LÖSUNG 3A

- Mit den Berechnungsfunktionen LG10(?) und SQRT(?) in "Transformieren", "Berechnen" können logarithmierte Werte sowie die Quadratwurzel von EINK bestimmt werden: LOG10EINK = LG10(EINK) und SQRTEINK = SQRT(EINK).

- Zur Prüfung einer Variablen auf Normalverteilung stehen mehrere Möglichkeiten zur Verfügung. Man sollte sich in erster Linie nicht auf statistische Tests stützen, sondern deskriptive Analysen, insbesondere auch grafische Darstellungen nutzen. Im Folgenden: Histogramme, Bewertung von Schiefe und Kurtosis, statistische Tests auf Normalverteilung, QQ-Diagramme und trendbereinigte QQ-Diagramme.

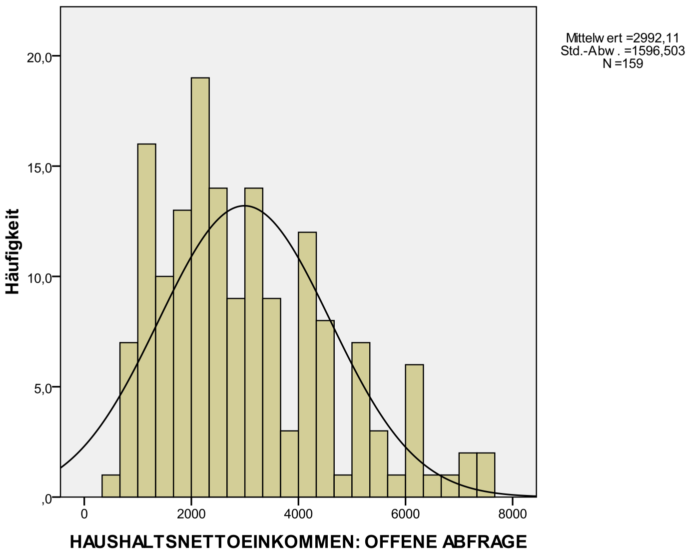
Histogramme:

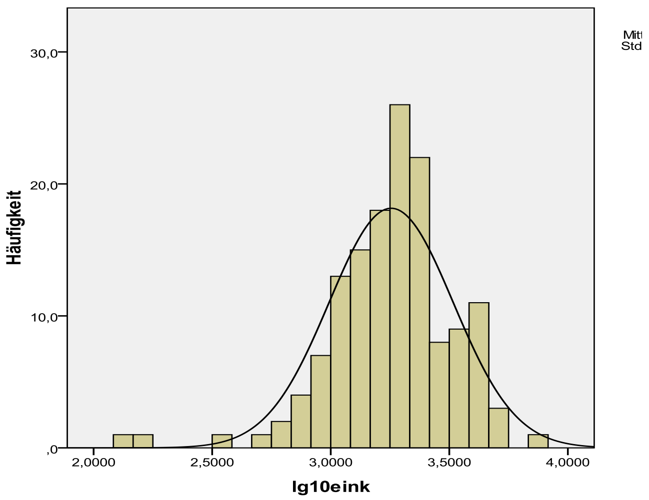
- Erstellen von Histogrammen mit überlagerter Normalverteilungskurve: „Grafik", „Diagrammerstellung…", durch Doppelklicken auf das Symbol für ein „Einfaches Histogramm" dieses in die Diagrammvorschau übertragen. EINK (bzw. LG10EINK bzw. SQRTEINK) auf „X-Achse?" ziehen. In der Dialogbox „Elementeigenschaften" in „Eigenschaften bearbeiten von:" „Balken1" markieren und „Normalverteilungskurve anzeigen" wählen.

