



**Doz. Dr. Stefan Weber & Team**  
Plagiats-, Titel- und Gutachtenprüfung

weber@plagiatsgutachten.de | <https://plagiatsgutachten.com>  
Science City | Schillerstraße 30 | 5020 Salzburg | Österreich | +43 664 13 13 444

**Gutachten zur Einhaltung der Regeln**  
**guter wissenschaftlicher Praxis in der Dissertation**  
***„Spatial Variability of Air Pollutants in the Basel Area and***  
***Carcinogenic and Non-Carcinogenic Health Risk“***  
**von Prof. Dr. Martin Rösli, Universität Basel, 2001**



## Inhaltsverzeichnis des Gutachtens

### **I Zusammenfassung**

Verfasser: Doz. Dr. Stefan Weber, <https://plagiatsgutachten.com/die-plagiatsprüfer>

### **II Studienrechtliche Einordnung**

Verfasser: Doz. Dr. Stefan Weber, <https://plagiatsgutachten.com/die-plagiatsprüfer>

### **III Ergebnis der Überprüfung auf Plagiatsfragmente**

Verfasserin: Birgit Kaiserlehner, MA, Team <https://plagiatsgutachten.com/die-plagiatsprüfer>

### **IV Ergebnis der Überprüfung der empirischen Teile der Dissertation**

Verfasserin: MMag. Dr. ETH Sabrina Dorn, MSc ETH, <https://www.knowledgeconsult.at>



## I Zusammenfassung

Verfasser: Doz. Dr. Stefan Weber, <https://plagiatsgutachten.com/die-plagiatsprüfer>

Dieses Gutachten dokumentiert eine studienrechtliche Unklarheit, **24 Plagiatsfragmente** sowie in einer gesonderten detaillierten Analyse – zum Teil auch auf Basis der Original-Rohdaten – „**leichte bis gröbere Fahrlässigkeiten**“ in den **empirischen Teilen** der gegenständlichen Dissertation, die nach Ansicht der analysierenden Statistikexpertin „gehäuft auftreten“.

Nach Rechtsauffassung des Gutachters ist für die Fallbeurteilung die gegenwärtige Promotionsordnung der Universität Basel heranzuziehen. In dieser heißt es in § 26:

### *Unlauteres Verhalten*

§ 26 Entsteht vor oder nach vollzogener Promotion der begründete Verdacht, dass die Kandidatin bzw. der Kandidat das Promotionsverfahren unlauter beeinflusst oder wissentlich irreführende Angaben gemacht hat, oder dass die Dissertation ganz oder teilweise aus einem Plagiat besteht, kann der Promotionsausschuss der Fakultätsversammlung beantragen, die Promotion zu widerrufen.

Quelle: Promotionsordnung der Philosophisch-Naturwissenschaftlichen Fakultät der Universität Basel vom 15. September 2015, S. 8



Das weitere Verfahren ist wie folgt geregelt:

- <sup>2</sup> Der Kandidatin bzw. dem Kandidaten ist Gelegenheit zu geben, zu den gegen sie bzw. ihn erhobenen Vorwürfen Stellung zu nehmen.**
- <sup>3</sup> Beschliesst die Fakultätsversammlung, die Promotion zu widerrufen, so verliert die Kandidatin bzw. der Kandidat die Berechtigung zur Führung des Titels «Dr. phil. des.» bzw. «Dr. phil.».**

*Quelle: Promotionsordnung der Philosophisch-Naturwissenschaftlichen Fakultät der Universität Basel vom 15. September 2015, S. 9*

Der Gutachter ersucht die Universität Basel um Einleitung eines solchen Verfahrens.



## II Studienrechtliche Einordnung

Verfasser: Doz. Dr. Stefan Weber, <https://plagiatsgutachten.com/die-plagiatsprüfer>

Die zum Zeitpunkt des Einreichens der Dissertation gültige Promotionsordnung normiert in § 5:

### *Manuskript der Dissertation*

§ 5. Die Dissertation muss eine Originalarbeit der Verfasserin oder des Verfassers sein. Die Hauptergebnisse sind kurz zusammenzufassen. Die graphischen Beilagen sollen in druckfertiger Form eingereicht werden. Geringfügige Änderungen des Manuskripts, die von der oder dem Fakultätsverantwortlichen oder von der Fakultät gutgeheissen werden, sind vor der Drucklegung anzubringen (vgl. Dissertationsdruckbestimmungen der Fakultät).

*Quelle: Reglement zur Erlangung des Doktorgrads an der Philosophisch-Naturwissenschaftlichen Fakultät der Universität Basel vom 22. Juni 1993, S. 3*

Die Dissertation von Prof. Dr. Martin Rösli ist jedoch im Kern eine kumulative Dissertation. Dieser Kern (S. 44 bis S. 130) besteht aus



fünf Papers mit jeweils mehreren Ko-Autorinnen und Autoren:

<b>PART IV</b>	<b>AIR QUALITY ANALYSIS</b>	<b>44</b>
CHAPTER 6	ARTICLE 1: Spatial Variability of Different Fractions of Particulate Matter Within an Urban Environment and Between Urban and Rural Sites <i>Rööslı M, Theis G, Künzli N, Oglesby L, Mathys P, Camenzind M, Staehelin J, Braun-Fahrländer C</i>	45
CHAPTER 7	ARTICLE 2: Airborne Particle Number Profiles, Particle Mass Distributions and Particle Bound PAH Concentrations Within the City Environment of Basel: an Assessment as Part of the Briska Project <i>Junker M, Kasper M, Rööslı M, Camenzind M, Künzli N, Monn C, Theis G, Braun-Fahrländer C</i>	65
CHAPTER 8	ARTICLE 3: Temporal and Spatial Variation of the Chemical Composition of PM <sub>10</sub> at Urban and Rural Sites in the Basel Area, Switzerland <i>Rööslı M, Theis G, Künzli N, Staehelin J, Mathys P, Oglesby L, Camenzind M, Braun-Fahrländer C</i>	84
<b>PART V</b>	<b>RISK ANALYSIS</b>	<b>105</b>
CHAPTER 9	ARTICLE 4: Which Effect Measure Should Be Used for Impact Assessment in a New Population Context? <i>Rööslı M, Künzli N, Schindler C, Braun-Fahrländer C</i>	106
CHAPTER 10	ARTICLE 5: Assessment of Cancer Risk Attributable to Air Pollution Using an Epidemiology Based Method <i>Rööslı M, Künzli N, Schindler C, Theis G, Mathys P, Oglesby L, Camenzind M, Braun-Fahrländer C</i>	116

*Quelle: Dissertation, S. iii*



Ausführungsbestimmungen, Sonderreglements etc. zu kumulativen Dissertationen wurden im Staatsarchiv Basel keine gefunden, siehe:

- ▼  **Philosophisch-Naturwissenschaftliche Fakultät**  
UNI-REG 17 | 1961-2013
- ▼  **Dekanat**  
UNI-REG 17a | 1962-1995
  -  **[leer] / Allgemeines und Einzelnes**  
UNI-REG 17a 1
  - ▶  **Protokolle Fakultät**  
UNI-REG 17a 2
  - ▶  **Dozenten**  
UNI-REG 17a 3
  -  **[leer] / Dekan**  
UNI-REG 17a 4
  - ▶  **Korrespondenz**  
UNI-REG 17a 5
  -  **[leer] / Ordnungen und Reglemente**  
UNI-REG 17a 6

Quelle: <https://dls.staatsarchiv.bs.ch/records/hierarchy?path=0001%2F0002%2F0002%2F0003%2F000E%2F0001&activeNode=413179>



Es ist daher unklar, ob die Dissertation den formalen Anforderungen entspricht, „*eine Originalarbeit der Verfasserin oder des Verfassers*“ zu sein. Streng genommen ist die Dissertation in Teilen keine Originalarbeit, da drei von fünf Papers bereits publiziert wurden. Auch ist die Dissertation in Teilen keine Arbeit des Verfassers, sondern eine Arbeit von mehreren Verfasserinnen und Verfassern.

Zur Klärung wird die Universität Basel ersucht, Bestimmungen zu partiell kumulativen Dissertationen bzw. Präzedenzfälle aus dem entsprechenden Zeitraum vorzulegen.





### III Ergebnis der Überprüfung auf Plagiatsfragmente

Verfasserin: Birgit Kaiserlehner, MA, Team <https://plagiatsgutachten.com/die-plagiatsprüfer>

(Links Martin Rööfli 2001, rechts unziterte Originalquelle)

I wish to thank the many persons, who contributed in different manners to the realization of this work. ~~My sincerest thanks are addressed to my supervisor PD Dr. med. Charlotte Braun.~~

S. iv

~~international setting.~~ I wish to thank the many persons, who contributed in different manners to the realisation of this work.

*OGLESBY 2000, o. S. (gemeinsame Forschungsgruppe und Ko-Autorin)*

~~thanks also to Dr. Hansruedi Moser (LHA) for helpful discussion~~ and all the other former colleagues of the ~~LHA~~ who contributed to the good and stimulating working climate (e.g. by

S. v

and all the other former ~~and current~~ colleagues, who contributed to the good and stimulating working climate ~~at the ISPM.~~

*OGLESBY 2000, o. S.*



~~On the other side in most environmental epidemiologic studies, investigating the effects of long term exposure to particulate matter on health, information on health outcomes and important covariates are collected on an individual level, whereas exposure is assigned on a group level, commonly based on data from one fixed monitoring station per study area. The appropriateness of this~~

### S. 3. In den References:

Künzli, N. and Tager, I. B. (1997): The semi-individual study in air pollution epidemiology: a valid design as compared to ecologic studies. *Environmental Health Perspectives*, **105**, 1078-1083.

