19. Zeitreihengestützte Prognose

Prognosen sind immer dann schwierig, wenn sie sich auf die Zukunft beziehen

Mark Twain

Es gibt zwei Prognosetypen:

- Bei der **unbedingten Prognose** sind alle Einflussfaktoren implizit in der Zeitreihe enthalten, so dass die Prognose durch direkte Analyse der Zeitreihe gewonnen wird.
- Bei der **bedingten Prognose** kennt man die zu prognostizierende Größe als Funktion von Einflussfaktoren oder Modellparametern. Falls diese sich leichter prognostizieren lassen, setzt man auch die Prognose als entsprechende Funktion der Einflussfaktoren bzw. als Modell an.[1]

Bemerkungen dazu:

- Die Aufspaltung von Zeitreihen in Trend, Saisonkomponente, etc. kann zum Zweck der bedingten Prognose angewendet werden: Jeder Einflussfaktor (Trend, glatter und saisonaler Anteil) kann für sich besser prognostiziert werden als die nicht aufgespaltene Zeitreihe als Ganzes.
- Typische Vertreter der unbedingten Prognosemethoden sind Voraussagen auf der Basis von gleitenden Mitteln, welche z.B. bei der "Chartanalyse" von Aktienkursen eingesetzt werden.

Verständnisfrage Erläutern Sie die bedingte Prognose anhand der Aufspaltung der Zeitreihe gemäß (i) dem additiven, (ii) dem multiplikativen Modell.

[1] Bei Aktien spricht man oft von "Technischer Analyse" bzw. "Chartanalyse" und "Fundamentalanalyse". Erstere ist Basis der unbedingten Prognose, letztere eine Methode der bedingte Prognose.

19.1 Unbedingte Prognose

Gegeben sei eine Zeitreihe $\{y_1, \cdots, y_t\}$, wobei t die Gegenwart bzw. den letzten Datenpunkt darstellt. Die **unbedingte Prognose** beinhaltet Schätzungen $\hat{y}_{t+1}$, $\hat{y}_{t+2}$, etc. ausschließlich anhand dieser Zeitreihe, ohne Berücksichtigung weiterer Informationen.

Verschiedene Prognoseannahmen sind gebräuchlich:

(1) **Elementaren Fortschreibung**
- des **Wertes**: $\hat{y}_{t+1} = y_t$. ("Das Wetter bleibt so, wie es ist"),
- der **Anstiegsrate**: $\hat{y}_{t+1} = 2y_t - y_{t-1}$. ("Im Oktober kühlt es gegenüber September um ebenso viele Grade ab wie September gegenüber August"),
- oder des **Wachstumsfaktors**: $\hat{y}_{t+1} = y_t^2 / y_{t-1}$. ("Die Steigerung der Autoproduktion ist dieselbe wie letztes Jahr").

(2) **Fortschreibung von gleitenden Mitteln**
- Wie bei der elementaren Fortschreibung ist eine Fortschreibung des Wertes, der Anstiegsrate oder des Wachstumsfaktors möglich, wobei aber nicht die Zeitreihe selbst, sondern ein davon abgeleitetes gleitendes Mittel als Basis dient.
- Vorteil: Fluktuationen/Ausreißer werden schwächer berücksichtigt.
- Nachteil: Das Verfahren ist träger im Vergleich zur elementaren Fortschreibung.

(3) **Lokalisierte (lokal gewichtete) lineare Regression** Diese funktioniert wie die Fortschreibung der Anstiegsrate des gleitenden Mittels, ist aber robuster.

19.1(b) Prognose durch Fortschreibung gleitender Mittel

Sie können immer nur am rechten Ende des Aktiencharts handeln, nie in der Mitte!

Ausspruch meines Brokers

Da beim gleitenden Durchschnitt der Ordnung τ auf beiden Seiten des betrachteten Punktes mindestens $\tau/2$ Nachbarn liegen müssen, sind diese am Rand der Zeitreihe nicht definiert. Für eine Prognose um einen Zeitschritt muss deshalb das gleitende Mittel nicht nur um einen, sondern um $\tau/2+1$ Zeitschritte fortgeschrieben werden!

