# SPSS Grundlagen online

Teil 3

# 1 Korrelation

Untersucht, ob ein kausaler Zusammenhang zwischen zwei Variablen besteht.
Die Variablen auf Normalverteilung untersuchen:

wenn ja: Pearsonscher Korrelationskoeffizient

wenn nein: Spearman oder Kendalls Tau

*Beispiel*: Es wird ein Zusammenhang zwischen der Vitalkapazität fvc und dem maximalen Ausatemstrom pef vermutet.

Beide Variablen folgen nicht der Normalverteilung.

**Analysieren → Korrelation → Bivariate Korrelationen**

Als Ergebnis erhalten wir als Spearman-Rho-Wert 0,86, was eine starke, positive Korrelation anzeigt, die auf dem 1% Niveau signifikant ist.

# 2 Lineare Regression

Untersucht den funktionalen Zusammenhang (zweier) voneinander unabhängiger Variablen.

*Beispiel:* Wir möchten z.B. anhand des Gewichts der Kinder deren Körpergröße „vorhersagen“. Eine Korrelation zeigt einen starken positiven Zusammenhang dieser beiden Variablen auf (Spearman-rho= 0,923).

**Analysieren → Regression → Linear** ,
abhängige Variable=Körpergröße, unabhängige=Gewicht.

**Modellzusammenfassung**

| Modell | R | R-Quadrat | Korrigiertes R-Quadrat | Standardfehler des Schätzers |
|---|---|---|---|---|
| 1 | .891[a] | .794 | .794 | 6.826 |

a. Einflußvariablen : (Konstante), gewi

Das $R^2$ beträgt 0.794, d.h. fast 80% der Gesamtstreuung der Variable „Körpergröße“ werden durch unser Modell, also die Variable „Gewicht“ erklärt, die übrigen 20% ergeben die Residuen (nicht erklärte Streuung).

# 3 Lineare Regression

Unser Modell ist mit einem p-Wert<0.001 signifikant.

**ANOVA**[a]

| Modell | | Quadratsumme | df | Mittel der Quadrate | F | Sig. |
|---|---|---|---|---|---|---|
| 1 | Regression | 275522.774 | 1 | 275522.774 | 5913.273 | .000[b] |
| | Nicht standardisierte Residuen | 71614.912 | 1537 | 46.594 | | |
| | Gesamt | 347137.686 | 1538 | | | |

a. Abhängige Variable: gross

b. Einflußvariablen : (Konstante), gewi

Ausgehend von den Koeffizienten ergibt sich ein Regressions-modell mit folgenden Werten:
(y = mx + b)
gross = 1.081*gewi + 102.278

**Koeffizienten**[a]

| Modell | | Nicht standardisierte Koeffizienten | | Standardisierte Koeffizienten | T | Sig. |
|---|---|---|---|---|---|---|
| | | RegressionskoeffizientB | Standardfehler | Beta | | |
| 1 | (Konstante) | 102.278 | .536 | | 190.882 | .000 |
| | gewi | 1.081 | .014 | .891 | 76.898 | .000 |

a. Abhängige Variable: gross

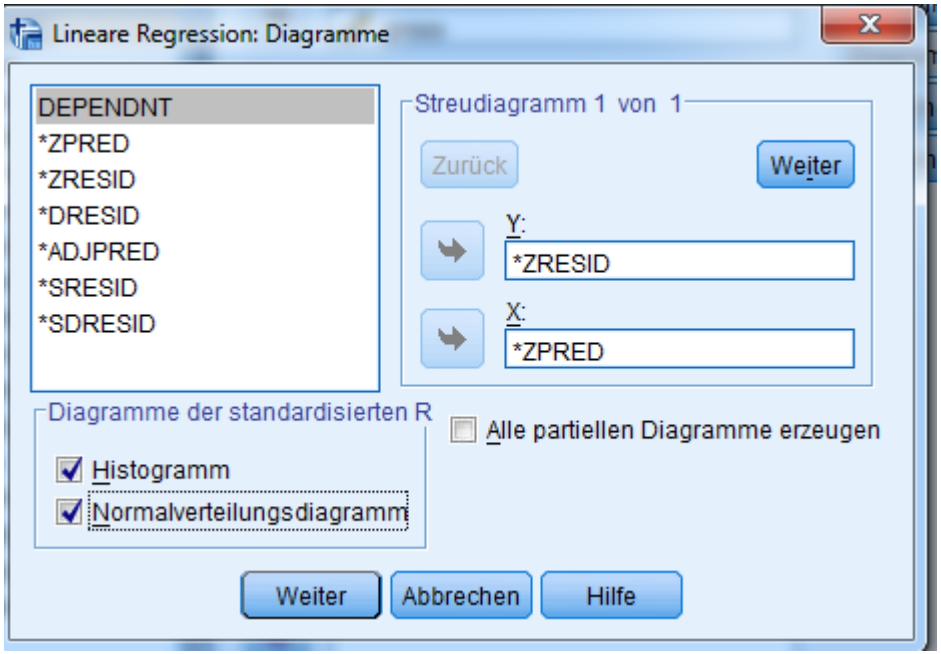
Anschließend muss aber noch untersucht werden, ob das Modell geeignet ist, um diesen Zusammenhang zu beschreiben.

