

Statistik-Klausur vom 25. April 2003

Bearbeitungszeit: 120 Minuten

Aufgabe 1

Performance-Vergleich weltweit angelegter Aktienfonds im Drei-Jahres-Vergleich vom 1.1.1999 bis 31.12.2001

Investmentfonds	Astra ¹⁾	FMM ¹⁾	Templeton Global Smaller Companies ²⁾
Wertpapier- kenn-Nummer	977 000	847 811	971 656
Wertzuwachs je Fonds-Anteilschein im Kalenderjahr ... (in %)			
1999	+12,5	+14,6	+27,5
2000	+10,0	+3,6	+2,3
2001	+6,5	+4,2	+2,2

1) Geführt von dem Vermögensverwalter Dr. Jens Erhardt

2) Ein Fonds der bekannten US-Investmentgesellschaft Franklin Templeton

Quelle: Euro am Sonntag, 6. Januar 2002, Seite 44

- Um wie viel Prozent lag der Wert eines Anteilscheins des Astra-Fonds am Jahresende 2001 höher als am Jahresanfang 1999?
- Berechnen Sie die durchschnittliche jährliche Wertsteigerung eines Anteilscheins des FMM-Fonds für den Zeitraum vom 1.1.1999 bis zum 31.12.2001.
- Welcher der drei Fonds wies vom 1.1.1999 bis zum 31.12.2001 die beste „Performance“ (=Wertzuwachs) auf? Begründen Sie Ihre Antwort durch Berechnung und Interpretation geeigneter statistischer Maßzahlen.

Lösung:

- a) Astra-Fonds

Jahr	Wert	Steigerungsrate (in %)	Steigerungsfaktor
31.12.1998		-	-
31.12.1999		+12,5	1,125
31.12.2000		+10,0	1,1
31.12.2001		+6,5	1,065

$$1,125 \cdot 1,1 \cdot 1,065 = 1,3179$$

d.h. er war um 31,8% höher.

- b) FMM-Fonds

Jahr	Wert	Steigerungsrate (in %)	Steigerungsfaktor
31.12.1998		-	-
31.12.1999		+14,6	1,146
31.12.2000		+3,6	1,036
31.12.2001		+4,2	1,042

$\sqrt[3]{1,146 \cdot 1,036 \cdot 1,042} = \sqrt[3]{1,2371} = 1,0735$
d.h. sie lag bei durchschnittlich 7,35% pro Jahr.

c) Templeton Global Smaller Companies

Jahr	Wert	Steigerungsrate (in %)	Steigerungsfaktor
31.12.1998		-	-
31.12.1999		+27,5	1,275
31.12.2000		+2,3	1,023
31.12.2001		+2,2	1,022

$1,275 \cdot 1,023 \cdot 1,022 = 1,3341$

d.h. der dritte Fonds hatte im Zeitraum vom 1.1.1999 bis 31.12.2002 mit 33,4% den größten Wertzuwachs gegenüber den beiden anderen Fonds (Wertzuwachs 31,8% bzw. 23,7%)

Aufgabe 2

Sie sind Vertriebs-Assistent bzw. Vertriebs-Assistentin in einem mittelständigen Unternehmen. Ihre Vertriebs-Leiterin möchte wissen, wie weit sie den Preis eines bestimmten Produkts senken muss, um den Absatz dauerhaft auf 1200 Stück zu erhöhen. In den vergangenen vier Quartalen hat sie bereits mit Preisveränderungen experimentiert. Daher liegen folgende Daten vor:

	Preis p (in Euro pro Stück)	abgesetzte Menge x (in Stück)
4. Quartal 2001	50	1 000
1. Quartal 2002	55	900
2. Quartal 2002	45	1 080
3. Quartal 2002	52	970

- Ermitteln Sie aus den vorgegebenen Daten eine lineare Preis-Absatz Funktion, die die abgesetzte Menge x in Abhängigkeit vom Preis p darstellt. Verwenden Sie dazu die Methode der kleinsten Quadrate. Geben Sie die Werte mit jeweils zwei Nachkommastellen an.
- Ihre Vertriebs-Leiterin glaubt nicht, dass sich die Preis-Absatz Funktion des Produkts durch eine lineare Funktion darstellen lässt. Mit welcher statistischen Maßzahl können Sie sie, zumindest bezüglich der vorliegenden Daten, überzeugen? Berechnen Sie die Maßzahl.
- Sie haben Ihre Vertriebs-Leiterin überzeugt. Jetzt möchte diese wissen, welchen Preis sie ansetzen muss, um 1200 Stück des Produkts abzusetzen. Ermitteln Sie die richtige Antwort.
- Ihre Vertriebs-Leiterin befolgt Ihren Rat und setzt den Preis auf den von Ihnen vorgeschlagenen Betrag herunter. Dennoch erreicht sie im nächsten Quartal nicht die gewünschte Absatzmenge. Was könnte schief gegangen sein? (Kurze Begründung)