Beispiel: Bei Prognose durch die Anstiegsrate des gleitenden arithmetische Mittels der ungeraden Ordnung $\tau = 2m+1$ werden die "aktuellsten" beiden Werte des Mittels um $m+1$ Schritte fortgeschrieben:

$$\hat{y}_{t+1} = \bar{y}^{(\tau)}_{t-m} + (m+1)\left(\bar{y}^{(\tau)}_{t-m} - \bar{y}^{(\tau)}_{t-m-1}\right)$$

Aufgabe: Gegeben sei die Zeitreihe

3 6 9 10 12 12 18 24.

(a) Prognostizieren Sie den nächsten Wert durch elementare Fortschreibung (i) des letzten Wertes, (ii) der letzten Anstiegsrate, (iii) des letzten Wachstumsfaktors.

(b) Zeigen Sie, dass bei Verwendung des gleitenden Mittels der Ordnung 3 und Prognose durch die Anstiegsrate dieses Mittels sich folgende explizite Prognoseformel ergibt:

$$\hat{y}_{t+1} = \frac{1}{3}\big(-2y_{t-3} + y_{t-2} + y_{t-1} + 3y_t\big).$$

Prognostizieren Sie damit den nächsten Wert obiger Zeitreihe.

19.1(c) Prognose mittels lokalisierter linearer Regression

Die explizite Prognoseformel für die gleitenden Durchschnitte der Ordnung 3 erscheint etwas willkürlich: Warum wurden der weit in der Vergangenheit liegende Wert y_{t-3} so stark gewichtet? Bessere statistische Eigenschaften hat die **lokalisierte lineare Regression**:

- Zunächst werden die letzten τ Punkte der Zeitreihe ausgewählt und *nur bezüglich dieser Datenpunkte* eine lineare Regression durchgeführt,
- die Prognose ist nun einfach durch die Regressionsfunktion $\hat{y}(t) = a + bt$ gegeben.

Es werden also nur die letzten τ Datenpunkte zur Voraussage verwendet, und zwar gleichgewichtet. Will man (analog zum exponentiellen gleitenden Mittel) die aktuellsten Daten stärker gewichten, verallgemeinert man dies zur **kernbasierten linearen Regression**:

$$a = \bar{y}^w - b\bar{x}^w, \quad b = \frac{\sum_i w_i x_i y_i - \bar{x}^w \bar{y}^w}{\sum_i w_i x_i^2 - (\bar{x}^w)^2}, \quad \bar{x}^w = \sum_i w_i x_i$$

wobei die τ Koeffizienten des Wichtungskerns lediglich die Bedingung

$$\sum_{i=1}^{\tau} w_i = 1$$

erfüllen müssen. Man kann also alle bisherigen Regressionsformeln verwenden, wenn man $\frac{1}{n}\sum_i$ durch $\sum_i w_i$ ersetzt! Umgekehrt ergibt sich die ungewichtete lokalisierte Regression aus der kernbasierten Regression durch $w_i = \frac{1}{\tau}$.

Beispiele zu 19.1(c)

Aufgabe:

(a) Führen Sie anhand der Zeitreihe von 19.1(b) eine Prognose durch lokalisierte lineare Regression der Ordnung $\tau = 4$ mit Gleichgewichtung durch.

(b) Verwenden Sie nun zur Prognose den Wichtungskern $\{w_t, w_{t-1}, w_{t-2}, w_{t-3}\} = \{0.4, 0.3, 0.2, 0.1\}$.

(c) Geben Sie für die gleichgewichtete lokalisierte lineare Regression für $\tau = 3$ eine explizite Prognoseformel folgender Form an:

$$\hat{y}_{t+1} = c_0 y_t + c_1 y_{t-1} + c_2 y_{t-2},$$

d.h., ermitteln Sie die Koeffizienten!

19.2 Bedingte Prognose durch Komponentenzerlegung

Das Vorgehen ist wie folgt:

1. Auswahl des Modells: Additive oder multiplikative Zerlegung
2. Analyse der Zeitreihe, d.h. Zerlegung in die Komponenten gemäß Abschnitt 16-18,
3. Prognose der Komponenten T und G bzw. K mit einer der Methoden von 19.1. (Die Saisonkomponente ist per Definition periodisch, der Verlauf bleibt also in der Prognose unverändert.)
4. Synthese, d.h. Addition/Multiplikation der prognostizierten Komponenten ohne den U-Anteil.

Aufgabe Folgende Zeitreihe gibt den Umsatz eines Unternehmens in den letzten Halbjahren an:

Jahr	2003	2004	2005	2006
1. Halbjahr	4	9	20	30
2. Halbjahr	3	6	15	18

Führen Sie die komponentenbasierte Prognose für die beiden Halbjahre 2007 durch.