# 4 Lineare Regression

Die Residuen müssen folgende Bedingungen erfüllen:

1. Unabhängig voneinander sein
2. Normalverteilt sein
3. Homogene Varianzen aufweisen

Um dieses zu überprüfen, wählt man bei dem Diagrammfenster der Regression die Schaltfläche „Diagramme“ aus:



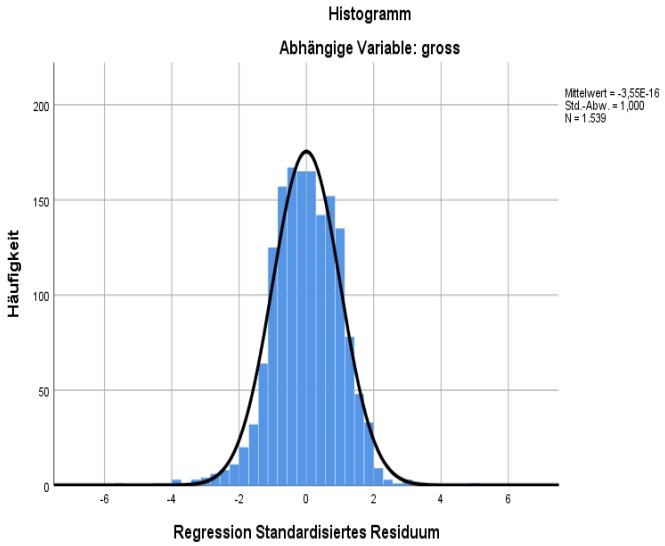

1. Häkchen setzen bei „Histogramm“ und/ oder „Normalverteilungsdiagramm“
2. Streudiagramm der standardisierten vorhergesagten Werte (ZPRED) und standardisierten Residuen (ZRESID)

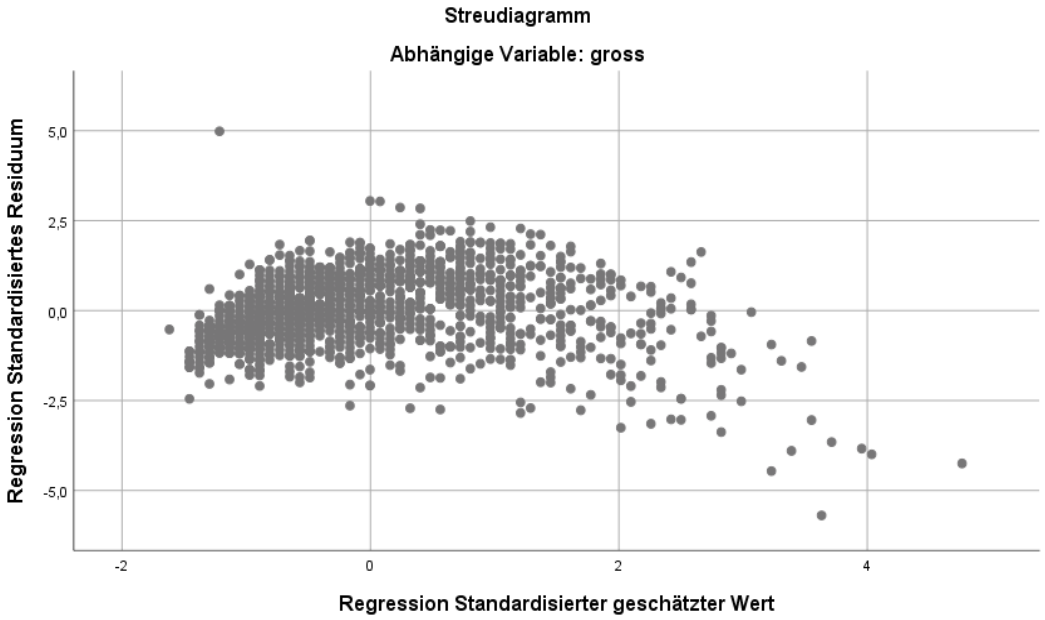
Zusätzlich unter Schaltfläche „Statistiken“ Auswahl des Durbin-Watson Tests.