~~Finally~~, information bias, or misclassification is related to the instruments and techniques used to collect information on exposure, health outcomes, or other study factors. Nondifferential information bias occurs when the likelihood of misclassification is the same for both groups being compared. When an effect exists, bias from nondifferential misclassification is always in the direction of no effect, and thus is of particular concern in studies that show no association between exposure and disease. Differential information bias occurs when the likelihood of misclassification is different for each comparison group. It can bias the observed effect estimate either toward or away from the null value.

S. 8

both, short-term<sup>1-3</sup> and long-term exposures<sup>1,2,4-6</sup>. A large number of these studies have in common that health outcome and important covariates were measured on an individual level, whereas exposure was assigned on a group or population level, based on fixed monitoring data.<sup>7</sup> ~~Since the early 1980ies, personal exposure measurements~~

7. Künzli, N. and Tager, I.B. "The semi-individual study in air pollution epidemiology: a valid design as compared to ecologic studies," *Environ Health Perspect.* **1997**, *105*, 1078-1083.

*OGLESBY 2000, S. 13*

**Information bias** — This bias, referred to as "misclassification", is related to the instruments and techniques used to collect information on exposure, health outcomes, or other study factors. Nondifferential information bias occurs when the likelihood of misclassification is the same for both groups being compared. When an effect exists, bias from nondifferential misclassification is always in the direction of the association stated in the null hypothesis (i.e., no effect), and thus is of particular concern in studies that show no association between exposure and disease. Differential information bias occurs when the likelihood of misclassification is different for each comparison group. This form of bias is potentially more problematic because it can bias the observed effect estimate either toward or away from the null value.

*MARSH 1995, S. 55 f.*



~~and field trials.~~ A clinical trial is an experiment with patients as subjects whereas a field trial deals with subjects who have not yet gotten the disease. ~~Usually such trials are conducted blinded, ideally double blind, i.e.~~ the individual who makes the assignment, the study subject, and the assessor of the outcome should all be ignorant of the treatment assignment. Blinding prevents certain biases that could affect assignment, assessment or compliance.

S. 9

---

~~epidemiologic.~~ On the one hand it is unethical for an investigator to expose a person to a potential cause of disease simply to learn about etiology. ~~On the~~

S. 9

---

~~The most intuitive study design in epidemiology is~~ the classic cohort study. Two or more groups of people are observed that are free of disease and that differ according to the extent of their exposure to a potential cause of disease.

S. 9

---

A clinical trial is an experiment with patients as subjects. ~~The goal of a clinical trial is~~

Field trials differ from clinical trials in that they deal with subjects who have not yet gotten disease ~~and therefore are not patients. Whereas the patients in a clinical trial may face~~ ~~treatment assignment. Ideally,~~ the individual who makes the assignment, the patient, and the assessor of the outcome should all be ignorant of the treatment assignment. Blinding prevents certain biases that could affect assignment, assessment, or compliance. ~~Most impor-~~

*ROTHMAN, GREENLAND 1998, S. 69 f.*

~~perimental studies in most circumstances.~~ While it is unethical for an investigator to expose a person to a potential cause of disease simply to learn about etiology, ~~people often~~

*ROTHMAN, GREENLAND 1998, S. 72*

---

In the classic cohort study, the investigator defines two or more groups of people that are free of disease and that differ according to the extent of their exposure to a potential cause of the disease. ~~These groups are referred to as the study cohorts (from the Latin~~

*ROTHMAN, GREENLAND 1998, S. 73*

---



~~group.~~ The only requirement is that information on populations such as schools, factories, and cities is available for the exposure and disease distribution in each group. ~~However, group aggregated data are insufficient to control~~

S. 10

~~aggregate studies.~~ The groups may be classes in a school, factories, cities, counties, or nations. The only requirement is that information on the populations studied is available to measure the exposure and disease distributions in each group. ~~Incidence and mortality~~

ROTHMAN, GREENLAND 1998, S. 77

~~for confounding in the analysis and moreover,~~ the degree of association between exposure and disease need not reflect individual-level associations (Greenland and Robins, 1994). Despite such problems ecologic studies can be useful for detecting associations, which may be worthy for further investigations.

S. 10

~~Because the data in ecologic studies are measurements averaged over individuals,~~ the degree of association between exposure and disease need not reflect individual-level associations (Morgenstern, 1982; Richardson et al., 1987; Greenland and Robins, 1994). ~~In questionable validity.~~ Despite such problems, ecologic studies can be useful for detecting associations ~~of exposure distributions with disease occurrence. Even if confounded by~~ worthy of further investigation. ~~A detailed discussion of ecologic studies is presented in~~

ROTHMAN, GREENLAND 1998, S. 78



~~disease occurrence with respect to exposure measurement~~ (Rothman and Greenland, 1998a). However, the prospective/retrospective distinction is sometimes also used to refer to the timing of subject identification, rather than measurement of exposure and covariates. With this usage, a retrospective (or historical) cohort study involves the identification and follow-up of subjects after their follow-up period has ended.

*S. 10. „Bauernopfer-Referenz“: Auf Rothman and Greenland 1998 wird nur einmal verwiesen, obwohl auch die folgenden Sätze aus dieser Quelle stammen.*

---

Two landmark prospective cohort studies reported that chronic exposure to particulate matter pollution increases the risk of premature mortality. ~~The~~  
*S. 12*

The prospective/retrospective distinction is sometimes used to refer to the timing of subject identification, rather than measurement of exposure and covariates. With this usage, a retrospective (or historical) cohort study involves the identification and follow-up of subjects, but the subjects are identified only after the follow-up period under study has ended. ~~The identification of the subjects, their exposure, and their outcome must be based~~

*ROTHMAN, GREENLAND 1998, S. 74 f.*

---

Two landmark prospective cohort studies reported that chronic exposure to particulate pollution increases the risk of premature mortality. ~~In the 1993 Six Cities Study, Harvard~~  
*American Lung Association (2000), <https://www.us-caw.org/pdf/AmerLung.pdf>, S. 2*

---



The time series studies were using different assumptions about the number of days following exposure to air pollution that effects will occur. Some studies have assumed that effects occur the day after exposures (1-day lag). However, toxicological evidence suggests that effects of exposure may be observed over several subsequent days. Schwartz, 2000 demonstrated by means of statistical models that assuming effects continue over several days roughly doubles the relative risk of premature mortality compared to a 1-day lag.

Some have argued that the association between day to day variations in mortality and air pollution represents a "harvesting" effect, that is, the advancement of death by a few days in people already about to die from other causes. If air pollution advances death of the very frail by only a few days, then you would expect that an increase in daily deaths would be followed by a decrease in deaths within a few days, which could not be confirmed by several analyses (Zeger et al., 1999, Schwartz, 2000b).

### S. 13. In den References:

Schwartz, J. (2000): The distributed lag between air pollution and daily deaths. *Epidemiology*, **11**, 320-326.

Zeger, S. L., Dominici, F. and Samet, J. (1999): Harvesting-resistant estimates of air pollution effects on mortality. *Epidemiology*, **10**, 171-175.

*Schwartz 2000b ist nicht in den References.*

Epidemiological studies have used different assumptions about the number of days following exposure to air pollution that effects will occur. Some studies have assumed that effects occur the day after exposures. However, toxicological evidence suggests that effects of exposure may be observed over several subsequent days. In an analysis using data from New Haven,

Spokane, and Seattle, Dr. Joel Schwartz, of the Harvard School of Public Health, has shown that statistical models that assume a one day lag grossly underestimate the effect of PM<sub>10</sub> on mortality. Assuming that effects continue over several days, as demonstrated by this analysis, roughly doubles the relative risk of premature mortality.

pollution are followed by an increased number of deaths. Some have argued that the associations between day to day variations in mortality and air pollution represent a "harvesting" effect, that is, the advancement of death by a few days in people already about to die from other causes. If air pollution advances death of the very frail by only a few days (the "harvesting" hypothesis), then you would expect that an increase in daily deaths would be followed by a decrease in deaths within a few days.

Schwartz, Joel. The Distributed Lag Between Air Pollution and Daily Deaths. *Epidemiology* 2000;11:320-326).

Zeger, S.L., Dominici, F, and Samet, J. Harvesting-Resistant Estimates of Air Pollution Effects on Mortality. *Epidemiology* 1999 Mar;10(2):171-5.

Schwartz, Joel. Harvesting and Long Term Exposure Effects in the Relation between Air Pollution and Mortality. *Am J Epidemiol* 2000; 151:440-8.

*American Lung Association (2000), <https://www.us-caw.org/pdf/AmerLung.pdf>, S. 4-6*





Numerous studies have focused on mortality because it is an easy to measure effect for which data is readily available. It is important to note that early deaths represent just the tip of the iceberg of particulate related health effects. It has been shown that for each death there are many more people admitted to the hospital (Schwartz, 1999), and for each hospital admission, many more visits to emergency departments and doctors offices (Delfino et al., 1997, Medina et al., 1997). Similarly, for each patient who visits an emergency clinic, many more experience uncomfortable respiratory symptoms or days when they must restrict their activity, increase their use of medication, or remain indoors (e.g. Ackermann-Liebrich et al., 1997, Zemp et al., 1999).

#### S. 14. In den References:

- Schwartz, J. (1999): Air pollution and hospital admissions for heart disease in eight US counties. *Epidemiology*, **10**, 17-22.
- Delfino, R. J., MurphyMoulton, A. M., Burnett, R. T., Brook, J. R. and Becklake, M. R. (1997): Effects of air pollution on emergency room visits for respiratory illnesses in Montreal, Quebec. *American Journal of Respiratory and Critical Care Medicine*, **155**, 568-576.
- Medina, S., LeTertre, A., Quenel, P., LeMoullec, Y., Lameloise, P., Guzzo, J. C., Festy, B., Ferry, R. and Dab, W. (1997): Air pollution and doctors' house calls: Results from the ERPURS system for monitoring the effects of air pollution on public health in Greater Paris, France, 1991-1995. *Environmental Research*, **75**, 73-84.