Lösung:

a) Lineare Regression von y auf x

	unabhängige Variable Preis x	abhängige Variable Menge y	x^2	$x \cdot y$	y^2
VI 2001	50	1000	2500	50 000	1 000 000
I 2002	55	900	3025	49 500	810 000
II 2002	45	1080	2025	48 600	1 166 400
II 2002	52	970	2704	50 440	940 900
Σ	202	3950	10 254	198 540	3 917 300

$$y = f(x) = a_1 + b_1 \cdot x$$

$$b_1 = \frac{4 \cdot 198\,540 - 202 \cdot 3\,950}{4 \cdot 10\,254 - (202)^2} = \frac{-3\,740}{212} = -17,6415$$

$$a_1 = \frac{3\,950 - (-17,6415) \cdot 202}{4} = 1\,878,3962$$

$$\text{d.h. } f(x) = 1878,40 - 17,64 \cdot x$$

$$\text{b) } b_2 = \frac{-3\,740}{4 \cdot 3\,917\,300 - (3\,950)^2} = \frac{-3\,740}{66\,700} = -0,0561$$

$$\text{Bestimmtheitsmaß } B = b_1 \cdot b_2 = 0,9892$$

d.h. 98,92% der Streuung der y -Werte wird erklärt durch die Streuung der Regressionsgeraden.

$$\text{oder Korrelationskoeffizient } r = -\sqrt{B} = -\sqrt{0,9892} = -0,9946$$

d.h. es liegt starker linearer (negativer) Zusammenhang vor.

c) Lineare Regression von x auf y

Jetzt ist die Menge y die unabhängige Variable und der Preis x ist die abhängige Variable

$$x = g(y) = a_2 + b_2 \cdot y$$

$$a_2 = \frac{202 - (-0,0561) \cdot 3\,950}{4} = 105,8711$$

$$g(1200) = 105,8711 - 0,0561 \cdot 1200 = 38,55$$

d.h. sie muss einen Preis von 38,55 Euro ansetzen.

d) Da der Wert 1200 außerhalb des beobachteten Bereichs [900;1200] liegt, handelt es sich hier um eine Extrapolation. Extrapolationen können schnell ungenau sein.

oder:

Lösung:

a) Lineare Regression von x auf p

	unabhängige Variable Preis p	abhängige Variable Menge x	p^2	$x \cdot p$	x^2
VI 2001	50	1000	2500	50 000	1 000 000
I 2002	55	900	3025	49 500	810 000
II 2002	45	1080	2025	48 600	1 166 400
II 2002	52	970	2704	50 440	940 900
Σ	202	3950	10 254	198 540	3 917 300

$$g(p) = a_2 + b_2 \cdot p$$

$$b_2 = \frac{4 \cdot 198\,540 - 202 \cdot 3\,950}{4 \cdot 10\,254 - (202)^2} = \frac{-3\,740}{212} = -17,6415$$

$$a_2 = \frac{3\,950 - (-17,6415) \cdot 202}{4} = 1\,878,3962$$

$$\text{d.h. } g(p) = 1878,40 - 17,64 \cdot p$$

$$\text{b) } b_1 = \frac{-3\,740}{4 \cdot 3\,917\,300 - (3\,950)^2} = \frac{-3\,740}{66\,700} = -0,0561$$

$$\text{Bestimmtheitsmaß } B = b_1 \cdot b_2 = 0,9892$$

d.h. 98,92% der Streuung der x -Werte wird erklärt durch die Streuung der Regressionsgeraden.

$$\text{oder Korrelationskoeffizient } r = -\sqrt{B} = -\sqrt{0,9892} = -0,9946$$

d.h. es liegt starker linearer (negativer) Zusammenhang vor.

c) Lineare Regression von p auf x

$$f(x) = a_1 + b_1 \cdot x$$

$$a_1 = \frac{202 - (-0,0561) \cdot 3\,950}{4} = 105,8711$$

$$f(1200) = 105,8711 - 0,0561 \cdot 1200 = 38,55$$

d.h. sie muss einen Preis von 38,55 Euro ansetzen.

d) Da der Wert 1 200 außerhalb des beobachteten Bereichs [900;1 080] liegt, handelt es sich hier um eine Extrapolation. Extrapolationen können schnell ungenau sein.