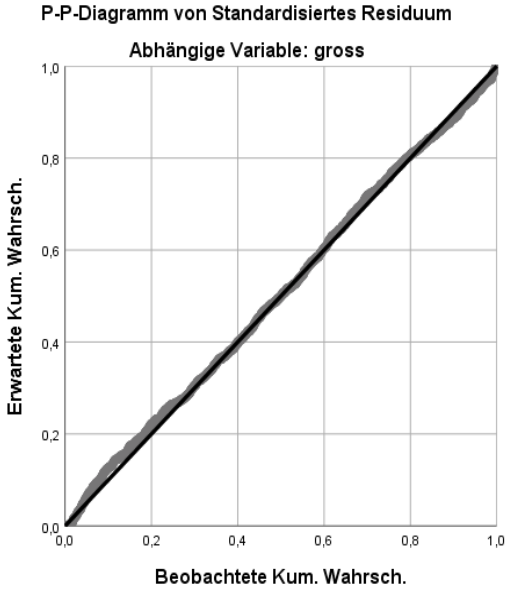
# 5 Lineare Regression

Histogramm sowie Normalverteilungs-diagramm (rechts) zeigen eine Normalverteilung der standardisierten Residuen.

Das Streudiagramm der stand. Residuen gegen die stand. geschätzten Werte (u.l.) zeigt keine extrem regelmäßige Verteilung der Punkte, die auf Autokorrelation hinweisen.



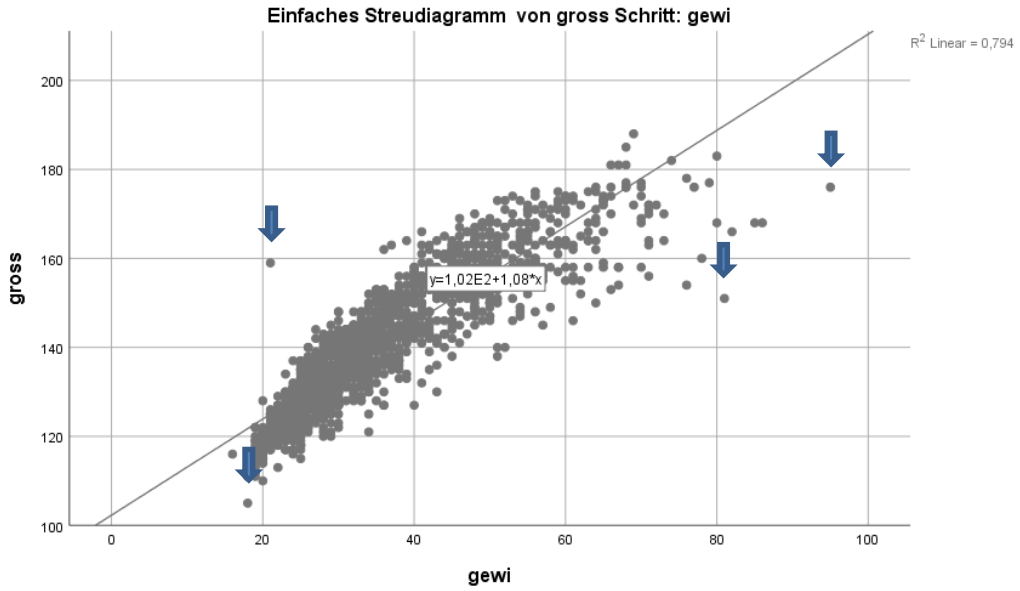



# 6 Lineare Regression

Durbin Watson Test auf Autokorrelation der Werte : 1,68 (Werte zwischen 1,5 und 2,5 werden als unauffällig angesehen).