Numerous studies have focused on mortality because it is an easy to measure effect for which data is readily available. It is important to note that early deaths represent just the tip of the iceberg of particulate related health effects. For each death, there are many more people admitted to the hospital, and for each hospital admission, many more visits to emergency departments and doctors offices. Similarly, for each patient who visits an emergency clinic, many more experience uncomfortable respiratory symptoms or days when they must restrict their activity, increase their use of medication, or remain indoors.

Schwartz, Joel. Air Pollution and Hospital Admissions for Heart Disease in Eight U.S. Counties. *Epidemiology* 1999; 10:17-22).

Delfino, R.J., Murphy-Moulton, A.M., Burnett, R.T., Brook, J.R., and Becklake, M.R. Effects of Air Pollution on Emergency Room Visits for Respiratory Illnesses in Montreal, Quebec. *Am J Respir Crit Care Med* 1997; 155:568-576.

Medina, S., Le Tertre, A., Quenel, P., Le Moullec, Y., Lameloise, P., Guzzo, J.C., Festy, B., Ferry, R., and Dab, W. Air Pollution and Doctors' House Calls: Results from the ERPURS System for Monitoring the Effects of Air Pollution on Public Health in Greater Paris, France, 1991-1995. *Environmental Research* 75, 73-84, 1997.

*American Lung Association (2000)*, <https://www.us-caw.org/pdf/AmerLung.pdf>, S. 10 f.



~~vascular system~~ Pope et al., 1999a found little evidence of pollution effects on the oxygen carrying capacity of the blood, but observed that a small elevation in pulse rate was associated with a rise in PM<sub>10</sub> levels. Other studies found that elevated concentrations of fine particulate matter were associated with lower heart rate variability (Pope et al., 1999c, Liao et al., 1999, Gold et al., 2000),

#### S. 14. In den References:

- Pope, C. A., Dockery, D. W., Kanner, R. E., Villegas, G. M. and Schwartz, J. (1999a): Oxygen saturation, pulse rate, and particulate air pollution - A daily time-series panel study. *American Journal of Respiratory and Critical Care Medicine*, **159**, 365-372.
- Liao, D. P., Creason, J., Shy, C., Williams, R., Watts, R. and Zweidinger, R. (1999): Daily variation of particulate air pollution and poor cardiac autonomic control in the elderly. *Environmental Health Perspectives*, **107**, 521-525.
- Gold, D. R., Litonjua, A., Schwartz, J., Lovett, E., Larson, A., Nearing, B., Allen, G., Verrier, M., Cherry, R. and Verrier, R. (2000): Ambient pollution and heart rate variability. *Circulation*, **101**, 1267-1273.

~~changed when suffering from a heart attack or cardiac illness~~. Further, increases in plasma viscosity were observed during an air pollution episode in Central Europe with elevated concentrations of sulfur dioxide, total suspended particulates and carbon monoxide (Peters et al., 1997). ~~Similar effects of~~

#### S. 14. In den References:

- Peters, A., Doring, A., Wichmann, H. E. and Koenig, W. (1997): Increased plasma viscosity during an air pollution episode: A link to mortality? *Lancet*, **349**, 1582-1587.

~~conducted during the winter months, when PM concentrations are highest~~. Researchers found little evidence of pollution effects on the oxygen carrying capacity of the blood, but observed that a small elevation in pulse rate was associated with a rise in PM<sub>10</sub> levels. ~~The medical and~~

~~taken on a weekly basis~~. Both the Baltimore and Boston studies found that elevated concentrations of fine particulate matter were associated with lower heart rate variability, and that

Pope, C.A., Dockery, D.W., Kanner, R.E., Villegas, G.M., and Schwartz, J. Oxygen Saturation, Pulse Rate, and Particulate Air Pollution: A Daily Time-Series Panel Study. *Am J Respir Crit Care Med* 1999; 159:363-372.

Liao, D., Creason, J., Shy, C., Williams, R., Watts, R., and Zweidinger, R. Daily Variation of Particulate Air Pollution and Poor Cardiac Autonomic Control in the Elderly. *Environ Health Perspect* 107:521-525 (1999).

Gold, D.R., Litonjua, A., Schwartz, J., Lovett, E., Larson, A., Nearing, B., Allen, G., Verrier, M., Cherry, R., and Verrier, R. Ambient Pollution and Heart Rate Variability. *Circulation*. 2000;101:1267.

American Lung Association (2000), <https://www.us-caw.org/pdf/AmerLung.pdf>, S. 7 f.

In January 1985, an air pollution episode occurred throughout central Europe, with elevated concentrations of sulfur dioxide, total suspended particulates, and carbon monoxide. ~~During the~~

Peters A., Doring A., Wichmann H.E., and Koenig, W. Increased Plasma Viscosity During an Air Pollution Episode: A Link to Mortality? *Lancet* 1997 May 31; 349(9065):1582-7.

American Lung Association (2000), <https://www.us-caw.org/pdf/AmerLung.pdf>, S. 8





A growing body of literature demonstrates that infants and children are especially sensitive to the effects of fine particle pollution. ~~Associations~~

S. 14

New studies demonstrate that infants and children are especially sensitive to the effects of fine particle pollution.

American Lung Association (2000), <https://www.us-caw.org/pdf/AmerLung.pdf>, S. 1

~~1997). Further it was observed that birth weight, the most important predictor for neonatal mortality, was influenced by the maternal exposure to sulfur dioxide and total suspended particles during the third trimester of pregnancy (Wang et al., 1997). Increased particle concentrations have been associated with acute reductions in lung function, and increased symptoms such as cough, phlegm production, and sore throat (Braun-Fahrlander et al., 1997, Vedal et al., 1998). Children with asthma were found to be more susceptible to these effects than other children. Further, some studies found increased~~

### S. 15. In den References:

Wang, X. B., Ding, H., Ryan, L. and Xu, X. P. (1997): Association between air pollution and low birth weight: A community-based study. *Environmental Health Perspectives*, **105**, 514-520.

Vedal, S., Petkau, J., White, R. and Blair, J. (1998): Acute effects of ambient inhalable particles in asthmatic and nonasthmatic children. *American Journal of Respiratory and Critical Care Medicine*, **157**, 1034-1043.

Low birth weight is the most important predictor for neonatal mortality in developed and developing countries, and is a significant determinant of infant health and survival. ~~A large study and outdoor air pollution in the study region.~~ Xiaobin Wang of the Boston University School of Medicine and colleagues found a significant exposure-response relationship between maternal exposure to sulfur dioxide and total suspended particles during the third trimester of pregnancy and low birth weight.

Increased particle concentrations have been associated with acute reductions in lung function and increased symptom reporting in children, including children with asthma. Dr. Sverre Vedal, ~~respiratory function,~~ and increased symptoms such as cough, phlegm production, and sore throat, after increases in relatively low 24-hour PM<sub>10</sub> concentrations. Children with asthma were found to be more susceptible to these effects than other children.

Wang, X., Ding, H., Ryan, L., and Xu, X. Association Between Air Pollution and Low Birth Weight: A Community-Based Study. *Environ Health Perspect* (1997); 105:514-520.

Vedal, S., Petkau, J., White, R., and Blair, J. Acute Effects of Ambient Inhalable Particles in Asthmatic and Nonasthmatic Children. *Am J Respir Crit Care Med* 1998, Vol. 157, No. 4, 1034-1043.

American Lung Association (2000), <https://www.us-caw.org/pdf/AmerLung.pdf>, S. 12 f.



Studies of migrants provide limited evidence in support of the hypothesis that air pollution is associated with lung cancer due to incomplete control for the effects of smoking and occupational exposure. Migrants from countries with higher lung cancer rates and higher levels of air pollution tend to develop lung cancer at rates higher than those of the new country of residence, suggesting that prior exposure was a risk factor (Speizer and Samet, 1994).

*S. 17. In den References:*

Speizer, F. E. and Samet, J. M. (1994): Air pollution and lung cancer. In: Epidemiology of Lung Cancer (Ed, Samet, J. M.) Marcel Dekker, New York, pp. 131-146.

*„Bauernopfer-Referenz“: Auf Cohen 2000 wird nur im vorhergehenden Absatz verwiesen.*

Studies of migrants reviewed elsewhere (26) provide limited evidence in support of the hypothesis that air pollution is associated with lung cancer risk. Migrants from countries with higher rates of lung cancer and higher levels of air pollution to countries with lower air pollution levels tend to develop lung cancer at rates higher than those of the new country of residence, suggesting that prior exposure was a risk factor. However, incomplete control for the effects of smoking and occupational exposure may account for these findings.

26. Speizer FE, Samet JM. Air pollution and lung cancer. In: Epidemiology of Lung Cancer (Samet JM, ed). New York:Marcel Dekker, 1994;131-146.

*COHEN 2000, S. 745, 750*



~~pollution on development of lung cancer.~~ Most studies found an overall excess risk in the order of 30-50% in the urban areas and larger relative excess risk among nonsmokers (e.g. Katsouyanni and Pershagen, 1997, Hemminki and Pershagen, 1994). The attribution of these results to the effect of ambient air pollution was strengthened by evidence of urban-rural differences in ambient levels of carcinogens such as benzo[a]pyrene or urban-rural gradients in the mutagenicity of airborne particulate matter across the United States (Hannigan et al., 1997, Pedersen et al., 1999). However, the influence of smoking on these findings is controversial (Doll and Peto, 1981). The urban factor may reflect influences instead of, or in addition to, outdoor air pollution. These could include indoor air pollution, patterns of migration occupational exposure or factors related to population density.

In ecologic studies of residential proximity to diverse industrial sources (e.g. petrochemical plants and steel mills) generally an increased rate of lung cancer was observed but in these studies it was not possible to control for confounders at the individual level (Pershagen, 1990). However, Archer, 1990 took advantage of a "natural experiment" by comparing two counties in Utah, similar in many respects: low smoking rates as well as low and nearly equal respiratory cancer mortality rates until a steel mill was constructed during World War II. The subsequent differences in incidences of lung cancer cases were substantial within about 15 years after the increase in air pollution and have persisted.