(a) Isolieren Sie zunächst die glatte und die Saisonkomponente nach dem multiplikativen Modell.

(b) Prognostizieren Sie die glatte Komponente durch elementare Fortschreibung der Anstiegsrate (bzw. des Wachstumsfaktors in den ursprünglichen Daten).

(c) Synthetisieren Sie die Einzelanteile durch Multiplikation.

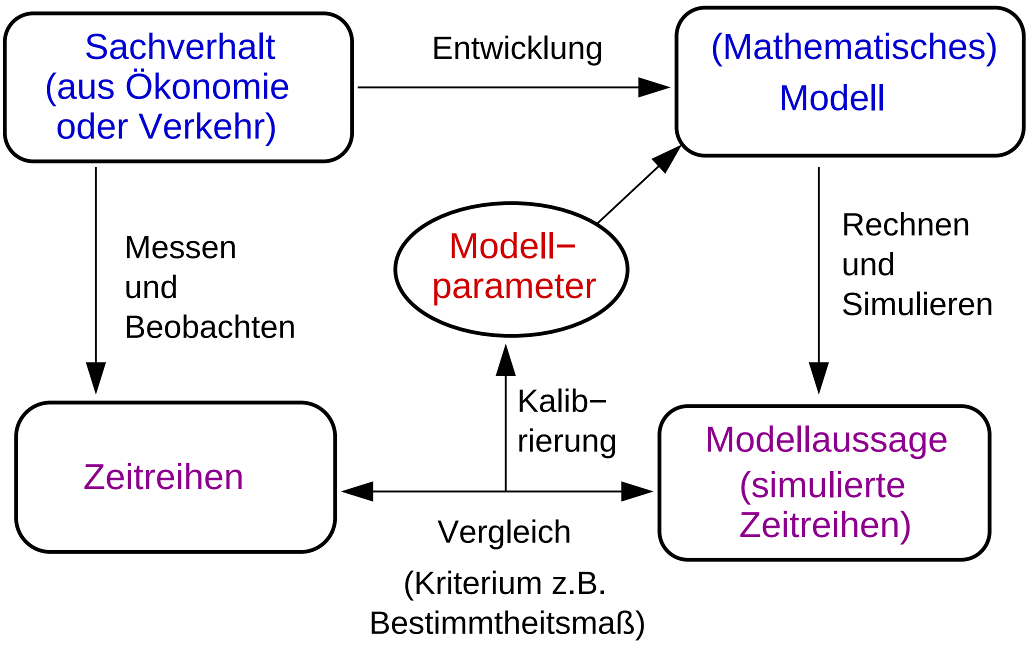
19.3 Modellgestützte Prognose

Das einzig Beständige ist die Veränderung als solche.

Heraklit, etwa 500 v. Chr.

Eine modellgestützte Prognose beruht i.A. darauf, dass die der Zeitreihe zugrundeliegende Dynamik mit **mathematischen Modellen** *modellierbar* ist. Die zu den Modellen gehörigen **Parameter** sind dabei entweder konstant oder ihre zukünftige Änderung lässt sich leichter prognostizieren. Vorgehensweise:

1. Formulierung des Modells
2. **Analyse:** Die Modellparameter weren durch Vergleich der Modellaussagen $\hat{y}_t$ mit den Daten y_t **kalibriert**
3. **Prognose:** Die kalibrierten Modellparameter werden mit einfachen Ansätzen in die Zukunft fortgeschrieben (in einfachsten Fall unverändert gelassen) und die zukünftige Zeitentwickung des Modells als Prognose identifiziert.