Aufgabe 3

Arbeitslosigkeit in Deutschland im Jahr 2002 nach Monaten

Zeitraum	Empirischer Wert	Saisonbereinigter Wert
	in 1 000	
Januar 2002	4 290	3 974
Oktober 2002	3 930	4 119

1) Jeweils Monats-Endstände

Quelle: Bundesanstalt für Arbeit, veröffentlicht in den Monatsberichten
der Deutschen Bundesbank

- Welche Komponenten (Bestandteile) von Zeitreihen sind in den empirischen Werten der Arbeitslosenzahlen enthalten? Nennen Sie die Komponenten.
- Welche Komponenten von Zeitreihen sind noch in den saisonbereinigten Werten der Arbeitslosenzahlen enthalten? Nennen Sie diese Komponenten.
- Worauf ist der Anstieg des saisonbereinigten Wertes der Arbeitslosenzahlen um rund 150 000 von Januar bis Oktober 2002 zurückzuführen?
- Worauf ist der Rückgang des empirischen Wertes der Arbeitslosenzahlen um 360 000 von Januar bis Oktober 2002 zurückzuführen?
- Warum geben die von der Bundesanstalt für Arbeit erfassten und der Öffentlichkeit bekannt gegebenen Zahlen der Arbeitslosen nicht das wahre Ausmaß der Arbeitslosigkeit in Deutschland wieder? Nennen Sie drei Fehlerquellen, von denen zwei dafür sorgen, dass die Arbeitslosigkeit zu niedrig angegeben wird, während die dritte Fehlerquelle für einen überhöhten Ausweis der Arbeitslosigkeit sorgt.

Lösung:

- a) Modell für die empirische Zeitreihe: $y_t = g_t + s_t + u_t$

g_t = Konjunktur (glatte Komponente)

s_t = Saison

u_t = unerklärter Rest

In der empirischen Zeitreihe sind im additiven Modell die Komponenten Konjunktur und Saison enthalten, außerdem die unerklärte Restkomponente.

- b) $\underbrace{y_t - s_t}_{\text{saisonbereinigte Reihe}} = g_t + u_t$

d.h. in der saisonbereinigten Zeitreihe sind sowohl die glatte Komponente als auch die Restkomponente enthalten.

- c) $4\,119\,000 - 3\,974\,000 = 145\,000 \approx 150\,000$

d.h. Anstieg um 150 000

Der Anstieg ist auf die glatte Komponente zurückzuführen, da u_t ein unerklärter Resteinfluss ist. Konkret bedeutet das, die schlechte Konjunktur-Lage war für den Anstieg verantwortlich.

d) $3\,930\,000 - 4\,290\,000 = -360\,000$

d.h. Rückgang um 360 000

$$\begin{array}{ccccccc}
 y_t & = & g_t & +s_t & & +u_t & \\
 \downarrow & & \downarrow & & & \downarrow & \\
 \text{Rückgang} & & \text{Anstieg} & & & \text{nicht erklärt} & \\
 \text{d.h. der Rückgang ist auf saisonale Einflüsse zurückzuführen.} & & & & & &
 \end{array}$$

- e) 1. Fehlerquelle: ABM-Stelleninhaber, Umschüler werden nicht als Arbeitslose erfasst, sondern den Erwerbstätigen zugerechnet, obwohl sie keine feste Stelle haben (ca. 1,7 Mio. Menschen)
2. Fehlerquelle: Personen, die gerne arbeiten würden, sich jedoch auf Grund der Tatsache, dass ihnen eh kein Arbeitslosengeld zusteht, nicht arbeitslos gemeldet haben, werden nicht erfasst (stille Reserve, ca. 0,5 bis 1,2 Mio. Menschen)
Die ersten beiden Fehlerquellen führen zu einer niedrigeren Anzeige der Arbeitslosigkeit.
3. Fehlerquelle: Personen, die nicht mehr arbeiten wollen, sich jedoch arbeitslos gemeldet haben, da die Chance gering ist, dass ihnen eine Stelle zugewiesen wird, erhöhen die Anzeige der Arbeitslosigkeit. (ca. 1 Mio Menschen)

