

## Auswertungen am Computer mittels SPSS

### Thema der Übung

- Einführung in das statistische Programmpaket SPSS
- Deskriptive Statistik, Transformationen, Testen.
- Korrelation und Regression

SPSS ist eine der gängigsten Software für medizinische Statistik. Sie ist auf jedem Uni-Rechner installiert und wird für wissenschaftliche Arbeiten empfohlen.

### Beschreibung des Datensatzes

Duchenne Muscular Dystrophy Data (from Rencher, 1995, Ex. 5.17)

The observations correspond to blood samples on patients collected in a project to develop a screening program for female relatives of boys with DMD. The program's goal was to inform a woman of her chances of being a carrier based on serum markers as well as her family pedigree. Another question of interest is whether age and season should be taken into account. Enzyme levels were measured in known carriers (n=34) and in a group of non-carriers (n=39).

### **Zusatz-Info:** *Duchenne muscular dystrophy*

*Duchenne muscular dystrophy (DMD) is one of more than 20 types of muscular dystrophies. All muscular dystrophies are due to faults in genes. They cause progressive muscle weakness because muscle cells break down and are gradually lost. The Duchenne type is named after Dr Duchenne de Boulogne who worked in Paris in the mid-19th century. He was one of the first to study muscular dystrophies.*

*The gene for DMD, found on the X chromosome, encodes a large proteindystrophin. Dystrophin is required inside muscle cells for structural support; it is thought to strengthen muscle cells by anchoring elements of the internal cytoskeleton to the surface membrane. Without it, the cell membrane becomes permeable, so that extracellular components enter the cell, increasing the internal pressure until the muscle cell „explodes“ and dies. The subsequent immune response can add to the damage. Unfortunately no cure has yet been discovered.*

*Since Duchenne muscular dystrophy is caused by an X-linked gene mainly boys are affected but their mothers may be carriers. Each subsequent son of a carrier has a 50:50 chance of being affected and each daughter has a 50:50 chance of being a carrier herself. A small number of female carriers of the gene have a mild degree of muscle weakness themselves and are then known as „manifesting carriers“.*

## Daten und Datei

Folgende Variablen sind in Stichproben von Nicht-Trägern (non-carrier) und Trägern (carrier) gemessen worden:

Group:	noncarrier = 1, carrier = 2	Age:	age
Month:	month of measurement	Crea:	creatine kinase
Hemo:	hemopexin	Lact:	lactate dehydrogenase
Pyru:	pyruvate kinase		

Die Messwerte sind in einem Excel-File (muscdys.xls) gespeichert. Excel-Dateien kann man in jede gängige Statistik-Software einlesen. Beim Erstellen einer Excel-Tabelle müssen die Daten sinnvoll strukturiert werden. Standardmässig wählt man für jeden Patienten eine Zeile und für jede Variable (Messung) eine Spalte.

## Fragestellung

Die zentrale Frage ist, ob einer oder mehrere der vier oben genannten Serum-Parameter geeignet sind für ein Screening der Duchenne Muskeldystrophie. Diese Frage kann man einerseits mit Hilfe der deskriptiven Statistik, andererseits mit Tests angehen. Wir werden nur die Variablen Crea und Pyru untersuchen.

### 1. Deskriptive Statistik

Eine Variable diskriminiert Träger und Nicht-Träger gut, wenn die beiden Stichproben wenig überlappen und schlecht, wenn sie ganz stark überlappen.

Tip: Mittels Box-Plot kann man visuell rasch beurteilen, ob zwei Stichproben stark überlappen. Was stellt man fest?

Danach möchte man diese Unterschiede genauer mittels statistischer Kennwerte analysieren.

Zu Beginn einer Datenanalyse muss geprüft werden, ob Ausreisser in den Daten sind und ob die Daten approximativ normalverteilt sind. Auch dazu eignet sich die deskriptive Statistik. Oft hat man bei Laborparametern eine rechtsschiefe Verteilung (asymmetrisch und gegen rechts gestreckt, siehe Histogramm oder Box-Plot). Es lohnt sich dann zu evaluieren, ob eine log- oder Wurzeltransformation zu einer approximativen Normalverteilung führt. Wie verhält es sich in unserem Beispiel?

### 2. Auf Unterschiede testen

Wenn man im 1. Schritt Unterschiede zwischen Trägern und Nicht-Trägern festgestellt hat, muss man diese auf statistische Signifikanz prüfen - es könnte ja auch ein Zufallsbefund sein. Welche statistischen Tests kommen für welche Situationen in Frage? Interpretieren Sie die Ergebnisse!

### 3. Korrelationen

Hängen die Serum-Variablen vom Alter ab? Untersuchen sie dies zunächst visuell. Muss man den Einfluss für jede Gruppe getrennt untersuchen?

(Hinweis: Heterogenitätskorrelation)