S. 17 f.

~~urban and rural environments (Table 36.4).~~ Most studies found overall excesses on the order of 30–40% in the urban areas. The attribution of these results to differences in air quality was strengthened by evidence of urban/rural differences in ambient levels of carcinogens such as benzo[a]pyrene and by the frequent persistence of the urban/rural differences after adjustment for cigarette smoking. ~~In addition, the studies of Buell et al.~~

~~urban/rural differences (IARC, 1997).~~ However, the so-called 'urban factor' may reflect other influences instead of, or in addition to, outdoor air pollution: these could include indoor air pollution, patterns of migration, or factors related to population density.

~~and 2.0 in males.~~ Ecologic studies of residential proximity to diverse industrial sources (e.g. petrochemical plants and steel mills) have generally observed increased relative risks of lung cancer, but have been unable to control for level confounders at the individual level, ~~such as cigarette smoking and employment at the industrial facility itself.~~

urban/rural studies. Taking advantage of a 'natural experiment,' Archer (1990) analyzed respiratory cancer mortality in two Utah counties with very low smoking rates. The counties were similar in many respects, with low and nearly equal respiratory cancer mortality rates, until a steel mill constructed during World War II ~~caused substantial~~

~~increases in air pollution in one of them.~~ The subsequent differences in lung cancer were substantial within about 15 years after the increase in air pollution and have persisted.

*SAMET, COHEN 1999, S. 855 f., 852, 856 f.*



Most published case-control studies found relative increases of lung cancer risks after adjustment for age, smoking, and occupational exposure similar to those observed in the urban-rural and ecologic studies (Pike et al., 1979; Vena,

S. 18

---

Biomarkers offer a new approach to quantifying the lung cancer risk associated with air pollution. Potential biomarkers for lung cancer include actual levels of the putative carcinogen in biologic materials, DNA adducts of potential carcinogens or metabolites, and antibodies against such adducts (Schulte and Perera, 1993). ~~This type of study has focused mainly on the~~

S. 18. In den References:

Schulte, P. A. and Perera, F. P. (1993): *Molecular Epidemiology: Principles and Practices*, Academic Press, New York.

~~the exposures of study subjects (Table 36.5). Most found relative increases of lung cancer risks after adjustment for age, smoking and occupational exposure similar to those observed in the urban/rural and ecologic studies. Dockery and colleagues (1993) recently~~

SAMET, COHEN 1999, S. 857

Exposure biomarkers – that is, indicators of exposure or dose measured in biological material – offer a new approach to quantifying the lung cancer risk associated with air pollution. Potential biomarkers for lung cancer risk include actual levels of the putative carcinogen in biological materials, adducts of potential carcinogens or metabolites down to DNA, and antibodies against such adducts (NRC, 1989; Schulte and Perera, 1993).

Schulte PA and Perera FP (1993) In: Schulte PA and Perera FP (eds) *Molecular Epidemiology: Principles and Practices*. New York: Academic Press.

SAMET, COHEN 1999, S. 854, 863



(Steenland and Savitz, 1997). In both the classical and Berkson models, one usually assumes a specific distribution for the error term in the model, for example a normal distribution with mean 0 and variance  $\sigma_e^2$  corresponding to a nondifferential misclassification. If so, it is well known that in ordinary linear regression with the classical error model the estimated exposure response relation is biased to the null value. This bias to the null will generally also hold for log linear regression models. However, if the assumptions of the classic error model are violated because the measurement error is correlated with the true value (differential misclassification), then the linear regression coefficient can be biased in every direction. In contrast, it is less well appreciated that if the Berkson model holds, then the estimate of exposure effect obtained by ordinary linear regression is in fact unbiased when the error is not related to the exposure level (Berkson, 1950). However the standard error is increased, resulting in less power or precision. On the other hand, when the variance of the errors is not constant, measurement error can lead to bias away from the null value in log linear regression (e.g. logistic, Cox, Poisson regression) when the Berkson error model holds. This would occur if the errors were greater for larger values of the observed variable, a situation that may occur in environmental studies. Of course the study result is also biased

*S. 22 f. Bauernopfer-Referenz“: Auf Steenland and Savitz 1997 wird nur einmal verwiesen, obwohl fast der gesamte restliche Absatz aus dieser Quelle stammt.*

In both the classical and Berkson models, one also usually assumes a specific distribution for the error term in the model, for example a normal distribution with mean 0 and variance  $\sigma_e^2$  (Thomas et al., 1993).

It is well known that in ordinary linear regression with the classic error model and normal error distribution, the estimate obtained by using the observed  $X$  values in regression is always biased toward the null value ~~servative” estimate of the true effect of  $Z$ . Furthermore, this same bias toward the null will generally hold for log linear regression (e.g., logistic~~ when the assumptions of the classic error model are violated because the measurement error is correlated with the true value (i.e.,  $X = Z + e$ , where ~~measure~~ (e.g., the exposure regression coefficient in a model relating exposure to disease) is always toward the null. When the correlation is negative, the resulting effect bias can be either toward the null, away from the null, or even in the opposite direction. ~~Wacholder advises using cau~~ ~~with the classic error model leads to bias toward the null,~~ it is less well appreciated that if the Berkson model holds, then the estimate of exposure effect (the coefficient for  $X$ ) obtained by ordinary linear regression obtained by using the observed  $X$  values is in fact unbiased (Berkson, 1950). However the standard error is increased, resulting in less power or precision.

On the other hand, in log linear regression (e.g., logistic, Cox, Poisson regression) when the Berkson error model holds, measurement error can lead to bias away from the null value, but only if the variance of the errors is not constant. For example, this would occur if the errors were greater for larger values of the observed variable  $X$ ; this situation would not be uncommon in environmental studies. ~~For example, this might occur for~~

*STEENLAND, DEDDENS (In: STEENLAND; SAVITZ) 1997, S. 22 f.*





~~into account that~~ for a long-term effect such as lung cancer, the benefit of lower air pollution levels would take years to be fully realized (Künzli et al., 2000). Further, one has to consider the fact that the attributable risk estimate does not take competing risk into account. Thus, it is well known in multicausal diseases that the sum of attributable cases across several risk factors may be larger than 100 percent (see Chapter 3.7). ~~It is therefore warranted of~~

*S. 137 f. „Bauernopfer-Referenz“: Auf Künzli et. al. 2000 wird nur einmal verwiesen, obwohl auch die folgenden Sätze aus dieser Quelle stammen. Verweise auf die bei Künzli et al. 2000 angegebenen Quellen fehlen.*

~~cancer cases can be caused by smoking.~~ Impact measures that take competing risks into account need to be developed, first.

*S. 138. Ein Verweis auf die bei Künzli et al. 2000 angegebene Quelle fehlt.*

~~with such an interpretation.~~ First, for long-term effects—the benefit of lower air-pollution levels—would take years to be fully realised.<sup>8</sup> Second, the attributable risk estimate does not take competing risks into account.

Accordingly, it is well known in multicausal diseases that the sum of attributable cases across several risk factors will not add up to 100% but may be larger.<sup>28</sup> ~~Impact~~

8 Sommer H, Chanel O, Vergnaud JCh, Herry M, Sedlak N, Seethaler R. Monetary valuation of road traffic related air pollution: health costs due to road traffic-related air pollution: an impact assessment project of Austria, France and Switzerland third WHO Ministerial Conference of Environment & Health. London: WHO, 1999.

28 Smith K, Corvalan C, Kjellström T. How much global ill health is attributable to environmental facators? *Epidemiology* 1999; **10**: 573–84.

*KÜNZLI et al. 2000, S. 800 f.*

~~will not add up to 100% but may be larger.~~<sup>28</sup> Impact measures that take competing risks into account need to be developed.<sup>29</sup> ~~It is to be emphasised that the economic~~

29 La D, Hardy R. Potential gains in life expectancy or years of potential life lost: impact of competing risks of death. *Int J Epidemiol* 1999; **28**: 894–98.

*KÜNZLI et al. 2000, S. 800 f.*



## Verzeichnis der plagiierten Quellen

Aaron J. COHEN, Outdoor Air Pollution and Lung Cancer. In: Environmental Health Perspectives 108, Supplement 4 (2000), 743-750. **Im Literaturverzeichnis angegeben.**

American Lung Association, Selected Key Studies on Particulate Matter and Health: 1997 – 2000. New Studies Confirm that Current Levels of Particulate Air Pollution are Harmful to Human Health (o. O. 2000), online unter: <https://www.uscaw.org/pdf/AmerLung.pdf> (abgerufen am 10.04.2023). **Nicht im Literaturverzeichnis angegeben.**

Nino KÜNZLI, Reinhard KAISER, Sylvia MEDINA, u. a. (2000), Public Health Impact of Outdoor and Traffic Related Air Pollution. In: The Lancet 356 (2000), 795-801. **Im Literaturverzeichnis angegeben.**

Gary M. MARSH, Statistical Issues in the Design, Analysis, and Interpretation of Environmental Epidemiologic Studies. In: Evelyn O. TALBOTT, Gunther F. CRAUN (Hrsg.), Introduction to Environmental Epidemiology (Boca Raton 1995), 47-62. **Nur einmal im Fließtext, aber nicht im Literaturverzeichnis angegeben.**

Lucy OGLESBY, Measures of exposure in air pollution epidemiology and health risk assessment (Dissertation, Eidgenössische Technische Hochschule Zürich 2000). **Nicht im Literaturverzeichnis angegeben.**

Kenneth J. ROTHMAN, Sander GREENLAND, Types of Epidemiologic Studies. In: Kenneth J. ROTHMAN, Sander GREENLAND (Hrsg.), Modern Epidemiology (2. Aufl., Philadelphia 1998), 67-78. **Im Literaturverzeichnis angegeben.**