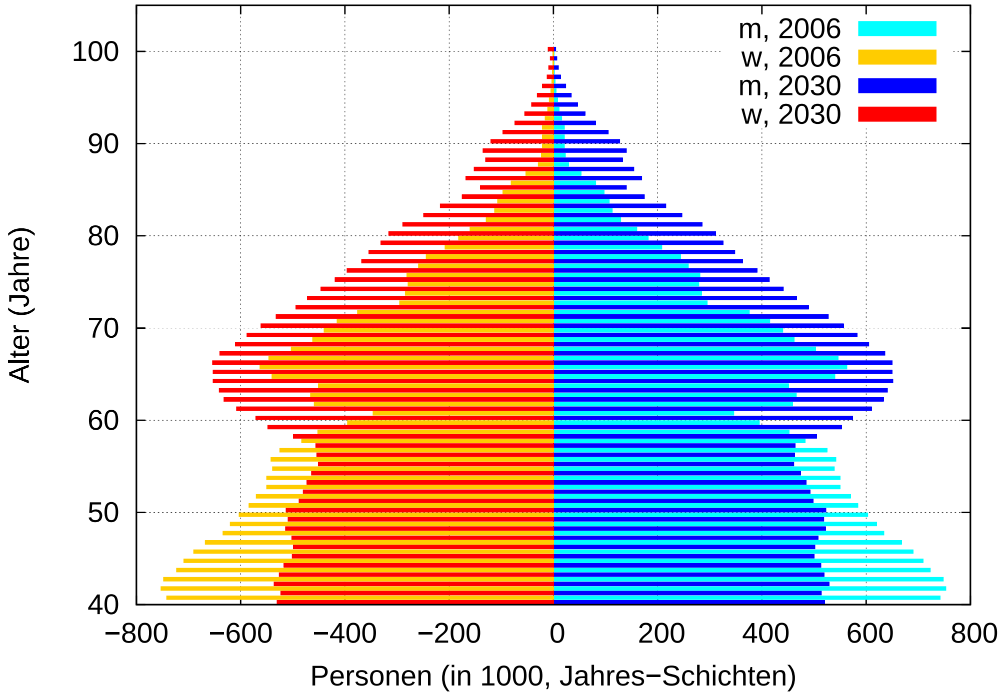
19.3(b) Bemerkungen zur modellgestützte Prognose

With four parameters, I can fit an elephant; with five parameters, I can make him wiggle his trunk

John von Neumann

- Die im vorhergehenden Abschnitt betrachtete Prognose durch Komponentenzerklegung ist ein Beispiel einer modellbasierten Prognose. Das Modell ist dabei die additive bzw. multiplikative Zerlegung selbst.
- Regressionsfunktionen (vor allem nichtlineare) können ebenfalls als Modell aufgefasst werden, z.B. ist
  - die e-Funktion ein Modell des ungebremsten Wachstums,
  - die Logistische Funktion ein Modell des Wachstums mit Sättigung.
- Bei der Kalibrierung wird die vom Modell errechnete Zeitreihe mit der realen verglichen und die Modellparameter so angepasst, dass die Übereinstimmung bezüglich einer **Zielfunktion** wie dem Bestimmtheitsmaß oder der Fehlerquadratsumme maximiert bzw. minimiert wird. Es wird also im Wesentlichen die nichtlineare Regressionsrechnung auf allgemeine (nicht notwendigerweise durch eine Formel darstellbare) Modellaussagen verallgemeinert.

19.3(c) Beispiel 1: Wandel der Altersstruktur



Vereinfachtes Modell (Detaillierteres unter `www.destatis.de`):

- Die Altersstruktur wird in jedem Jahr t durch die Zahlen $n_T(t)$ der in diesem Jahr T Jahre alten Personen beschrieben. Die Prognose wird separat für Frauen und Männer durchgeführt.
- Die Wahrscheinlichkeit u_T, das nächste Jahr zu überleben, hängt nur vom Alter und dem Geschlecht ab. Diese Bilanz der Verschiebung der Altersschichten wird durch Zuzüge vom Ausland und Wegzüge zum Ausland ergänzt:

$$n_T(t+1) = u_{T-1} n_{T-1}(t) + n_T^{\text{zu}} - n_T^{\text{ab}}, \quad T \geq 1$$

- Jede Frau im Alter T gebiert mit der Wahrscheinlichkeit p_T^{geb} im nächsten Jahr ein Kind. Dieses ist mit der Wahrscheinlichkeit $p_{♀}$ weiblich:

$$n_0(t+1) = \sum_{T=T_{\min}}^{T_{\max}} p_T^{\text{geb}} n_T^{♀}(t) + n_0^{\text{zu}} - n_0^{\text{ab}}.$$

19.3(d) Beispiel 1: Wandel der Altersstruktur (Fortsetzung)

Das Modell hat folgende Parameter:

