

Multivariate Statistik

SPSS-Befehle

Deskriptive Statistik

- *Analysieren -> Deskriptive Statistiken -> Häufigkeiten...*
- Übertragen Sie die zu untersuchende Variable in das Feld *Variablen(n) Statistiken*

- Markieren Sie

Quartile, Mittelwert, Median, Modalwert, Standardabweichung, Varianz, Bereich, Minimum, Maximum, Schiefe, Kurtosis

Weiter

- *Diagramme*

Markieren Sie

Histogramme: Normalverteilungskurve im Histogramm anzeigen

Weiter

- *OK.*

Test auf Normalität

- ***Analysieren -> Deskriptive Statistiken -> Explorative Datenanalyse...***

Übertragen Sie die Variable, die auf Normalität untersucht werden soll, in das Feld ***Abhängige Variablen***.

Statistiken

- ***Weiter***
- ***Diagramme***
Keine, Normalverteilungsdiagramm mit Tests
Weiter
- ***OK***.

In der Tabelle ***Test auf Normalverteilung*** gibt es zwei Tests:

- Kolmogorov-Smirnov-Test (für größere Datendateien)
- Shairo-Wilk (für kleinere Datendateien, z.B. etwa mit 50 Datensätzen)

Für $Sig > \alpha$ (bei jedem Test) wird die Hypothese, dass die Variable normalverteilt ist, bestätigt.

Bivariate Korrelation und Regression

Streudiagramm anfordern

- ***Grafik -> Alte Dialogfelder -> Streu-Punktdiagramm...***
- ***Einfaches Streudiagramm
Definieren***
- **Bringe die abhängige Variable in das Feld *Y-Achse* und die unabhängige Variable in das Feld *X-Achse*.**
Titel
Weiter
- ***OK.***

Bivariate Korrelation

- ***Analysieren -> Korrelation -> Bivariat...***
- **Bringe die beiden Variablen in das Feld *Variablen*
Markiere *Pearson*
*OK***

Einfache Lineare Regression

- ***Analyse -> Regression -> Linear***
- **Übertragen Sie die abhängige Variable in das Eingabefeld *Abhängige Variable* und die unabhängige Variable in das Eingabefeld *Unabhängige Variable*

*Statistiken...***
- ***Konfidenzintervalle, Kovarianzmatrix, Deskriptive Statistik, Teil- und partielle Korrelationen und Kollinearitätsdiagnose.***

Weiter.

- **Diagramme**

Übertragen Sie die Variable *ZRESID in das Feld *Y* und die Variable *ZPRED in das Feld *X* und wählen Sie die Optionen **Histogramm**, **Alle partiellen Diagramme erzeugen** und **Normalverteilungsdiagramm**.

Weiter

- **Speichern**

Wählen Sie bei

Vorhersagewerten: Nicht standardisiert

Residuen: Nicht standardisiert, Studentisiert

Vorhersageintervallen: Mittelwert, Individuell

Weiter

- **OK**

Multiple Lineare Regression

- **Analysieren -> Regression -> Linear...**

- Übertragen Sie die abhängige Variable in das Eingabefeld **Abhängige Variable** und die unabhängigen Variablen in das Eingabefeld **Unabhängige**.

Statistiken...

(Im Dialogfeld **Lineare Regression: Statistiken** sind **Schätzer** und **Anpassungsgüte des Modells** voreingestellt.)

Konfidenzintervalle, Kovarianzmatrix, Deskriptive Statistik, Teil- und partielle Korrelationen und **Kollinearitätsdiagnose**.

Weiter.

- **Diagramme**

Übertragen Sie die Variable *ZRESID in das Feld *Y* und die Variable *ZPRED in das Feld *X* und wählen Sie die Optionen **Histogramm**, **Alle partiellen Diagramme erzeugen** und **Normalverteilungsdiagramm**.

Weiter

- **Speichern**

⟨ Durch Klicken auf die Schaltfläche **Speichern** lassen sich dann eine ganze Reihe im Zusammenhang mit einer Regressionsgleichung berechenbarer Variablen anfordern und zu den Variablen des Datensatzes hinzufügen.