Jonathan M. SAMET, Aaron J. COHEN, Air Pollution and Lung Cancer. In: Robert L. MAYNARD, Stephen T. HOLGATE, Hillel S. KOREN u. a. (Hrsg.), Air Pollution and Health (San Diego 1999), 841-864. ***Nicht im Literaturverzeichnis angegeben.***

Kyle STEENLAND, James A. DEDDENS, Design and Analysis of Studies in Environmental Epidemiology. In: Kyle STEENLAND, David A. SAVITZ (Hrsg.), Topics in Environmental Epidemiology (New York 1997), 9-27. ***Im Literaturverzeichnis angegeben.***





# IV Ergebnis der Überprüfung der empirischen Teile der Dissertation

**Verfasserin: MMag. Dr. ETH Sabrina Dorn, MSc ETH, <https://www.knowledgeconsult.at>**

## 1. Inhaltlicher Überblick

Die vorliegende Dissertation untersucht die zeitliche und räumliche Variabilität von Luftschadstoffen in einer städtischen Umgebung (Region Basel, Schweiz), die chemische Zusammensetzung von Feinstaub sowie das von der Belastung mit krebserzeugenden Luftschadstoffen (mittels PM10 als Indikator) ausgehende Gesundheitsrisiko (siehe S. 3-4, S. 5 und S. 131). Die in dieser Dissertation untersuchten Daten zu Luftschadstoffen wurden im Rahmen der BRISKA-Studie (Basler Risikostudie Aussenluft unter Beteiligung des Instituts für Sozial- und Präventivmedizin der Universität Basel und des Lufthygieneamts Basel) erhoben. Darüber hinaus wurden zusätzlich Daten (Referenzmessungen) zwischen 1997 und Mai 1999 an der permanenten NABEL-Station und an der fixen Messstation St. Johann des Lufthygieneamts Basel miteinbezogen (siehe S. 36-37). Aus vorangegangener Literatur wird angenommen, dass sich Luftverschmutzung gesundheitlich auswirkt und, dass dabei Feinstaubpartikeln die stärkste gesundheitliche Auswirkung zukommt.

Vor dem Hintergrund, dass in der Schweiz 1998 die Luftqualitätsnorm für TSP durch eine für PM10 ersetzt wurde, spricht sich die Arbeit als regulatorische Empfehlung für eine weniger auf einzelne Schadstoffe gerichtete Luftschutzpolitik aus (siehe S. 145).

Kapitel 1 bis 5 motivieren die Fragestellung, erklären die Ziele der Dissertation, geben einen Überblick über vorangegangene Literatur und erklären das technische Konzept zur Messung von Feinstaub. Die Kapitel 6 bis 8 präsentieren und analysieren mittels statistischer Methoden die im Rahmen des BRISKA Projekts zuzüglich Referenzmessungen erhobenen Daten. Kapitel 6 wurde in Journal of Air & Waste Management Association 50:



1115-1124 (2000), Kapitel 7 in *Atmospheric Environment* 34: 3171-3181 (2000) und Kapitel 8 in *Atmospheric Environment* 35: 3701-3713 (2001) publiziert. Die Kapitel 9 und 10 enthalten eine Schätzung des Gesundheitsrisikos durch PM10. Dafür wurde in Kapitel 9 ein Modell entwickelt (als *epidemiology-based risk method* bezeichnet) und in Kapitel 10 auf ein extensiveres Datenbeispiel zu Basel angewandt. Kapitel 9 wurde in *Human and Ecological Risk Assessment: An International Journal*, 9:3, 709-719 (2003) und Kapitel 10 in *Journal of Occupational and Environmental Medicine* 45(7): 715-723 (2003) publiziert. Nach verdeckter Anfrage beim Autor der Dissertation wurden Datenfragmente übermittelt. Diese erwiesen sich jedoch als zu unvollständig um die Ergebnisse der Kapitel 6 bis 8 im Detail zu replizieren. Die Berechnungen aus diesen Kapiteln lassen sich jedoch intersubjektiv im Groben nachvollziehen und es besteht auch kein offensichtlicher Zweifel an deren grundsätzlicher Plausibilität. Mithilfe der übermittelten Daten ließen sich jedoch Behelfs-Datensätze konstruieren, anhand derer die wesentlichen empirischen Hypothesen aus den Kapiteln 6 bis 8 bezüglich PM10 hinsichtlich ihrer qualitativen Implikationen empirisch überprüft werden konnten.

In der Untersuchung zu Kapitel 6 zeigte sich ein robust stark positiver und signifikanter Zusammenhang zwischen der PM10-Messung an der Referenzstation NABEL und der am jeweiligen Ort platzierten mobilen Station. Dies entspricht qualitativ den Ergebnissen der Dissertation. Dieser Zusammenhang ist auch robust, wenn man – wie in der hier unternommenen empirischen Plausibilitätsprüfung – einen autoregressiven Zusammenhang für PM10 über die Zeit erlaubt. Wird für Unterschiede in meteorologischen Bedingungen nicht-parametrisch für Monate und durch einen linearen Trend kontrolliert, so steigt PM10 robust, sowohl mit steigendem Personen- als auch Lastverkehr. Kapitel 6 kommt diesbezüglich zu einer anderen Schlussfolgerung, deren empirische Robustheit damit zumindest teilweise anzuzweifeln ist. Auch die Replikationsübung mit den Daten zu Kapitel 8 legt die gleiche empirische Schlussfolgerung nahe, dass Verkehrsbelastung eine robust signifikant positive Determinante der PM10-Konzentration ist.



Zur Bewertung des Gesundheitsrisikos durch Luftverschmutzung in der Region Basel in den Kapiteln 9 und 10 wird ein Ansatz mit PM10 als Indikator für Kanzerogenität gewählt und mit einem anderen Ansatz auf der Grundlage von unit risk factors verglichen. Auf der Grundlage der Ergebnisse amerikanischer Langzeitstudien und eines meta-analytischen Ansatzes wird der Zusammenhang zwischen PM10 und Gesundheitsfolgen (Lungenkrebs) für Basel untersucht (siehe S. 40). Das mittels Excess Rate (ER) ausgedrückte Wirkungsmaß aus Kapitel 9 wurde dazu in Kapitel 10 in einer Kalibrierung auf die Basler Bevölkerung angewandt.

Die Plausibilität des in den Kapiteln 9 und 10 verwendeten epidemiologischen Ansatzes hängt dabei wesentlich davon ab, ob die Daten den für die verwendete Taylorreihen-Approximation in dem Sinn zulässigen Wertebereich einnehmen, dass davon auszugehen ist, dass der Näherungsfehler durch diese nur gering ist. Wie im Rahmen einer hier angestellten Simulation festgestellt wird, erfüllen zwischen rund 60-90% der Daten diese wesentliche Annahme nicht. Dies wäre somit auch eine plausible Erklärung dafür, dass die von den Autoren vorgeschlagene *epidemiology-based risk method* ein erheblich höheres Krebsrisiko prognostiziert als Vergleichsschätzungen aus der Literatur (vgl. S. 117). Die mögliche Überschätzung des Krebsrisikos mittels des epidemiologiebasierten Ansatzes könnte durch diesen Bias einer für den Wertebereich der Daten zu einer schlechten Annäherung führenden Approximation erklärt werden (vgl. *ibid.*).

Insgesamt ziehen sich Fehler und Ungenauigkeiten bei der Berechnung der statistischen Ergebnisse – insbesondere bei der Inferenz – durch die gesamte Arbeit. Ebenso ist die statistische Modellierung nicht stets nachvollziehbar bzw. wird deren Robustheit nicht entsprechend seitens der Autoren belegt. Dabei dürfte es sich insgesamt um leichte bis größere Fahrlässigkeiten handeln, welche die inhaltlichen Ergebnisse der Kapitel 6 bis 8 teilweise berühren. Die Schlussfolgerung einer stark positiven Korrelation zwischen Referenz- und mobiler Messstation bleibt jedenfalls erhalten, jedoch wird durch die Replikationsübungen anhand der Behelfs-Datensätze ein stark signifikant positiver Effekt des



Verkehrsaufkommens auf die PM10-Belastung nahegelegt. Ein solcher wird in der Dissertation nur wage bestätigt bzw. werden Unterschiede (qualitativ bzw. narrativ) eher mit meteorologischen Phänomenen in Zusammenhang gebracht.

Tatsächlich inhaltlich in Zweifel zu ziehen sind die Berechnungen aus Kapitel 9 und 10 aufgrund der unpassenden Taylorreihen-Approximation. Es ist nicht klar, warum dies übersehen wurde, aber eine bewußte Manipulation ist dem eher nicht zu entnehmen, obwohl dieses Problem bei sorgfältiger Vorgehensweise eigentlich hätte auffallen müssen (zumindest im Peer-Review). Anzumerken ist, dass die Autoren die diesbezüglichen Ergebnisse stets sehr vorsichtig und damit angemessen interpretieren.

Auf die Datenanfrage hin, hat Prof. Rösli entsprechend guter wissenschaftlicher Praxis unmittelbar reagiert. Auch wenn der Datensatz nicht mehr zur Gänze vorhanden zu sein scheint, haben sich daraus die qualitativen Aussagen aus den Kapiteln 6 bis 8 im Großen und Ganzen replizieren bzw. plausibilisieren lassen.

## **2. Überblick Überprüfungsverfahren**

### **2.1 Äußerer Eindruck der Ergebnisse**

Jede detaillierte Untersuchung zu den Kapiteln 6 bis 10 enthält eine Auflistung der in der Dissertation an der jeweiligen Stelle verwendeten Tabellen und Abbildungen mit kurzen und längeren Anmerkungen, falls sich solche ergaben. Dies gibt einen Überblick bezüglich der intersubjektiven Nachvollziehbarkeit der präsentierten Ergebnisse.