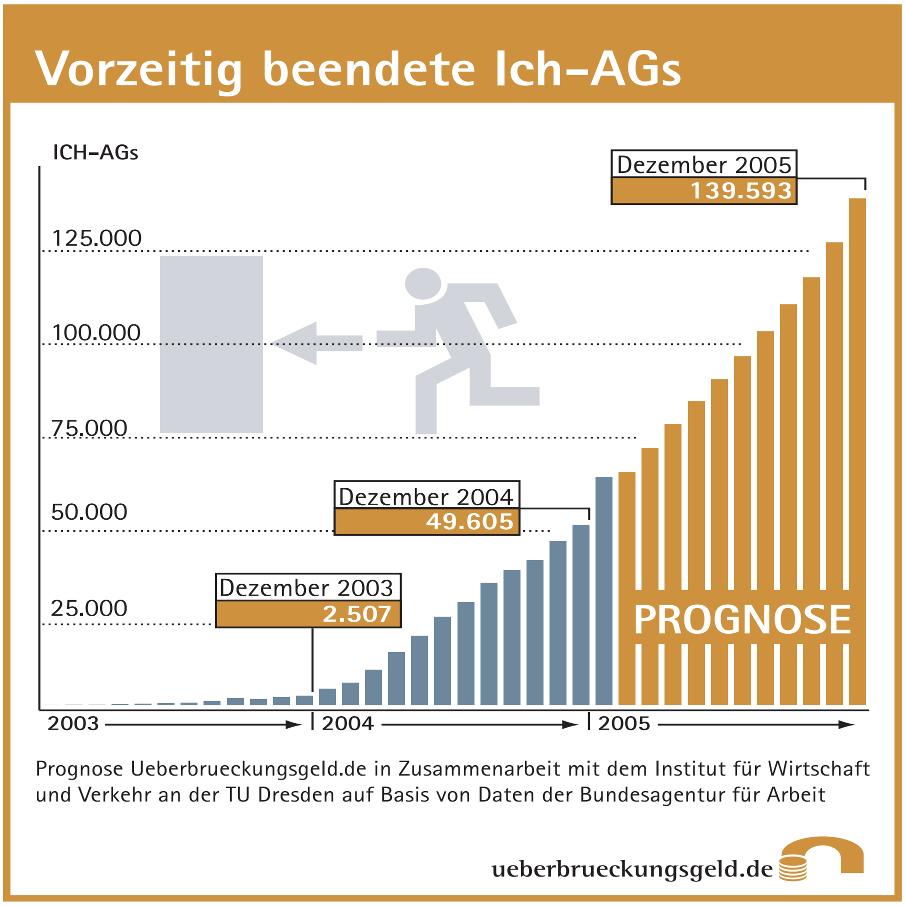
- Die bedingten Überlebenswahrscheinlichkeiten u_T, getrennt nach Geschlecht. Diese sind bei Versicherungen wohl bekannt und als sogenannte *Sterbetafeln* tabelliert .
- Die Altersstruktur p_T^{geb} der werdenden Mütter (zur Zeit gilt etwa $\sum_T p_T^{\mathrm{geb}} = 1.4$), sowie $p_{♀}$ (ungefähr = 0.5),
- Die Zu- und Fortzugsraten n_T^{zu} und n_T^{ab}.

Bemerkungen

- Bereits die einfachste Prognose, bei der all diese Parameter nach Kalibrierung für die Zukunft konstant gelasen werden, ergibt selbst für Jahrzehnte eine bemerkenswert gute Prognose, vor allem für die bereits geborenen Jahrgänge,
- Verschiebt man die Sterbetafeln der älteren Jahrgänge um 1 Lebensjahr pro Jahrzehnt, wird die Tatsache berücksichtigt, dass die Bevölkerung immer älter wird.
- Eine analoge lineare Verschiebung der Altersstruktur p_T^{geb} berücksichtigt die Tatsache, dass die Eltern immer älter werden.

19.3(c) Beispiel 2: Prognose der Zahl der Ich-AGs

Ich-AGs sind/waren eine spezielle Form von Einmannunternehmen, für die besondere staatliche Unterstützungen ("Existenzgründerzuschüsse") gewährt wurden: Am intensivsten die ersten 12 Monate, danach weniger. Anfang 2005 bestand die Frage, ob man den sich abzeichnenden Anstieg der Ausstiegszahlen (und damit den Erfolg oder Misserfolg dieser Maßnahme) prognostizieren könnte.



Dieses Beispiel einer konkreten modellgestützten Prognose (vgl. `http://www.gruendungszuschuss.de/service-menue/presse/pressemitteilungen.html` vom 04.02.2005) wurde analog zur Demografie-Dynamik modelliert und lieferte bei einer 1-Jahr Prognose einen Prognosefehler von weniger als 10%.

