Stockholms universitet

innehåll

- Regressionsutskrifter med ANOVA-tabell inkl F-test
- Grafer för konfidens- och prediktionsintervall
- Grafer för residualanalys
- Korsvalidering av regressionmodeller för modellval
- Simulering från linjär regressionsmodell

Installera

Installera paketet (en gång per dator)

```
install.packages("remotes")
library(remotes)
install_github("StatisticsSU/sda1paket")
```

Läsa in paketet (en gång per session)

```
library(sda1)
```

Hjälp

Hjälpsidan för varje funktion nås genom frågetecken

```
?reg_summary
```

Regression - utskrift

Skriva ut regressionsresultat med ANOVA-tabell

Valfritt variabelnamn

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

ARGUMENT

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

Prediktionsintervall

Plotta prediktionsintervall för enkel regression

obs! **enkel** regression

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

ARGUMENT

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

Residualanalys

Plotta residualer för att undersöka antaganden

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

Regression - simulering

Simulera data från linjär multipel regression

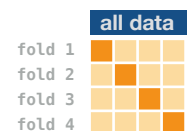
$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p + \varepsilon$$

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

ARGUMENT

n	antal observationer
sigma_eps	standardavvikelse för feltermen epsilon
betavect	vektor med populationsparametrar
intercept	om TRUE, intercept i modellen
covdist	Fördelning för x-variabler, 'normal' eller 'uniform'
rho_x	korrelationen mellan alla x-variabler

Korsvalidering



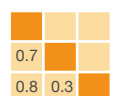
K-fold korsvalidering av regressionsmodeller

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

ARGUMENT

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

Korrelationsanalys



Parvisa korrelationer och test för korrelation

```
corr_matrix(data)
```