Aber: -leichte Krümmungstendenz der Punktwolke
-größerer Abstand der Punkte zur Nulllinie bei größeren Werten
-einige Ausreißer (s. Pfeile)

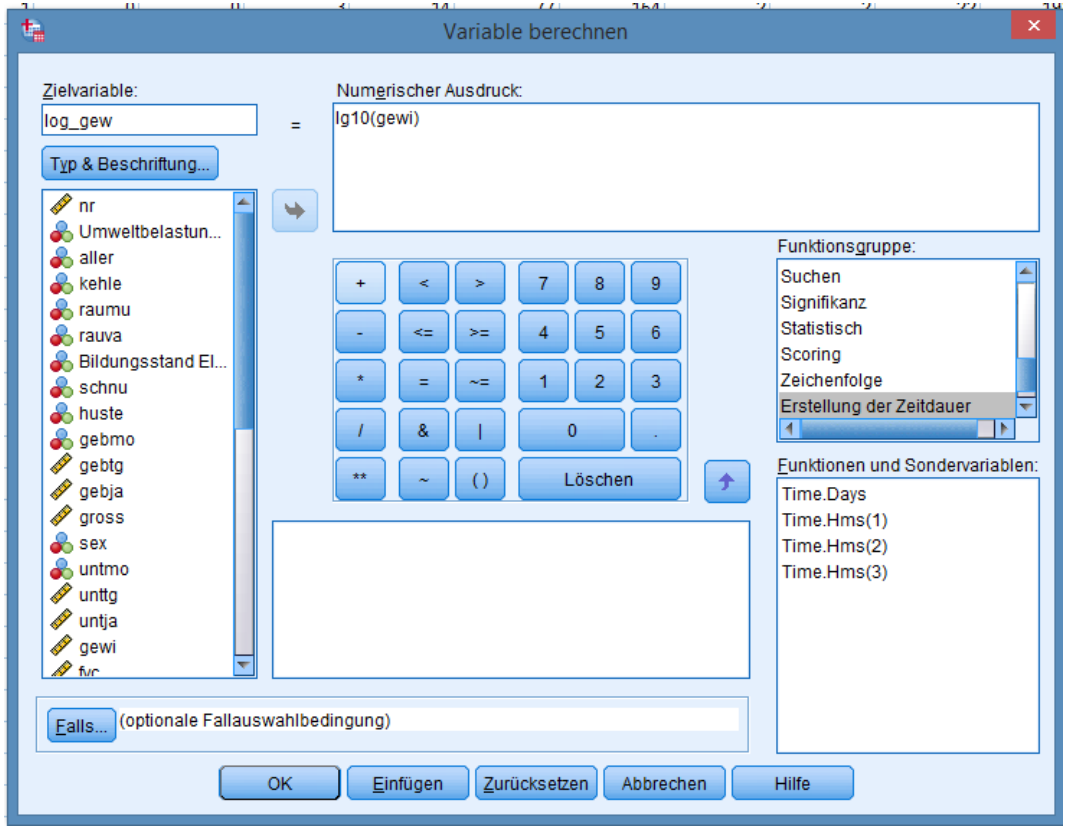


# 7 Lineare Regression

Krümmung eliminieren: über eine Logarithmierung. **Transformieren → Variable berechnen** als Funktion z.B. den Zehnerlogarithmus lg10 nehmen.