Folgende Variablen können dem Datensatz hinzugefügt werden (In Klammern werden die jeweiligen Namen mit ihren Labels aufgeführt):

a) Vorhergesagte Werte:

- **Nicht standardisiert.** Nicht standardisierte Vorhersagewerte y_i^* (PRE_, „Unstandardised Predicted Value“)
- **Standardisiert.** Standardisierte vorhergesagte Werte (ZPR_, „Standardised Predicted Value“)
- **Korrigiert.** Vorhersagewert bei Ausschluss des jeweiligen Falles *i* bei Berechnung der Regressionsgleichung (ADJ_, „Adjusted Predicted Value“)
- **Standardfehler des Mittelwertes.** Standardfehler des mittleren Vorhersagewertes y_i^* (SEP_, „Standard Error of Predicted Value“)

b) Distanzen

- **Mahalanobis.** Dieses Distanzmaß *misst*, wie stark ein Fall vom Durchschnitt der anderen Fälle hinsichtlich der erklärenden Variablen abweicht. Ein hoher Distanzwert für einen Fall *i* signalisiert, dass dieser hinsichtlich der erklärenden Variablen ungewöhnlich ist (Ausreißer) und

damit eventuell einen hohen Einfluss auf die Modellergebnisse hat (MAH_, „Mahalanaobis Distance“).

(Es gibt noch weitere Distanzen, wie z.B. *Cook* und *Leverage*, auf die hier nicht eingegangen wird.)

c) Vorhersageintervalle

- **Mittelwert.** Intervallschätzwerte für das durchschnittliche y_i^* (LMCI und UMCI, „Lower and Upper Mean for Confidence Interval“)
- **Individuell** (LICI und UICI, „Lower and Upper Individual for Confidence Interval“)

d) Residuen

- **Nicht standardisiert.** Die Residualwerte u_i (RES_, „Unstandardised Residual“)
- **Standardisiert.** Standardisierte Residualwerte. Die Residualwerte werden durch ihre Standardabweichung dividiert (ZRE_ „Standardised Residual“)
- **Studentisiert.** Die Residualwerte u_i , dividiert durch den Schätzwert ihrer Standardabweichung, wobei diese je nach der Distanz zwischen den Werten der unabhängigen Variablen des Falls i und dem Mittelwert der unabhängigen Variablen von Fall zu Fall variiert (SRE-, „Studentised Residual“)
- **Ausgeschlossen.** Residualwert bei Ausschluss des jeweiligen Falles i bei Berechnung der Regressionsgleichung (DRE_, „Deleted Residual“)
- **Studentisiert, ausgeschl.** Studentisierte Residuen bei Ausschluss des jeweiligen Falles i bei Berechnung der Regressionsgleichung (SDR_, „Studentised Deleted Residual“)

e) Einflussstatistiken (Maße zur Identifizierung einflussreiche Fälle).

f) Koeffizientenstatistik

- Es werden die Regressionskoeffizienten sowie weitere zu diesen Schätzungsergebnissen gehörende Informationen entweder in die *SPSS*-Arbeitsdatei oder eine *SAV*-Datei ausgegeben.

g) Modellinformationen in XML-Datei exportieren

- Mit dieser Option können Regressionskoeffizienten und (wahlweise) ihre Kovarianzen in eine Datei exportiert werden. }

Wählen Sie **folgende** Optionen:

Bei

Vorhersagewerten: Nicht standardisiert

Residuen: Nicht standardisiert, Studentisiert

Vorhersageintervallen: Mittelwert, Individuell

Weiter

- **OK**

Einfaktorielle Varianzanalyse

- *Analysieren -> Mittelwerte Vergleichen -> Einfaktorielle Varianzanalyse...*
- Bringe die abhängige Variable in das Feld *Abhängige Variablen* und den Faktor in das Feld *Faktor*.

Optionen

- Markiere
Deskriptive Statistik, Test auf Homogenität der Varianzen, Diagramm der Mittelwerte

Weiter

- *Post Hoc...*

LSD

Weiter

- *OK*
- *Analysieren -> Deskriptive Statistiken -> Explorative Datenanalyse...*
- Bringe die abhängige Variable in das Feld *Abhängige Variablen* und den Faktor in das Feld *Faktorenliste*.

Diagramme

- Wähle
*Normalverteilungsdiagramm mit Tests
Nicht transformiert*

Weiter

- *OK.*

- **Graf -> Alte Dialogfelder -> Fehlerbalken**

- **Definieren**

Bringe die abhängige Variable in das Feld **Variable** und den Faktor in das Feld **Kategorieachse**.

Bei **Bedeutung der Balken** wähle **Standardfehler des Mittelwertes**.

Setze den **Multiplikator** auf 1.

OK.

Clusteranalyse (K-Means Cluster)

- ***Analysieren -> Klassifizieren -> K-Means-Cluster***
- **Bringe die relevanten Variablen in das Feld *Variablen* und den Fall in das Feld *Fallbeschriftung*.**

Bestimme die *Anzahl der Cluster*.

Optionen

- ***Cluster Informationen für jeden fall***

Weiter

- ***Speichern***

Clusterzugehörigkeit

Weiter

- ***OK.***

(Letzte Aktualisierung: 21.09.2019)