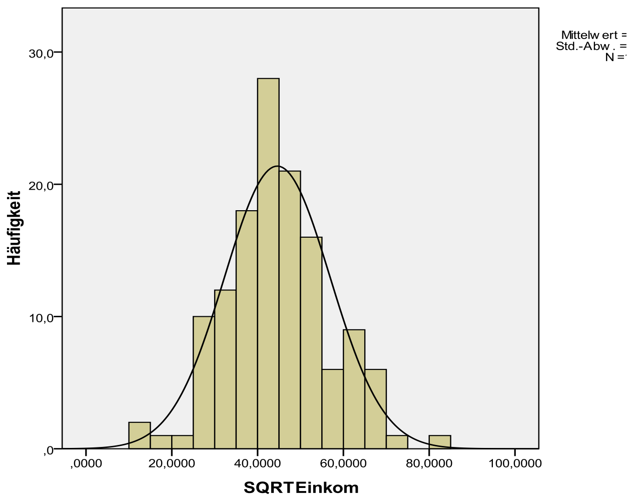
- "Histogramme" mit überlagerter Normalverteilungskurve sind auch mit „Analysieren", "Deskriptive Statistiken", „Häufigkeiten…" zu erstellen. Dann müssen im Untermenü „Diagramme", der Optionsschalter „Histogramme" und das Auswahlkästchen „Normalverteilungskurve im Histogramm anzeigen" markiert werden.

- Es ist ersichtlich, dass EINK typischerweise eine linkssteile Verteilung hat.

- Aus einem Vergleich der Histogramme für EINK, LOG10EINK und SQRTEINK wird deutlich, dass durch beide Transformationen eine Annäherung an die Normalverteilung erreicht wird. Die Annäherung von SQRTEINK ist etwas besser als für LOG10EINK.







Schiefe und Kurtosis:

- Berechnen: „Analysieren“, "Deskriptive Statistiken", "Häufigkeiten…“ und die Variable EINK, LG10EINK und SQRTEINK in „Variable(n)“ übertragen. Durch Klicken auf die Schaltfläche „Statistiken…“ die entsprechende Dialogbox öffnen und dort „Schiefe“ und „Kurtosis“ markieren. Zusätzlich sind im Beispiel auch der Mittelwert und der Zentralwert angefordert worden, um auch aus dem Vergleich dieser beiden Lagemaße Aussagen über eine Abweichung von der Symmetrie zu gewinnen.

  Soll keine Häufigkeitstabellen ausgeben werden, muss man auf der Dialogbox „Häufigkeiten“ „Häufigkeitstabellen anzeigen“ ausschalten.

Statistiken

		eink	lg10eink	sqrteink
N	Gültig	143	143	143
	Fehlend	158	158	158
Mittelwert		2096,78	3,254804	44,176757
Median		1900,00	3,278754	43,588989
Schiefe		1,186	-1,079	,302
Standardfehler der Schiefe		,203	,203	,203
Kurtosis		2,000	3,497	,464
Standardfehler der Kurtosis		,403	,403	,403

- Die Schiefemaße von SQRTEINK liegt am nächsten bei 0 und signalisiert die größte Nähe zur Symmetrie der Verteilung.

  Das Schiefemaß von LG10EINK ist negativ und zeigt damit eine rechtssteile Verteilung an. Dies ist auch aus dem Histogramm für LG10EINK zu erkennen. Auch aus dem Vergleich des Mittelwerts mit dem Median wird dies deutlich: für LG10EINK ist Median > Mittelwert.

  Ergänzend kann man sich auch ein 95-%-Konfidenzintervall für das Maß der Schiefe unter Nutzung des Standardfehlers (SF) in der Tabellenausgabe berechnen (zu Konfidenzintervallen s. Kapitel 8.4). Es zeigt sich, dass für EINK der Konfidenzbereich den Wert 0 nicht einschließt und daher EINK nach dieser Betrachtung nicht symmetrisch ist.