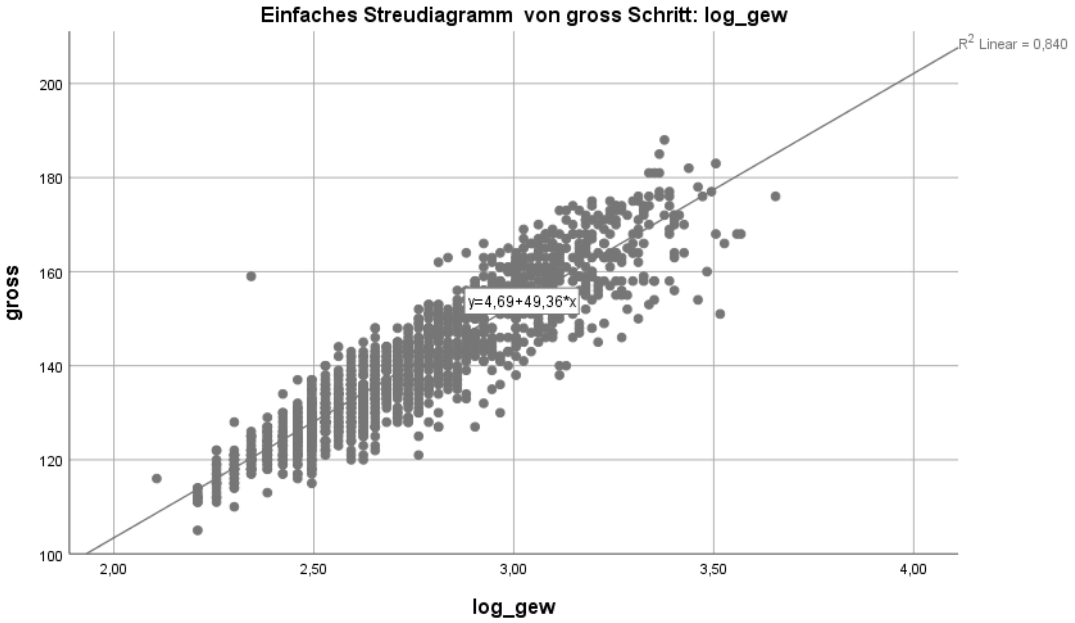
Dann die Regression erneut durchführen mit „log_gew“ als unabhängige Variable.

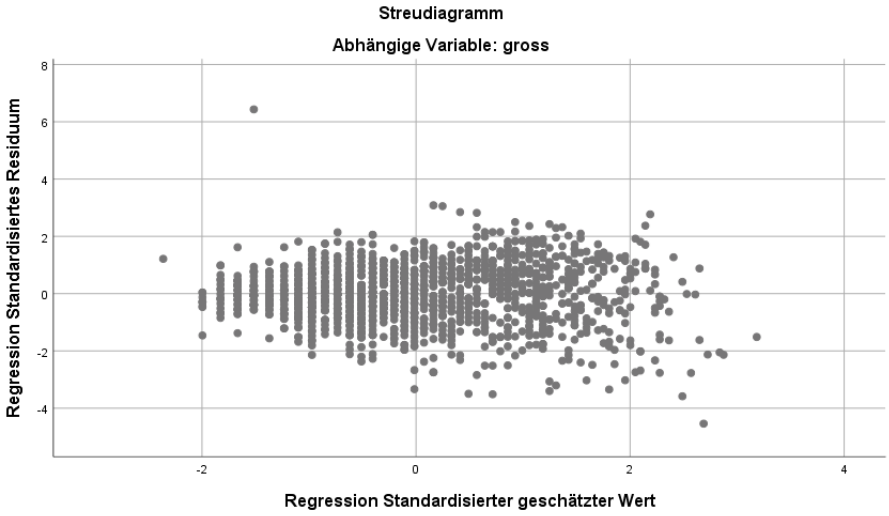
Die größere Streuung bei größeren Kindern lässt sich nicht durch Transformation ändern, dieses ist ein Effekt, der häufig bei „biologischen“ Messungen anzutreffen ist.



# 8 Lineare Regression

Die lineare Regression liefert auch ein deutlich höheres $R^2$ (0,84, vorher 0,794), der Wert des Durbin-Watson Tests liegt mit 1,766 auch im unauffälligen Bereich und das Streudiagramm der standardisierten




Residuen gegen die standardisierten geschätzten Werte zeigt auch eine relativ gleichmäßige Punktwolke (linkes Bild)

# 9 Lineare Regression

*Ausreißer überprüfen*:
z.B. in dem nicht linearisierten Diagramm die visuell auffälligen Punkte heraus suchen:

Diagramm doppelt klicken, um den Diagrammeditor zu öffnen

**Elemente → Datenbeschriftungsmodus,** das erscheinende Viereck auf den betreffenden Punkt setzen und klicken, es wird die Zeilennummer des Punktes im Datensatz angezeigt.