19.3(d) Beispiel 2: Prognose der Zahl der Ich-AGs (Forts.)

Das Prognosemodell (ausführliche Informatinen unter diesem Link) basierte auf folgenden Annahmen:

- Die "Ich-AG's" wurden nach ihrem "Alter" (in Monaten seit der Gründung) aufgegliedert.
- Die Ausfallwahrscheinlichkeit innerhalb eines Monats wurde für alle Monate außer dem 12-ten und 13-ten mit α_0, für die beiden kritischen Monate (im 13-ten Monat wird dir Förderung reduziert!) mit α_c angesetzt.
- "Zu- und Fortzüge" entfallen.
- Die "Geburtenrate" (Neugründungen) wurde nach den Mitteln der unbedingten Prognose fortgeschrieben. (Ab Juli 2006 lief dieses Modell aus, die dadurch verschwindende "Geburtenrate" erleichterte die Prognose zusätzlich.)

Das Modell enthält die beiden Parameter α_0 und α_c, welche durch Minimierung der Abweichungsquadrate der Aussteigerzahlen im Analysezeitraum auf die Werte $\alpha_0 = 0.4\%$ und $\alpha_c = 24.4\%$ kalibriert wurden.

Neben einer genaueren Voraussage der Ausstiegszahlen machte das Modell insbesondere die Aussage, dass eine Ich-AG nur zu 50% drei Jahre oder älter werden wird und damit als erfolgreich angesehen wird. Diese Aussage (zu einem Zeitpunkt, als es noch keine dreijährigen Ich-AGs gab) ist mit einer unbedingten Prognosemethode überhaupt nicht möglich.

19.3(e) Beispiel 3: CO_2-Verbrauchsprognose (Aufgabe)

Der Zeitverlauf des in Deutschland durch private Kfz insgesamt verursachten jährlichen CO_2 - Ausstoßes soll mit folgendem Modell berechnet werden:

$$C_{\mathrm{tot}}(t) = \sum_T n_T(t) p_T^{\mathrm{FS}}(t) l_T(t) C_T^{\mathrm{Kfz}}(t)$$

Hierbei bedeuten

- $n_T(t)$ die Zahl der T Jahre alten Einwohner,
- $p_T^{\mathrm{FS}}(t)$ der Führerscheinanteil der jeweiligen Altersgruppe,
- $l_T(t)$ die jährliche Fahrleistung T Jahre alter Führerscheinbesitzer als Fahrer,
- $C_T^{\mathrm{Kfz}}(t)$ der CO_2 - Ausstoß pro km, gemittelt über die in der jeweiligen KAltersklasse gefahrenen Fahrzeuge.

Der Einfachheit halber werden nur 6 Altersklassen mit je 18 Jahren Klassenbreite gemäß folgender Tabelle unterschieden:

Altersklasse k	n_k (Mio.)	p_k^{FS}	l_k(1000 km/J)	C_k^{Kfz}(g/km)
1 (0-18 J)	15	0	-	-
2 (18-36 J)	15	0.9	10	130
3 (36-54 J)	22	0.9	10	140
4 (54-72 J)	16	0.8	8	150
5 (72-90 J)	10	0.6	4	150
6 (>90 J)	2	0.4	1	150

19.3(f) Beispiel 3: CO_2-Verbrauchsprognose (Forts.)

Zur Prognose für 18 Jahre (= 1 Altersklasse) in die Zukunft werden folgende Annahmen getroffen:

- Keine Zu- und Fortzüge von/ins Ausland
- Die Überlebenswahrscheinlichkeit für die nächsten 18 Jahre beträgt 100% für die Altersklassen 1-3, 75% für die Klasse 4, 40% für die Klasse 5 und 0% für die Klasse 6 (diese wären ja zum Prognosezeitpunkt mindestens 108 Jahre alt).
- 90% aller 18-jährigen machen den Führerschein. Danach wird weder ein Führerschein erworben noch einer abgegeben.
- Die Jahresfahrleistung als Fahrer erhöht sich über alle Altersklassen um 10%, die CO_2 Emissionen der jeweiligen FLotten reduzieren sich um 20%.

Um wieviel Prozent erhöht bzw. erniedrigt sich der CO_2 Ausstoß gemäß diesem Szenarium?