Schiefe			
	Schiefe – 1,96*SF	Schiefe	Schiefe + 1,96*SF
EINK	0,778	1,186	1,584
LG10EINK	-1,477	-1,079	0,681
SQRTEINK	-0,100	0,302	0,700

- Das Wölbungsmaß Kurtosis ist für die Verteilungen der drei Variablen positiv: in allen drei Fällen sind die Wölbungen der Verteilungen spitzer als die einer Normalverteilung. Hinsichtlich der Wölbung entspricht die Verteilung von SQRTEINK am ehesten einer Normalverteilung.

  In der nachfolgenden Tabelle sind 95-%-Konfidenzintervalle für die Kurtosis aufgeführt. Es zeigt sich, dass nur für SQRTEINK der Konfidenzbereich den Wert 0 einschließt. Nach dieser Betrachtung entspricht nur die Verteilung von SQRTEINK der Wölbung einer Normalverteilung.

Kurtosis			
	Kurtosis – 1,96*SF	Kurtosis	Kurtosis + 1,96*SF
EINK	1,210	2,00	2,790
LG10EINK	2,707	3,497	4,287
SQRTEINK	-0,326	0,464	1,254

- Zusammenfassend kann man festhalten: die Analyse der Daten legt nahe, dass die Verteilung von SQRTEINK einer Normalverteilung entspricht.

Statistisches Testen:

- Mit „Analysieren“, "Deskriptive Statistiken", "Explorative Datenanalyse…" und dort der "Schaltfläche Diagramme" und der Auswahl von "Normalverteilungsdiagramm mit Tests" können statistische Tests auf Normalverteilung für EINK, LG10EINK und SQRTEINK aufgerufen werden (Test nach Kolmogorov-Smirnov mit Signifikanz nach Lilliefors sowie nach Shapiro-Wilk, s. Kapitel 9.3.2). Verbunden damit werden zwei Grafiken erstellt, die auf den Vergleich mit einer Normalverteilung abstellen (QQ-Diagramm und trendbereinigtes QQ-Diagramm). (s. Kapitel 9.3.2 und Kapitel 29.17).

  Da der Test nach Shapiro-Wilk wegen größerer Trennschärfe gegenüber dem nach Kolmogorov-Smirnov mit Signifikanz nach Lilliefors im Allgemeinen zu bevorzugen ist, sollte man diesen verwenden (D’Agostino in seinem Beitrag „Tests for Normal Distribution” in: R. B. D’Agostino und M. Stephen (Hrg.), Goodness-of-fit Techniques: “The Kolmogorov-Smirnov test is only a historical curiosity. It should never be used.”)

- Geht man von einem Signifikanzniveau von 5 % aus, so kommt der Shapiro-Wilk-Test zum Ergebnis, dass nur für SQRTEINK die Hypothese H_0 (die Variable entspricht einer Normalverteilung) nicht abgelehnt werden kann. Dieses ergibt sich daraus, dass der in der Tabelle gegebene Wert für Signifikanz größer ist als das Signifikanzniveau in Höhe von 5 % (0,13 > 0,05).

  Insofern entspricht dieses Ergebnis der oben gemachten Beurteilung.

- Man darf bei der Verwendung von statistischen Tests nicht vergessen, dass die Ergebnisse vom Stichprobenumfang abhängen. Bei kleinen Stichproben führen Normalverteilungstests meistens zur Annahme von $H_{0,}$ weil die Trennschärfe des Tests nicht ausreicht, um zu entscheiden, ob die wenigen Fälle aus einer Normalverteilung stammen. Bei großen Stichproben werden schon kleine Abweichungen von der Normalverteilung als signifikant ausgegeben.

- Kleine Abweichungen von der Normalverteilung sind aber keine Einschränkung für die Anwendung von z.B. eines t-Tests oder einer ANOVA. Daher sollte man sich bei der Prüfung auf Normalverteilung in erster Linie nicht auf statistische Tests, sondern auf die anderen Analysen stützen.