### **2.2 Daten zur Replikation der empirischen Ergebnisse der Kapitel 6 bis 8**

Mittels eines Mystery-Shopping-Ansatzes wurde Prof. Rösli um die Daten zu den Aufsätzen in den Kapiteln 6 bis 8 gefragt. Dieser retournierte daraufhin vier xls-Files (IN-HALT.XLS, Matrixpm.xls, NABEL98.XLS, PM10-98.XLS) mit der Anmerkung, dass er sich selbst nicht mehr so ganz erinnern könnte. Festzuhalten ist, dass keines der Files die in Kapitel 6 bezeichneten meteorologischen Variablen enthält, sodass die Ergebnisse zu



Gleichung (6-1) auf S. 51 nicht direkt überprüft werden können. Die Daten weisen teilweise Lücken auf, lassen sich jedoch mit Hilfe der Informationen aus der Dissertation zuordnen.

Die übermittelten Daten wurden in Folge gemeinsam mit einschlägigen in der Dissertation gegebenen Informationen dazu verwendet Behelfs-Datensätze (**Datensatz 1**, **Datensatz 2** und **Datensatz 3** im Folgenden) zu konstruieren um anhand dieser einige der qualitativen Ergebnisse der Kapitel 6 bis 8 bezüglich PM10 empirisch zu überprüfen:

- Die Daten im File Matrixpm.xls können mittels der Informationen aus Abbildung 5-1, S. 36 auf die sechs Messorte und auf eine Charakterisierung nach dem Verkehrsaufkommen mit Hilfe von Tabelle 8-1 auf S. 87 zurückgeführt werden. Daten für die Referenzmessung zur permanenten Station NABEL sind ebenfalls gegeben. (Im Folgenden **Datensatz 1** bezeichnet.)
- Die Daten im File Matrixpm.xls können mittels der Informationen aus Tabelle 7-1, S. 69 auf unterschiedlich starke Verkehrsaufkommen für die Stationen Wasgenring, Luzernerring und Sevogelstrasse zurück geführt werden. Daten für die Referenzmessung zur permanenten Station NABEL sind ebenfalls gegeben. Um das Verkehrsaufkommen zu messen, wird aus den Informationen aus Tabelle 7-1 eine ordinale Variable konstruiert, welche mit dem Verkehrsaufkommen laut Tabelle 7-1 (Sevogelstrasse < Wasgenring < Luzernerring) zunimmt. (Im Folgenden als **Datensatz 2** bezeichnet.)
- Zusätzlich dazu kann aus den mitgelieferten Datumsinformationen eine binäre Variable für Wochenenden (Samstage und Sonntage) versus Werkstage konstruiert werden (vgl. insbes. Kapitel 6 und 8)
- Um für unterschiedliche meteorologische Bedingungen zumindest approximativ zu kontrollieren, wird eine binäre Variable für einen hohen Temperatur-Gradienten (vgl. Abbildung 6-5, S. 61) in den Monaten November bis März erzeugt, sowie alternativ werden Dummy-Variablen für jeden Monat (mit Bezugsmonat April in den Regressionen unten)



spezifiziert und schließlich wird zudem eine Trend-Variable erzeugt um für einen nicht näher zu bezeichnenden lineare zeitlichen Trend zu kontrollieren.

- Zeitreihendaten, die eine genauere Analyse von PM10 über die Zeit enthalten, sind im File NABEL98.XLS enthalten. (Im Folgenden **Datensatz 3** bezeichnet.)

### **2.3 Simulation**

Um die Ergebnisse der Risikobewertung in den Kapiteln 9 und 10 zu evaluieren, wird anhand des Wertebereichs der empirischen Daten, wie in der Dissertation genannt, die Qualität der den Berechnungen im Appendix zu Kapitel 9 der Dissertation (S. 114-115) zugrunde liegenden Approximation überprüft. Auch können Werte in den Kapiteln 9 und 10 allgemein mittels der gegebenen Formeln und empirischen Größen durch “manuelle Berechnungen” überprüft werden.

## **3. Spatial Variability of Different Fractions of Particulate Matter Within an Urban Environment and Between Urban and Rural Sites (Kapitel 6)**

### **3.1 Zusammenfassung und Würdigung**

In Kapitel 6 sollen die nicht zeitgleich an den temporären Standorten gemessenen Feinstaubkonzentrationen durch eine Regressionsanalyse vergleichbar gemacht werden (siehe S. 38). Auf S. 33 wird darauf verwiesen, dass laut einer anderen Studie des Autors (Rööslü et al., 1999) die tägliche Variabilität der Feinstaubkonzentration besonders stark von meteorologischen Parametern beeinflusst wird, die sich auf die Verdünnung auswirken und damit die täglichen Messungen von Partikeln in der Luft nicht nur von der Emissionsquelle, sondern auch in hohem Maße von den vorherrschenden meteorologischen Bedingungen beeinflusst werden. Die vom Autor übermittelten Daten erlauben keine exakte Replikation dieser Ergebnisse, da Werte für die meteorologischen Variablen fehlen.

Im Ergebnis wurde die räumliche Variabilität von PM4, PM10 und TSP im Gebiet von Basel als eher beschränkt beurteilt und fast zur Gänze durch die Station NABEL erklärbar. Diese



Beobachtung lässt sich anhand von Datensatz 1 exemplarisch für PM10 qualitativ replizieren. Der hohe Erklärungsgehalt der NABEL-Messwerte ergibt sich ebenso hier und scheint auch nicht durch eine mögliche Autokorrelation der Messwerte getrieben zu sein; siehe dazu 3.3, Abschnitt B) unten.

Auch wenn in Kapitel 6 ein nur schwach positiver Effekt des Verkehrsaufkommens auf die PM10-Konzentration behauptet wird, wobei eingeräumt wird, dass menschliche Aktivitäten, einschließlich des Verkehrs, Auswirkungen auf die PM-Konzentrationen in der Luft haben können, lässt sich dieser Befund nicht so aus Datensatz 1 replizieren. Unter der Einschränkung, dass für meteorologische Bedingungen anhand von Proxy-Variablen kontrolliert wird, ergibt sich für die Messstation aus Datensatz 1, dass die Verkehrsexposition einen hoch signifikant positiven Effekt auf die PM10-Konzentration hat (siehe die zugehörigen Berechnungen in 3.3, Abschnitt B) unten). Aufgrund der Ordinalität der konstruierten Behelfsvariable, kann dem jedoch keine sinnvolle quantitative Schätzung des Effekts entnommen werden.

Die Argumentation der Autoren erfolgt überwiegend qualitativ und die statistischen Modelle haben einige Schwächen betreffend unterstellter funktionaler Form und Inferenz (Verdacht auf Heteroskedastie und Autokorrelation der Residuen, Nicht-Normalverteilung der Residuen; keine Transformation der abhängigen Variable). Seitens der in Kapitel 6 durch die Autoren angestellten Regressionen, ist es nicht nachvollziehbar, warum in keiner Spezifikation Verkehrsexposition simultan mit meteorologischen Variablen als mögliche Gründe für die Konzentrationen von PM4, PM10 und TSP untersucht wurde. Es ist daher anzunehmen, dass der in Kapitel 6 behauptete Zusammenhang teilweise nicht robust ist.

### **3.2 Übersicht**

1. Tabelle 6-1, S. 49; Inhalt: Datenbeschreibung
2. Tabelle 6-2, S. 54; Inhalt: Feinstaubkonzentration mobile Messstationen (1-6) mit Verkehrsexposition versus Messstation am Berg (NABEL); Anmerkungen: Zwei Arten von  $R^2$ , die nicht genauer erläutert sind; Regressionsmodell ohne Angabe der weiteren



erklärenden Variablen; nicht replizierbar wegen fehlender erklärender Variablen; Ergebnisse wirken aufgrund der extrem hohen positiven Korrelation ungewöhnlich, können jedoch qualitativ repliziert werden; Autokorrelation in den Residuen wahrscheinlich; Log-Transformation der abhängigen Variable wegen Normalverteilungsannahme scheint aus Rohdaten erboten.

Die in Tabelle 6-2 auf S. 54 dargestellten Regressionsergebnisse beruhen auf der in Gleichung (6-1) auf S. 51 gegebenen Spezifikation.

Diese unterstellt einen linearen Zusammenhang zwischen der gemessenen Feinstaubkonzentration (PM<sub>4</sub>, PM<sub>10</sub>, TSP) durch eine mobile Messstation (Verkehrsexposition) und der am Berg aufgestellten permanenten Messstation (NABEL; keine Verkehrsexposition), sowie zwischen der gemessenen Feinstaubkonzentration und anderen meteorologischen Variablen. Es wird nicht weiter gezeigt, ob diese funktionale Annahme durch die Daten auch erfüllt ist (ggf. durch eine weniger restriktive Spezifikation oder eine Residuenanalyse der geschätzten Regressionsmodelle). Dieser Zusammenhang kann aus Datensatz 1 qualitativ repliziert werden (siehe Abschnitt 3.3 unten).

Zur Ermittlung der Ergebnisse in Tabelle 6-2 auf S. 54 und bezüglich der dieser vorangegangener Variablenselektion ist unklar, warum hier sechs Einzelmodelle mit jeweils 52 Beobachtungen und nicht ein gepooltes (d.h. gemeinsames) Modell (mit 312 Beobachtungen für 6 mobile Messungen; ggf. zuzüglich 52 für NABEL) geschätzt wurde. Aufgrund der jeweils sehr kleinen Stichprobengrößen für die Variablenauswahl (7 meteorologische Variablen insgesamt, vgl. S. 51) von je nur 52 Beobachtungen, ist es nicht unwahrscheinlich, dass (getrieben durch die Stichprobengröße) eigentlich relevante Variablen nicht erkannt wurden (52 Beobachtungen pro Modell mit Parametern  $a$ ,  $b$  und 7 weiteren in „ $m$ “, d.h. 9 Parameter mit 43 Freiheitsgraden der Residuen, was auch in Hinblick auf die Approximation durch die Normalverteilung für die Inferenz kritisch zu sehen ist). Darüber hinaus ist nicht auszuschließen, dass durch multiples Testen Variablen fälschlicherweise als





signifikant erkannt wurden, die es eigentlich nicht sind.