Aufgabe 4

Die Produktion eines Einproduktunternehmens wird im Wirtschaftsjahr 2003 auf drei Standorte A, B, C verteilt. Jedes Produkt wird genau nur an einem der drei Standorte produziert. In Standort A werden 45% und in Standort B 35% aller Produkte gefertigt. Der Rest wird in Standort C produziert. Die Ausschussraten an den einzelnen Standorten werden für das Wirtschaftsjahr 2003 wie folgt prognostiziert:

Standort	Ausschussrate
A	5%
B	7%
C	4%

- a) Welcher Anteil der Produkte wird im Wirtschaftsjahr 2003 am Standort C gefertigt?
- b) Mit welchem Ausschussanteil an der Gesamtproduktion muss das Unternehmen im Wirtschaftsjahr 2003 auf Grund der vorliegenden Daten rechnen?
- c) Im März 2003 bekommt ein Kunde ein defektes Produkt geliefert. Dieses wurde am 28. Februar produziert. Mit welcher Wahrscheinlichkeit wurde das defekte Produkt am Standort C gefertigt?

Lösung:

B_1 = Produkt wurde am Standort A produziert

B_2 = Produkt wurde am Standort B produziert

B_3 = Produkt wurde am Standort C produziert

A = Produkt ist Ausschuss

Bekannt sind die folgenden Wahrscheinlichkeiten:

$$0,45 = P(B_1) \quad P(A | B_1) = 0,05$$

$$0,35 = P(B_2) \quad P(A | B_2) = 0,07$$

$$0,20 = P(B_3) \quad P(A | B_3) = 0,04$$

a) $P(B_3) = 1 - 0,45 - 0,35 = 0,20$

d.h. 20% aller Produkte werden am Standort C produziert.

b) $P(A) = P(A \cap B_1) + P(A \cap B_2) + P(A \cap B_3)$

$$P(A | B_1) = \frac{P(A \cap B_1)}{P(B_1)} \Rightarrow P(A \cap B_1) = P(A | B_1) \cdot P(B_1) = 0,05 \cdot 0,45 = 0,0225$$

$$P(A \cap B_2) = P(A | B_2) \cdot P(B_2) = 0,07 \cdot 0,35 = 0,0245$$

$$P(A \cap B_3) = P(A | B_3) \cdot P(B_3) = 0,04 \cdot 0,20 = 0,008$$

$$P(A) = 0,0225 + 0,0245 + 0,008 = 0,055$$

Arbeitstabelle:

	B_1	B_2	B_3	Σ
A	0,0225	0,0245	0,0080	0,0550
\bar{A}				
	0,45	0,35	0,20	1

d.h. der Ausschussanteil an der gesamten Produktion beträgt 5,5%.

c) $P(B_3 | A) = \frac{P(A \cap B_3)}{P(A)} = \frac{0,008}{0,055} = 0,1455$

d.h. die Wahrscheinlichkeit beträgt 0,1455.

Aufgabe 5

Eine Studentin der Fakultät Wirtschaftswissenschaften plant, am Ende des laufenden Semesters an vier Prüfungen A,B,C,D teilzunehmen. Auf Grund einer Selbsteinschätzung kommt sie zu folgendem Ergebnis:

Prüfung	Wahrscheinlichkeit, die Prüfung	
	zu bestehen	nicht zu bestehen
A	0,85	0,15
B	0,85	0,15
C	0,85	0,15
D	0,72	0,28

Darüber hinaus geht sie davon aus, dass die Ausgänge der einzelnen Prüfungen stochastisch unabhängig sind.

- Berechnen Sie die Wahrscheinlichkeit, dass die Studentin genau zwei der ersten drei Prüfungen (A,B,C) besteht.
- Berechnen Sie die Wahrscheinlichkeit, dass die Studentin höchstens zwei der ersten drei Prüfungen (A,B,C) besteht.
- Berechnen Sie die Wahrscheinlichkeit, dass die Studentin alle vier Prüfungen (A,B,C,D) besteht.

- d) Berechnen Sie die Wahrscheinlichkeit, dass die Studentin genau drei der vier Prüfungen (A,B,C,D) besteht.
- e) Berechnen Sie die Wahrscheinlichkeit, dass die Studentin mindestens drei der vier Prüfungen (A,B,C,D) besteht.