Tests auf Normalverteilung

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistik	df	Signifikanz	Statistik	df	Signifikanz
eink BEFR.: MONATLICHES NETTOEINKOMMEN	,128	143	,000	,923	143	,000
lg10eink	,071	143	,073	,940	143	,000
sqrteink	,077	143	,036	,985	143	,130

a. Signifikanzkorrektur nach Lilliefors

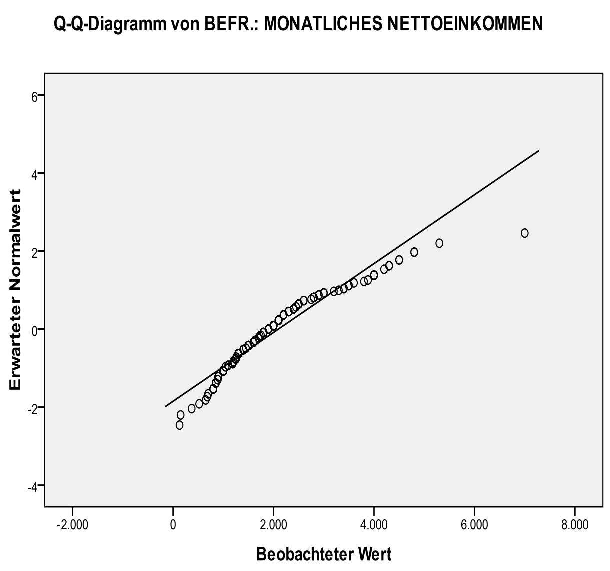
QQ-Diagramme und trendbereinigte QQ-Diagramme:

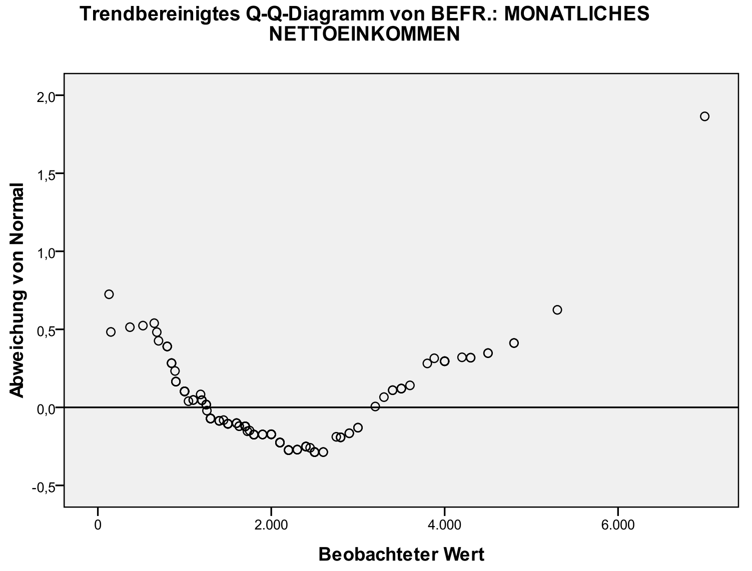
- Mit der Ausgabe der Tests auf Normalverteilung werden auch QQ-Diagramme und trendbereinigte QQ-Diagramme ausgegeben. Diese können aber auch mit dem Menü „Diagramme", „Diagrammerstellung" erstellt werden (s. Kapitel 29.17).