Beim Durchführen der Regression→ Schaltfläche „Speichern“, Häkchen bei „Distanzen“ setzen, alle drei Werte (Cook, Mahalanobis, Hebelwerte) werden im Datenblatt gespeichert und dienen zur Identifikation extremer Werte in der Regression.

# Übung 4

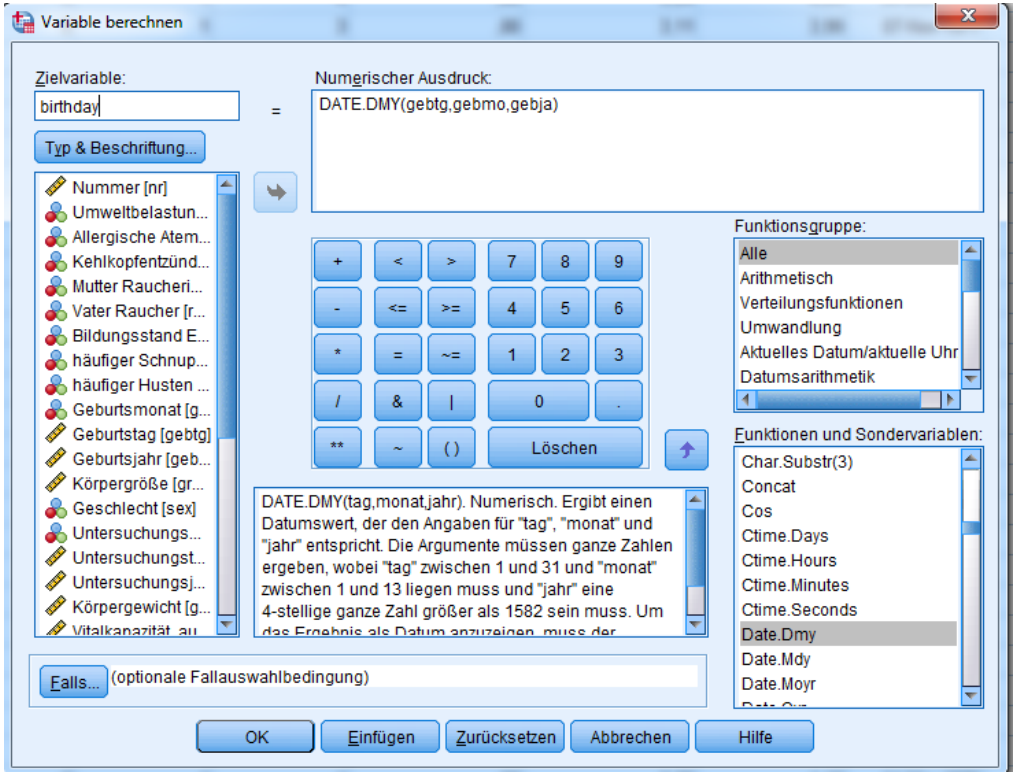
Zum Datensatz „atemwege“:

1) Die Vermutung liegt nahe, dass die Körpergröße (gross) eng mit der Vitalkapazität (fvc) zusammenhängt. Zeigen Sie dies durch eine Korrelationsanalyse.
2) Führen Sie eine lineare Regression durch: unabhängige Variable: „gross“, abhängige Variable „fvc“ (die Vitalkapazität anhand der Größe vorhersagen).

# 10 Exkurs Datumserstellung

In der Atemwegs-Datei sind das Geburtsjahr, der Geburtsmonat und der Tag als separate Variablen angegeben. Um daraus eine einzige Variable „birthday“ zu erzeugen, verfährt man wie folgt:

**Transformieren → Variable berechnen** Hier muss man sich in der Funktionsliste (rechts) eine passende Funktion heraus suchen, z.B. Date.Dmy und die Daten aus der Quell-variablenliste eintragen.



# 11 Exkurs Datumserstellung

Ein Blick auf die neue Variable „birthday“ in der Datenansicht zeigt eine wilde Zahlenfolge. Daher muss noch in der Variablenansicht der Datentyp von numerisch auf Datum geändert werden, man erhält dann eine sinnige Datumsangabe!

Es existieren auch Funktionen, die eine umgekehrte Extraktion von einzelnen Daten zulassen!