

Hinweise zur 5. Übung

1.] Der Korrelationskoeffizient mißt die lineare Abhängigkeit von zwei Zufallsgrößen. Nichtlineare Zusammenhänge kann man mit ihm nicht erfassen.

a) Es sei X eine symmetrische Zufallsgröße und $Y = a + bX^2$. Zeigen Sie, dass $\rho_{X,Y} = 0$ ist. Wie sieht dann die beste Anpassung von Y durch X aus?

b) Es sei $U \sim U(-1, 1)$ und $Y = 1 - U^2$. Erzeugen Sie 100 Zufallszahlen zu U und Y . Bestimmen Sie den empirischen Korrelationskoeffizienten zwischen diesen Merkmalen und danach die beste quadratische Anpassung.

a) Es sei X eine symmetrische Zufallsgröße meint, daß $E X=0$, dann ist

$$\begin{aligned} Cov(X, Y) &= Cov(X, a + b X^2) = E((X - E X)((a + b X^2) - E(a + b X^2))) \\ &= b E(X^3) - (E X)E(X^2) = 0 \end{aligned}$$

b) Mit $U=RV.UNIFORM(-1,1)$ und $Y = 1 - U^2$ ergibt sich ein Korrelationskoeffizient ρ von 0,25 ?? Die Auflösung des scheinbaren Widerspruches liegt in der zu geringen Zahl der verwendeten Werte von U , die noch nicht "richtig schön" gleichverteilt sind. Verwenden wir 1000 Werte, ergibt sich $\rho=0,049$, also schon eine gute Annäherung an den theoretischen Wert Null von a). Wobei $\rho=0,25$ auch schon eine sehr "schlechte" Korrelation ist.

2.] a) Erzeugen Sie mit Hilfe der Funktionen $RV.NORMAL(MW,\sigma)$ 100 nach $N(0,1000)$ verteilte Zufallszahlen als Variable NN .

b) Erzeugen Sie die Variable Fallnummer mit der Systemvariablen \$casenum.

c) Klassifizieren Sie die Variable NN in 10 Klassen gleicher Breite unter der Variablen NG .

d) Berechnen und zeichnen Sie die empirischen Verteilungsfunktionen von NN und NG , und die "richtige" Verteilungsfunktion der Normalverteilung $N(0, 1000)$ in einem gemeinsamen Bild, und bestimmen Sie den Abstand der Verteilungsfunktionen von NN und $N(0, 1000)$. Hinweis: die Verteilungsfunktionen vieler Verteilungen stehen im Feld -->Berechnen unter der Abkürzung $CDF.name(parameter)$ bereit.

a) Zuerst sind 100 Fälle zu aktivieren, damit kann die Berechnung starten

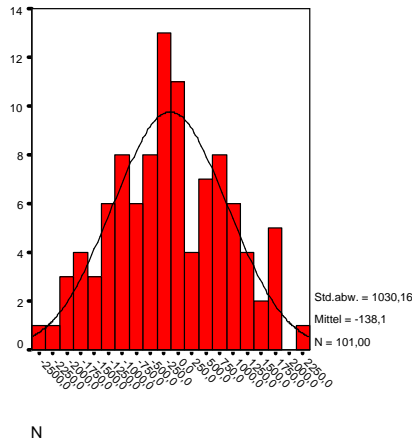
-> Transformieren -> Berechnen, Neue Variable: N, Numerischer Ausdruck: $RV.Normal(0,1000)$

Das Histogramm zeigt, dass 100 Zufallszahlen noch nicht perfekt die Normalverteilung ergeben. Durch Doppelklick kann man es noch weiter bearbeiten, z.B.die Klassenzahl 10 einstellen, so dass Teil c) graphisch gelöst werden kann.

b) ->Transformieren->Berechnen, Neue Variable: Fallnummer, Numerischer Ausdruck: \$casenum

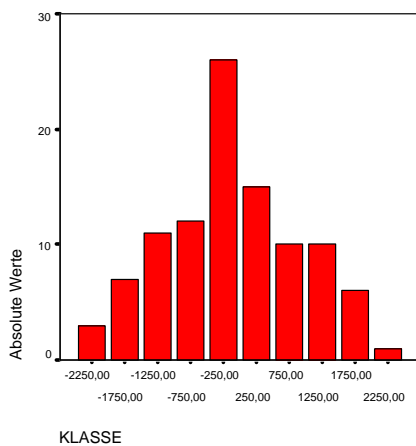
c) Zur Klassifizierungsproblematik siehe vorangehende Serien.

