

## R-Funktionen

a lm package stat (immer vorhanden): lm

```
> r.lm <- lm(log10(ersch) ~ log10(dist),  
             data = d.spreng)
```

b **Funktion** `summary` produziert Resultate, die man üblicherweise will.

```
> summary(r.lm)
```

Genauer: `print` zeigt die Resultate.

(„generic function“, „method“ `print.summary.lm`)

```
> r.lm <- summary(r.lm)
```

```
> str(r.lm)
```

enthält u.a. `$sigma`, `$R.square`, ...

`r.lm$coef`: geschätzte Koeffizienten

`r.lm$coef`: Tabelle mit `Std. Error`, `t value` und `P-Wert`

c **Funktion** `drop1`: Signifikanz von „Faktoren“ testen.

d **Funktion** `predict`: Generische Funktion, Hilfe durch `?predict.lm`

```
> predict(r.lm)
```

```
> t.pred <- predict(r.lm,
```

```
  newdata=data.frame(dist=seq(20,150,10)),  
  interval="prediction")
```

e **Funktion** `plot`: Residuen-Analyse, 4 Diagramme (s. später)

## 3.4

f Mühsam? → Package `regr0`, **Funktion** `regr`.

```
> r.regr <- regr(log10(ersch) ~ log10(dist) +
  log10(ladung) + Stelle, data = d.spreng,
  subset = as.numeric(Stelle) <= 4)
```

Wie `lm` zu verwenden.

- Ruft `lm`, `summary` und `drop1` auf und sammelt Ergebnisse.
- Gleicher Aufruf für verschiedene Modelle.
- Erweiterte Residuen-Analyse durch `plot.regr`

```

> r.regr
Call:
regr(formula = log10(ersch) ~ log10(dist) + log10(ladung) + Stelle,
      data = d.spreng, subset = as.numeric(Stelle) <= 4)
Fitting function: lm

Terms:
(Intercept)      2.510      NA      4.409      NA      1      NA
log10(dist)     -1.338     -0.686     -4.711     0.435     1     0.000
log10(ladung)    0.692     0.150     1.156     0.048     1     0.025
Stelle          NA      NA      0.899     0.170     3     0.093

Coefficients for factors:
$Stelle
  1      2      3      4
0.0000 0.1643 0.0217 0.1108

St.dev.error:  0.147  on 42 degrees of freedom
Multiple R^2:  0.832  Adjusted R-squared: 0.812
F-statistic:   41.7  on 5 and 42 d.f.,  p.value: 3.22e-15

```

### g Resultate von `regr`

- Aufruf
- Haupttabelle, s. unten
- Falls Faktoren vorkommen, folgen ihre geschätzten Koeffiz.
- Tabelle mit
  - St.dev.error =  $\hat{\sigma}$ , Freiheitsgr. der Residuen
  - Multiple  $R^2$  und Adjusted R-squared
  - F-statistic: Gesamttest, mit Freiheitsgraden und P-Wert
- Falls `correlation=TRUE`, folgt die Korr.mx der  $\hat{\beta}_j$

h „Haupttabelle“ hat die Spalten

- `coef`: gesch. Koeffiz.  $\hat{\beta}_j$  (ausser Faktoren)
- `stcoef`: standardisierte Koeffiz.  $\hat{\beta}_j^* = \hat{\beta}_j \cdot \text{sd}\langle X^{(j)} \rangle / \text{sd}\langle Y \rangle$ ,
- `R2.x`: Mass  $R_j^2$  für Kollinearität (später)
- `df`: Anzahl Freiheitsgrade, für Faktoren  $> 1$
- `signif`:  $= T / q_{0.975}^{(t_k)}$   
Quotient t-Test-Statistik / Signifikanzgrenze.  
 $\beta_j = 0$  abgelehnt, wenn `signif`  $> 1$ .  
Faktoren: eine analoge Grösse
- `p value` .

i **Modell-Formeln**

$\log_{10}(\text{ersch}) \sim \log_{10}(\text{dist}) + \log_{10}(\text{ladung}) + \text{Stelle}$

**Modell-Formeln allgemein**

Klasse von S-Objekten, charakterisiert durch  $\sim$

Regression: Zielgrösse  $\sim$  Regressor-Terme

$Y \sim X_1 + X_2$  sieht wie Mathematik aus!

Bedeutet in der lin. Regression:

$$Y_i = \beta_0 + \beta_1 X_i^{(1)} + \beta_2 X_i^{(2)} + E_i$$

Syntax hat eigene Regeln, die

nicht immer den math. Zeichen entsprechen!

## j Zielgrösse $\sim$ Regressor-Terme

Terme (rechte Seite):

- quantitative Variable
- Faktor
- Funktion von Eingangsvariablen
- Wechselwirkung zwischen solchen Termen

Linke Seite:

- Funktionen von (einzelnen) Variablen
- mehrere Variable (logistische, multivariate, ... Regr.)
- fehlt für multivariate Verfahren (Hauptkomponenten, ...)

## 3.4

k Viele Funktionen brauchen Formeln.

`plot(formula, ...)` benützt linke Seite vertikal, rechte horiz.

l **Erweiterung:**  $Y \sim X | Z$

– `coplot`

– gemischte Modelle der Varianzanalyse. Ausserdem:

$Y \sim X | Z$ ,  $Y \sim X / Z$ ,  $Y \sim X \%in\% Z$

m **Wo werden Variable gesucht?**

Funktionen, die `formula` als Argument haben, haben auch `data .`

Variable in der Formel sollen Spalten-Namen von `data` sein.

... sonst wird im `search`-Pfad gesucht, also zuerst im akt. `workspace`.

## n Abkürzungen

- $Y \sim ., \text{ data} = t.d$ 
  - steht für „alle anderen Variablen“ (untransformiert)
- Wechselwirkungen:
 
$$X1 * X2 \iff X1 + X2 + X1 : X2$$
- $( X1 + X2 + X3 ) ^ 2 :$   
alle Haupteffekte & alle Wechselwirk. 1. Ord.

- **Komplikation:** Die Zeichen +, \*, ^ haben eine neue Bedeutung.

Manchmal möchte man die ursprüngliche Bedeutung haben.

→ Funktion  $I ( \dots ) : \dots$  nicht als Formel interpretieren!

$I ( X1 ^ 2 ) , I ( X1 * ( X2 - 4 ) )$

(Innerhalb von Funktionen unnötig, z.B.  $+ \text{sqrt} ( X1 ^ 2 + X2 ^ 2 )$ )