Auch ist zu berücksichtigen, dass Abbildung 6-4 auf S. 60 nahe legt, dass die Schadstoffwerte über die Zeit korreliert sind, sodass auch die Residuen der Regressionen in Tabelle 6-2 auf S. 54 korreliert sein könnten, was für die Ermittlung der statistischen Signifikanz bzw. Berechnung der Standardfehler Berücksichtigung finden muss. Auf diesen die Inferenz möglicherweise verzerrenden Zusammenhang wird in der Arbeit nicht eingegangen. Weiters ist zu berücksichtigen, dass auch die in Tabelle 6-2 auf S. 54 präsentierten Korrelationskoeffizienten nur lineare Zusammenhänge erfassen können.

3. Tabelle 6-3, S. 56; Inhalt: Durchschnittskonzentrationen, Variationskoeffizient; Anmerkungen: Schadstoffe Unklarheit wie dies ermittelt wurden; keine genauen Angaben über die Berechnung der Konfidenzintervalle oder statistisches Validieren einer etwaigen Verteilungsannahme

Tabelle 6-3 auf S. 56 gibt laut Autoren die Durchschnittskonzentrationen der Schadstoffe für die sechs mobilen Messstationen laut Regressionsmodell wieder. Es ist dabei aber unklar, wie diese ermittelt wurden. Um Durchschnittswerte für konstante meteorologische Bedingungen und einen gegebenen Messwert von NABEL zu ermitteln, müssten alle Regressionsmodelle für ein einheitliches Set an erklärenden Variablen prognostiziert werden. Wurden diese Durchschnittswerte aber anhand des jeweiligen Fits der abhängigen Variable in der jeweiligen Unterstichprobe ermittelt, so sind diese Werte nicht untereinander vergleichbar. Wiederum ist nicht klar, warum hier kein gepooltes Modell geschätzt wurde bzw. warum hier keine Variablen für die Verkehrsbelastung miteinbezogen wurden.

4. Tabelle 6-4, S. 57; Inhalt: Regressionsergebnisse Verkehrsaufkommen; Anmerkungen: Unklarheit bezüglich exakter Spezifikation (welche weiteren erklärenden Variablen?)



5. Abbildung 6-1, S. 57; Inhalt: Zusammenhang Partikel und Verkehrsaufkommen; Anmerkungen: Keine Konfidenzbänder, Überprüfung mit nicht parametrischer Spezifikation?

Tabelle 6-4 und Abbildung 6-1 auf S. 57 legen nahe, dass zwischen Feinstaub (PM<sub>10</sub>, PM<sub>4</sub> und TSP) und der Verkehrsdichte nur ein schwach positiver Zusammenhang besteht. Zu Tabelle 6-4 wird nicht erläutert auf Basis welcher Stichprobengröße und welcher Spezifikation diese Ergebnisse zu Stande kommen. Es macht den Eindruck, als hätten die als bedeutsam ermittelten meteorologischen Variablen aus Tabelle 6-2 (S. 54) keine Berücksichtigung gefunden. Darüber hinaus unterstellt die Spezifikation einen linearen Zusammenhang zwischen Verkehrsdichte und Schadstoffkonzentration, jedoch wird nicht weiter dargelegt, ob ein solcher Zusammenhang empirisch überhaupt angenommen werden darf. Auch eine Residuenanalyse, inwiefern das zugrunde liegende empirische Modell die Daten adäquat beschreibt bzw. ob die Inferenz korrekt ist, wurde nicht angestellt.

Abbildung 6-3 (S. 57) legt für die Daten für die Messorte A, B, C und D nahe, dass die Varianz mit der Messstation variiert. Dies könnte auch auf die Daten aus Tabelle 6-4 zutreffen. Ist dies so, dann liegt Heteroskedastie in den Daten vor, was im Rahmen der Inferenz berücksichtigt werden muss, da dies sonst zu verzerrten Schätzungen für die Standardfehler bzw. Konfidenzintervalle führt. Wie oben bereits in anderem Kontext erwähnt, könnten die Residuen aus den dort erläuterten Gründen zeitlich korreliert sein, was wiederum zu verzerrten Konfidenzintervallen führt, wenn dies keine Berücksichtigung findet.

Die Autoren argumentieren, dass sich die Messungen zwischen den sechs temporären Messstationen mit Verkehrsexposition (1-6 in Tabelle 6-1) nur geringfügig von der fixen Messstation NABEL (ohne Verkehrsexposition unterscheiden) und darüber hinaus nur ein schwach positiver Zusammenhang zwischen Verkehrsdichte und Schadstoffkonzentration besteht. Es ist unklar warum hierzu ein Variationskoeffizient geschätzt wird und nicht



zusätzlich auf statistisches Testen bzw. ein gepooltes Modell zurückgegriffen wurde.

6. Abbildung 6-2, S. 58; Inhalt: Partikelkonzentration und Wochentage; Anmerkungen: Ohne Originaldaten nicht überprüfbar; keine Angabe über Annahmen für Inferenz oder Art des Konfidenzintervalls
7. Abbildung 6-3, S. 59; Inhalt: PM10-Konzentrationen an Messstationen A, B, C, D
8. Abbildung 6-4, S. 60; Inhalt: Monatliche Werte für PM10 für Messstationen A, B, C, D Anmerkungen: keine Konfidenzbänder
9. Abbildung 6-5, S. 61; Inhalt: Zusammenhang Temperaturgradient und PM10-Unterschied zwischen Messstation A und C; Anmerkungen: Keine Konfidenzbänder; keine detaillierte Regressionsanalyse und nur bivariater Zusammenhang veranschaulicht

### 3.3 Berechnungen

#### A) Deskriptive Statistik Datensatz 1

Die Konstruktion des Datensatzes erfolgte entsprechend der Ausführungen in Abschnitt 2.2; Datensatz 1 besteht aus 315 Beobachtungen für 6 Messorte, die mithilfe von Abbildung 5-1 und Tabelle 6-1 dem Verkehrsaufkommen nach Personenverkehr (PV) und Lastverkehr (LV) (ordinal) zugeordnet werden konnten.

##	Messort	Datum	PM10	NABEL	
##	Claraplatz	:52 Min. :1997-01-09	Min. : 3.70	Min. : 3.30	
##	Im Langen Loch	:52 1st Qu.:1997-04-17	1st Qu.: 14.70	1st Qu.: 15.20	
##	Klingelbergstrasse:51	Median :1997-07-14	Median : 22.30	Median : 23.90	
##	Luzernerring	:49 Mean :1997-07-09	Mean : 27.22	Mean : 28.13	
##	Sevogelstrasse	:52 3rd Qu.:1997-10-06	3rd Qu.: 32.70	3rd Qu.: 34.00	
##	Wasgenring	:59 Max. :1997-12-31	Max. :115.30	Max. :107.50	
##					
##	LV	PV	Wochenende	Gradienthoch	Monat



```
## 260 :52 5000 :52 Mode :logical Mode :logical Length:315
## 370 :52 5200 :52 FALSE:218 FALSE:182 Class :character
## 500 :52 6200 :52 TRUE :97 TRUE :133 Mode :character
## 1130:51 16300:59
## 3410:59 27100:51
## 4920:49 29500:49
##
## trend L1.PM10
## Min. : 0.0 Min. : 3.70
## 1st Qu.: 98.5 1st Qu.: 14.22
## Median :186.0 Median : 22.10
## Mean :181.5 Mean : 27.21
## 3rd Qu.:270.5 3rd Qu.: 33.10
## Max. :356.0 Max. :115.30
## NA's :25
## tibble [315 × 11] (S3: tbl_df/tbl/data.frame)
## $ MESSort : Factor w/ 6 levels "Claraplatz","Im Langen Loch",...: 6 6 6 6 6 6 6 6 6 ...
## $ Datum : Date[1:315], format: "1997-01-09" "1997-01-10" ...
## $ PM10 : num [1:315] 94.9 86.3 46.6 41.4 58.6 69.7 70.7 75.3 85.6 72.3 ...
## $ NABEL : num [1:315] 95.8 98.5 55.6 45.4 58.5 74.9 78.5 81.8 90.7 80.4 ...
## $ LV : Ord.factor w/ 6 levels "260"<"370"<"500"<...: 5 5 5 5 5 5 5 5 5 ...
## $ PV : Ord.factor w/ 6 levels "5000"<"5200"<...: 4 4 4 4 4 4 4 4 4 ...
## $ Wochenende : logi [1:315] FALSE FALSE TRUE TRUE FALSE FALSE ...
## $ Gradienthoch: logi [1:315] TRUE TRUE TRUE TRUE TRUE TRUE ...
## $ Monat : chr [1:315] "Jänner" "Jänner" "Jänner" "Jänner" ...
## $ trend : num [1:315] 0 1 2 3 4 5 6 7 8 9 ...
## $ L1.PM10 : num [1:315] 15.7 37.2 52.8 55.4 57.8 72.4 68.7 75.8 72.6 24.4 ...
```



```
## - attr(*, "na.action")= 'omit' Named int [1:17] 1 109 110 111 302 321 322 323 324 325
...
## .. attr(*, "names")= chr [1:17] "1" "109" "110" "111" ...

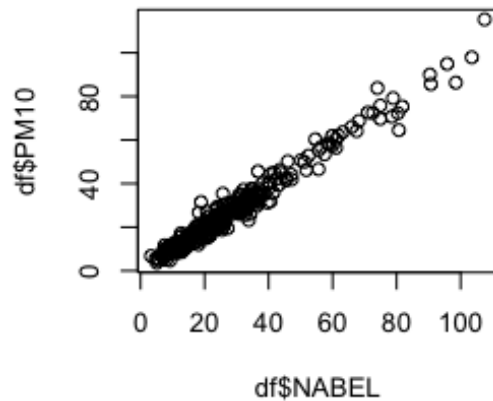
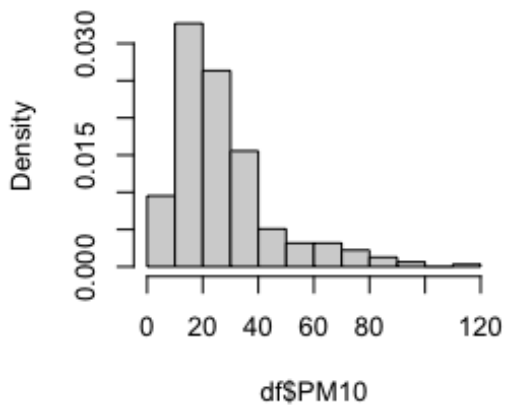
## # A tibble: 6 × 11
## Messort Datum PM10 NABEL LV PV Woche...1 Gradi...2 Monat trend L1.PM10
## <fct> <date> <dbl> <dbl> <ord> <ord> <lgl> <lgl> <chr> <dbl> <dbl>
## 1 Wasgen... 1997-01-09 94.9 95.8 3410 16300 FALSE TRUE Jänn... 0 15.7
## 2 Wasgen... 1997-01-10 86.3 98.5 3410 16300 FALSE TRUE Jänn... 1 37.2
## 3 Wasgen... 1997-01-11 46.6 55.6 3410 16300 TRUE TRUE Jänn... 2 52.8
## 4 Wasgen... 1997-01-12 41.4 45.4 3410 16300 TRUE TRUE Jänn... 3 55.4
## 5 Wasgen... 1997-01-13 58.6 58.5 3410 16300 FALSE TRUE Jänn... 4 57.8
## 6 Wasgen... 1997-01-14 69.7 74.9 3410 16300 FALSE TRUE Jänn... 5 72.4
## # ... with abbreviated variable names 1Wochenende, 2Gradienthoch
```