Lösung:

- A = Studentin besteht Prüfung A
 B = Studentin besteht Prüfung B
 C = Studentin besteht Prüfung C
 D = Studentin besteht Prüfung D

Bekannt sind die folgenden Wahrscheinlichkeiten:

$$\begin{aligned} P(A) &= 0,85 \\ P(B) &= 0,85 \\ P(C) &= 0,85 \\ P(D) &= 0,72 \end{aligned}$$

- a) $ABC\bar{C}$ oder $A\bar{B}C$ oder $\bar{A}BC$
 $0,85 \cdot 0,85 \cdot 0,15 + 0,85 \cdot 0,15 \cdot 0,85 + 0,15 \cdot 0,85 \cdot 0,85 = 0,3251$
d.h. die Wahrscheinlichkeit beträgt 0,3251.

- b) höchstens zwei = genau zwei oder genau eine oder keine

Ereignis	Wahrscheinlichkeit
genau zwei	0,3251
genau eine	$A\bar{B}\bar{C}$ oder $\bar{A}B\bar{C}$ oder $\bar{A}\bar{B}C$ $0,85 \cdot 0,15 \cdot 0,15 + 0,15 \cdot 0,85 \cdot 0,15 + 0,15 \cdot 0,15 \cdot 0,85 = 0,0574$
keine	$\bar{A}\bar{B}\bar{C}$ $0,15 \cdot 0,15 \cdot 0,15 = 0,0034$
\sum	$0,3251 + 0,0574 + 0,0034 = 0,3859$

d.h. die Wahrscheinlichkeit beträgt 0,3859.

- c) $ABCD$

$$0,85 \cdot 0,85 \cdot 0,85 \cdot 0,72 = 0,4422$$

d.h. die Wahrscheinlichkeit beträgt 0,4422.

- d) $ABC\bar{D}$ oder $AB\bar{C}D$ oder $A\bar{B}CD$ oder $\bar{A}BCD$

$$0,85 \cdot 0,85 \cdot 0,85 \cdot 0,28 + 0,85 \cdot 0,85 \cdot 0,15 \cdot 0,72 + 0,85 \cdot 0,15 \cdot 0,85 \cdot 0,72 + 0,15 \cdot 0,85 \cdot 0,85 \cdot 0,72 = 0,4060$$

d.h. die Wahrscheinlichkeit beträgt 0,4060.

- e) mindestens drei = genau drei oder genau vier

Ereignis	Wahrscheinlichkeit
genau drei	0,4060
genau vier	0,4422
\sum	0,8482

d.h. die Wahrscheinlichkeit beträgt 0,8482.

Aufgabe 6

Um die Zufrieden- bzw. Unzufriedenheit der Studierenden mit ihrem Studium an der FH Köln abzuklären, entschließt sich der Rektor der FH Köln, eine Befragung durchzuführen.

Einer seiner Mitarbeiter schlägt ihm Folgendes vor:

„Wir führen eine Befragung der Studierenden bei einer repräsentativen Stichprobe an der FH Köln immatrikulierten Studierenden durch. Mit einer gewissen Wahrscheinlichkeit und innerhalb gewisser Grenzen können wir dann darauf vertrauen, dass die Ergebnisse der Stichprobe vom Umfang 500 aussagekräftig für alle 18 000 Studierenden der FH Köln sind.“

Der Rektor fragt nach:

„Wie kann man von den Ergebnissen der Stichprobe auf die wahren, unbekannt, aber eigentlich interessierenden Merkmale einer Grundgesamtheit schließen?“

Was würden Sie ihm antworten?

Lösung:

p = Anteil der zufriedenen Studierenden an der FH Köln

Der Wert von p ist unbekannt.

\hat{p} = Anteil der zufriedenen Studierenden in der Stichprobe

Der Wert von \hat{p} lässt sich anhand der Stichprobe berechnen.

$n = 500$ Stichprobenumfang

Statistisches Verfahren = Anhand der Stichprobe ein 0,95-Konfidenzintervall für p berechnen:

$$\left[\hat{p} - 1,96 \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{500}} ; \hat{p} + 1,96 \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{500}} \right]$$

Die Breite des Konfidenzintervalls beträgt höchstens:

$$\text{Obergrenze} - \text{Untergrenze} = 2 \cdot 1,96 \cdot \sqrt{\frac{\hat{p}(1-\hat{p})}{500}} \leq \underbrace{2 \cdot 1,96 \cdot \sqrt{\frac{0,5 \cdot 0,5}{500}}}_{\text{worstcase}} = 0,0877$$

d.h. die maximale Abweichung vom wahren Parameter nach oben oder unten beträgt etwa 4,4%-Punkte . Das ist gut!