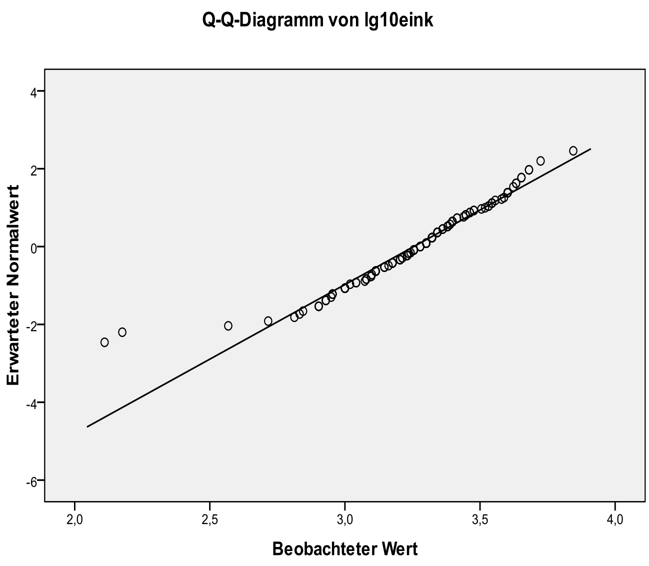
  Im QQ-Diagramm zeigen die Abweichungen der Punkte von der Geraden (diese entspricht bei der QQ-Verteilung den bei einer Normalverteilung zu erwartenden Werten), wie stark die Abweichung von der Normalverteilung ist. Man kann erkennen, dass mit den beiden Transformationen eine Annäherung an die Normalverteilung erzielt wird. Welche der beiden Transformationen besser ist, ist aber nicht so eindeutig.

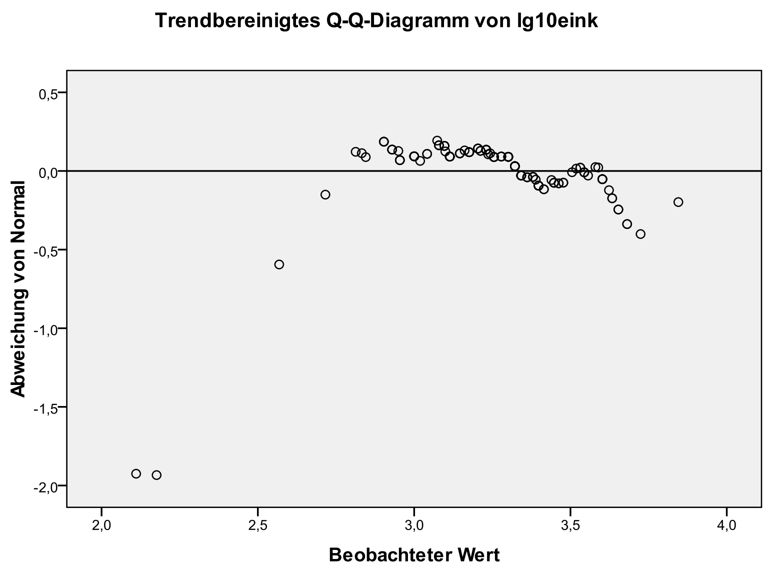
- Im trendbereinigten QQ-Diagramm werden die Punkte auf der Geraden im QQ-Diagramm als Basiswerte genommen. Abweichungen der Punkte im QQ-Diagramm erscheinen im trendbereinigten QQ-Diagramm nun als Abweichungen von einer Nulllinie. Insofern handelt es sich nur um eine andere Darstellungsform des QQ-Diagramms. Die Abweichungen werden noch pointierter sichtbar.

