

# Statistik och dataanalys 1 • sda1 • cheat sheet



## sda1

är ett **R-paket** för grundkursen  
Statistik och dataanalys I, 15 hp  
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

```
?regsummary
```

## Regression - utskrift

Skriva ut regressionsresultat med ANOVA tabell etc.

Valfritt variabelnamn

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

### ARGUMENT

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

## Prediktionsintervall

Plotta prediktionsintervall för enkel regression

obs! **enkel** regression

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

### ARGUMENT

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

## Residualanalys

Plotta residualer för att undersöka antaganden

```
fit <- residuals4in1(fit)
```

## Regression - simulering

Simulera data från linjär multipel regression

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

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

### ARGUMENT

<b>N</b>	antal observationer
<b>beta_vect</b>	vektor med populationsparametrar
<b>intercept</b>	om TRUE, intercept i modellen
<b>covdist</b>	Fördelning för x-variabler. "Normal" eller "uniform"
<b>rho_x</b>	korrelationen mellan alla x-variabler

## Korsvalidering

K-fold korsvalidering av regressionsmodeller

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

### ARGUMENT

<b>formula</b>	R formula av samma typ som i <code>lm()</code>
<b>data</b>	dataframe med data
<b>nfolds</b>	antalet folds i korsvalideringen
<b>obs_order</b>	ordningen på observationerna före splits obs_order = "random" för slumpmässig ordning.

## Korrelationsanalys

```
corr_matrix(data)
```