Um die mühselige Einteilung in 10 Klassen per Hand zu umgehen, kann man sich eine formalisierte Klassenzuweisung mittels Modulo-Rechnung überlegen:



1. Berechne Klasse = $N - \text{MOD}(N, 500) + 250$
2. Berechne (Falls $N < 0$) Klasse = Klasse - 500

Das erzeugt eine Klasseneinteilung in (hier 10 oder 11) Klassen der Breite 500. Das Aggregieren geht unter `-> Daten` mit *N* als Aggregierungsvariable und *Klasse* als Breakvariable. Wenn man noch die Fallzahl je Breakgruppe anklickt, erhält man in der neuen Datentabelle die kumulierten Anzahlen je Klasse, siehe Balkendiagramm.



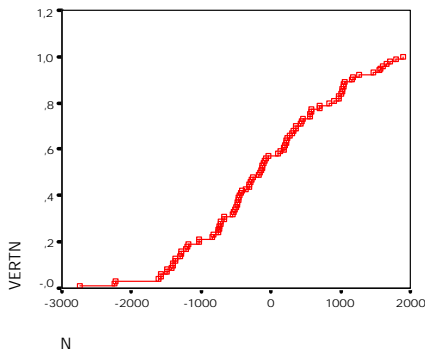
d) Verteilungsfunktionen: Zuerst sind die Variablen *N* zu sortieren. Dann kann wieder die empirische Verteilungsfunktion berechnet werden: `-> Transformieren -> Berechnen`, Neue Variable: Verteilung von N, Numerischer Ausdruck: $\$casenum/100$. `-> Grafik -> Streuplot` gibt die empirische Verteilung.

Analog zu $N(0,1000)$: Die Variablen *N* sind schon sortiert. Dann `-> Transformieren -> Berechnen`, Neue Variable: Verteilung von $N(0,1000)$, Numerischer Ausdruck: $\text{CDF.NORMAL}(N,0,1000)$ `-> Grafik -> Streuplot` gibt die exakte Verteilungsfunktion.

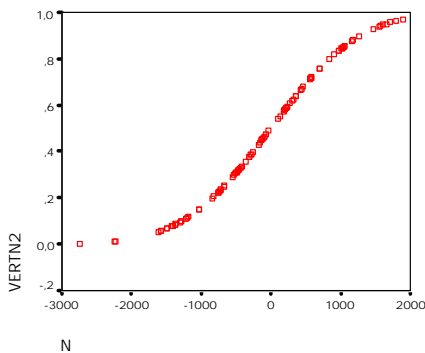
Bemerkung: Die Verteilungsfunktionen vieler Verteilungen stehen im Feld `--> Berechnen` unter der Abkürzung $\text{CDF.name(parameter)}$ bereit.

Mit den Befehlsdateien `spss052.sps` und `klass100n.sps` kann diese Aufgabe gelöst werden.

Verteilungsfunktion von N



Verteilungsfunktion von (N(0,1000))



3.] Erstellen Sie in SPSS die Variable ARGUMENT x mit den Werten $x_j = j/20$ mit $j = -100, \dots, 100$. Zeichnen Sie den Graph der Funktionen *SIN*, *COS*, *ARCTAN*, *ARCSIN*, und *EXP* mit dem Grafkbefehl Streudiagramm.

Es sind 201 Zeilen im SPSS-Datenfenster zu aktivieren. Mit -> Transformieren -> Berechnen,
Neue Variable: x, Numerischer Ausdruck: ($\$casenum-101$)/20, ergibt sich die Achseneinteilung von $[-5,5]$.

Die Funktionswerte sind je:

-> Transformieren -> Berechnen,

Neue Variable: SIN

Numerischer Ausdruck: SIN(x)

Neue Variable: COS

Numerischer Ausdruck: COS(x)

Neue Variable: ARCTAN

Numerischer Ausdruck: ARTAN(x)

Neue Variable: ARCSIN

Numerischer Ausdruck: ARSIN(x)

Neue Variable: EXP

Numerischer Ausdruck: EXP(x)

In -> Grafik-> Streuplot sind dann entsprechende Variablenpaare einzubringen.
oder man verwende hier auch

-> Grafik-> Linienplot, da die x-Werte alle äquidistant sind.

Beim ARCSIN stellt sich heraus, dass der Definitionsbereich nur $[-1,1]$ ist. Das ist kor-

rekt! Andere Werte sind nicht zugelassen. Im Variablenfeld stehen ansonsten “missing values”, die der Graphikbefehl unterdrückt.
(Mit der Befehlsdatei spss053.sps kann diese Aufgabe gelöst werden.)