Summary Statistics

Summarising, Correlations

Summary Statistics

summary()

`summary` is a generic function used to produce result summaries of the results of various model fitting functions. Its general form is

```
summary(object, ...)
```

where object is an object for which a summary is desired. Object can be a data frame, matrix or a model.

Summary Statistics

summary()

Do it yourself: Use the summary function for data sets `smoking` and `mtcars`

```
> summary(smoking)
> summary(mtcars)
```

`summary()` calculates min, 1st and 3rd quartiles, median, mean and max of each variable in the data frame.

Summary Statistics

sd() and var()

`sd()` and `var()` calculate standard deviation (SD) and the variance of a single vector. Variance-covariance matrix of a data frame can be calculated using the var() function.

Do it yourself: Calculate SD and variance mpg in `mtcars` data set.

```
> sd(mtcars$mpg)
> var(mtcars$mpg)
```

Summary Statistics

cor() and cor.test()

- `cor()` calculates correlation coefficient ("pearson", "kendall", "spearman") for a pair of variables and the correlation matrix for more than two variables.
- The significance of the linear relationship between two variables can be tested using `cor.test()`.

Summary Statistics

cor() and cor.test()

Do it yourself: Calculate correlation coefficient of `mpg` and `wt` from the `mtcars` data set. Is there a significant relationship between these two variables?

```
> cor(mtcars$mpg, mtcars$wt)
> cor.test(mtcars$mpg, mtcars$wt)
```

Do it yourself: Calculate the correlation matrix for the `mtcars` data set.

```
> cor(mtcars)
```

Summary Statistics

A bit advanced: aggregate()

- Sometimes you need to obtain the summary statistics of a *data frame* x grouped by a *list* of grouping elements.
- For example death rate grouped by smokers and non-smokers.
- Use aggregate() for this purpose

```
aggregate(x, by, FUN )
```

Summary Statistics

cor() and cor.test()

Do it yourself: Calculate the mean for the epilepsy grouped by the treatment/no treatment?

```
> aggregate(mtcars, by =
list(row.names(mtcars)), FUN = "mean")
```

or

```
> aggregate(mtcars$mpg, by =
list(row.names(mtcars)), FUN = "mean")
```