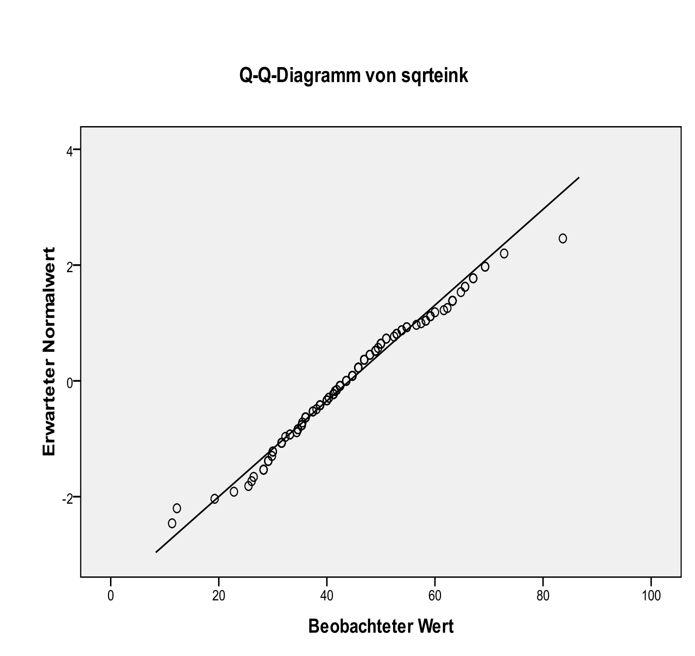
- Die Diagramme sind für die Beurteilung der Abweichung von der Normalverteilung besser geeignet als statistische Tests, weil sie den Grad der Abweichung unabhängig vom Stichprobenumfang zeigen. Generell muss man davon ausgehen, dass eine perfekte Normalverteilung selten gegeben ist und die statistischen Tests, die eine Normalverteilung voraussetzen, bei nicht zu gravierender Abweichung hinreichend zuverlässig sind. Leider fehlt es an Konventionen darüber, welcher Grad der Abweichung für die Tests toleriert werden kann. Bei großen Abweichungen sollte man eventuell Datentransformationen zur Annäherung an die Normalverteilung verwenden.

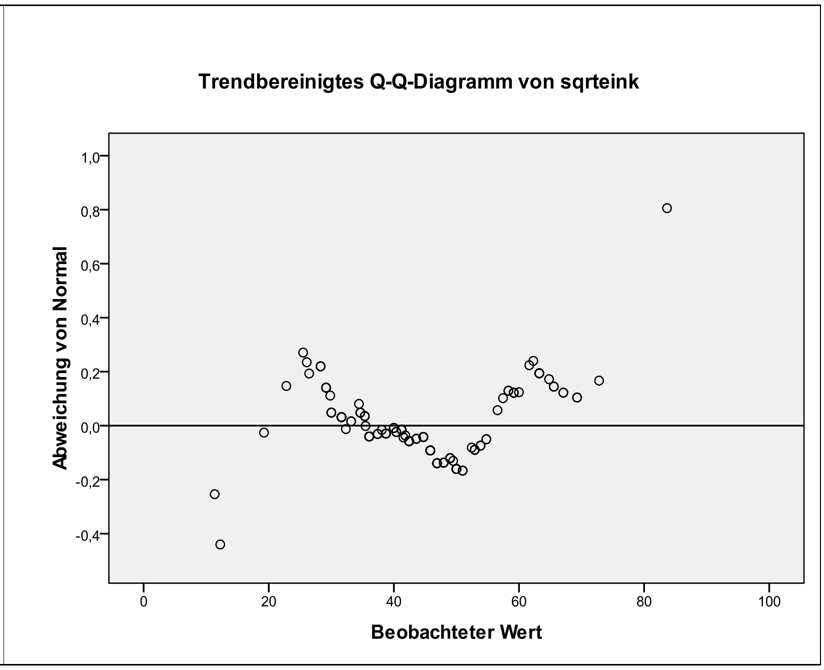

Q-Q-Diagramm von BEFR.: MONATLICHES NETTOEINKOMMEN
Erwarteter Normalwert
Beobachteter Wert


Trendbereinigtes Q-Q-Diagramm von BEFR.: MONATLICHES NETTOEINKOMMEN
Abweichung von Normal
Beobachteter Wert


Q-Q-Diagramm von lg10eink
Erwarteter Normalwert
Beobachteter Wert


Trendbereinigtes Q-Q-Diagramm von lg10eink
Abweichung von Normal
Beobachteter Wert


Q-Q-Diagramm von sqrteink
Erwarteter Normalwert
4
2
0
-2
0
20
40
60
80
100
Beobachteter Wert


Trendbereinigtes Q-Q-Diagramm von sqrteink
Abweichung von Normal
1,0
0,8
0,6
0,4
0,2
0,0
-0,2
-0,4
0
20
40
60
80
100
Beobachteter Wert