Ein Blick auf die Verteilung der Variable PM10 in der Stichprobe (oberes Histogramm in nachstehender Grafik) zeigt die Rechtsschiefe der empirischen PM10-Verteilung. Zwecks Normalverteilungsannahme für die Inferenz in einem linearen Regressionsmodell empfiehlt es sich daher die Variable PM10 logarithmisch zu transformieren. Die Verteilung ist dadurch symmetrischer, wie das untere Histogramm zeigt. Unabhängig davon, ob PM10 und die Referenzmessung NABEL beide logarithmisch transformiert wurden oder nicht, ist die Annahme eines linearen Zusammenhangs, wie durch Gleichung (6-1) auf S. 51 unterstellt, durch die Daten gedeckt (siehe die beiden Abbildungen im rechten Panel in

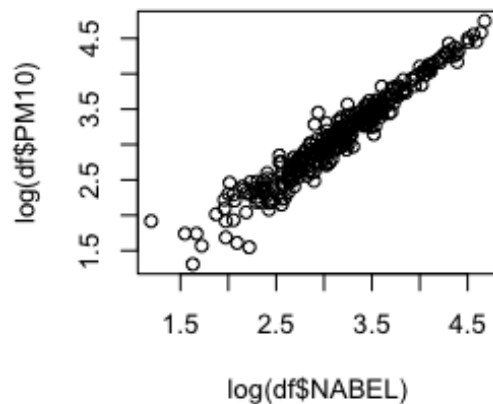
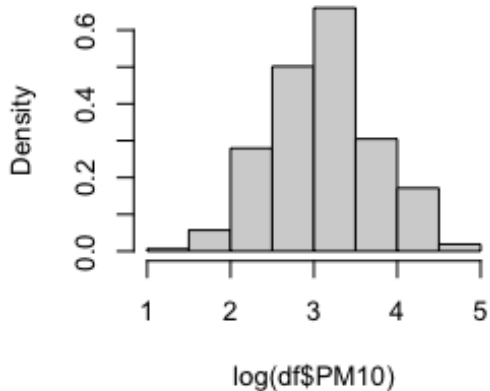


nachfolgender Grafik).

**Histogram of df\$PM10**



**Histogram of log(df\$PM10)**



## B) Qualitative Replikation der Ergebnisse

Der empirischen Haupt-Spezifikation der Dissertation in Gleichung (6-1) auf S. 51 folgend, wird für Datensatz 1 das nachstehende Regressionsmodell berechnet:

$$C_{it} = a + bC_{NABEL,t} + c * t + Monat_t + WE_t + Verkehr_i,$$

wobei Index  $i$  eine der sechs mobilen Messtationen bezeichnet und Index  $t$  den Messzeitpunkt. Die Variable  $t$  bezeichnet einen linearen Trend,  $Monat_t$  sind monats-spezifische fixe Effekte (alternativ wurde hier eine binäre Variable für einen hohen vs. niedrigen



Temperaturgradienten mit den Monatsangaben laut Abbildung 6-5 auf S. 61 spezifiziert),  $WE_t$  ist eine Indikatorvariable dafür, ob es sich beim Messzeitpunkt um ein Wochenende handelt oder nicht, und die Variable Verkehr basiert auf den für Personen- und Lastverkehr gegebenen Durchschnittswerten für die jeweilige Messstation (siehe Tabelle 6-1 auf S. 49).

Nachstehend wurden auf Basis obiger auf Gleichung (6-1) beruhender Spezifikation lineare Regressionen für den aus 315 Beobachtungen bestehenden Datensatz 1 geschätzt.

Im Ergebnis zeigt sich ein robust stark positiver Zusammenhang zwischen der PM10-Messung an der Referenzstation NABEL und der jeweiligen mobilen Station. An dieser empirischen Schlussfolgerung ändert sich auch dann nichts, wenn man einen autoregressiven Zusammenhang für PM10 über die Zeit in die Spezifikation inkludiert. Wird für Unterschiede in meteorologischen Bedingungen nicht-parametrisch für Monate und durch einen linearen Trend kontrolliert, so steigt PM10 robust sowohl mit steigendem Personen- als auch Lastverkehr.

Die Inferenz unten stehender Berechnungen beruht auf Cluster-robusten Standardfehlern nach Messort (vgl. Residuenanalyse, welche einen Hinweis auf Heteroskedastizität und Autokorrelation liefert); adjustiertes  $R^2$  der Schätzungen jeweils unterhalb angegeben.

*Abhängige Variable in Levels mit unterschiedlichen Maßen für die Verkehrsbelastung*

Oben:  $PM10 \sim NABEL + trend + Wochenende + Monat + PV$ ; unten:  $PM10 \sim NABEL + trend + Wochenende + Monat + LV$

```
##
## t test of coefficients:
##
##          Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.733109  3.439765  1.3760 0.1698521
## NABEL        0.975071  0.025042 38.9370 < 2.2e-16 ***
## trend       -0.025903  0.026303 -0.9848 0.3255077
## WochenendeTRUE -1.378152  0.411395 -3.3499 0.0009119 ***
```



```
## MonatAugust -1.542951 2.978881 -0.5180 0.6048670
## MonatDezember 2.329013 6.278831 0.3709 0.7109516
## MonatFebruar -3.765670 1.632419 -2.3068 0.0217491 *
## MonatJänner -9.193202 2.681062 -3.4289 0.0006911 ***
## MonatJuli -2.048674 2.272082 -0.9017 0.3679566
## MonatJuni -2.987030 1.886027 -1.5838 0.1143035
## MonatMai -1.912849 0.710726 -2.6914 0.0075159 **
## MonatMärz -2.435221 1.503595 -1.6196 0.1063726
## MonatNovember 1.726608 4.952033 0.3487 0.7275850
## MonatOktober 0.999674 4.157128 0.2405 0.8101290
## MonatSeptember -1.195516 4.152729 -0.2879 0.7736328
## PV 0.505087 0.123637 4.0852 5.661e-05 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## [1] 0.97806

##
## t test of coefficients:
##
## Estimate Std. Error t value Pr(>|t|)
## (Intercept) 3.660077 3.516469 1.0408 0.2987913
## NABEL 0.974670 0.025236 38.6215 < 2.2e-16 ***
## trend -0.019515 0.026318 -0.7415 0.4589784
## WochenendeTRUE -1.387044 0.411743 -3.3687 0.0008542 ***
## MonatAugust -1.824681 2.913528 -0.6263 0.5316099
## MonatDezember 1.341837 6.292388 0.2132 0.8312792
## MonatFebruar -2.878984 1.680679 -1.7130 0.0877513 .
## MonatJänner -8.403508 2.832762 -2.9665 0.0032547 **
```





```
## MonatJuli -2.234937 2.326647 -0.9606 0.3375382
## MonatJuni -2.776968 1.813100 -1.5316 0.1266751
## MonatMai -1.604783 0.646598 -2.4819 0.0136180 *
## MonatMärz -1.512189 1.484934 -1.0184 0.3093327
## MonatNovember 0.872884 4.910582 0.1778 0.8590353
## MonatOktober 0.108755 4.070437 0.0267 0.9787022
## MonatSeptember -1.426746 3.966969 -0.3597 0.7193579
## LV 0.495412 0.137135 3.6126 0.0003555 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## [1] 0.9779158
```



*Abhängige Variable und Referenzmessung NABEL in Logs mit unterschiedlichen Maßen für die Verkehrsbelastung*

Oben:  $\log(\text{PM}_{10}) \sim \log(\text{NABEL}) + \text{trend} + \text{Wochenende} + \text{Monat} + \text{PV}$ ; unten:  $\log(\text{PM}_{10}) \sim \log(\text{NABEL}) + \text{trend} + \text{Wochenende} + \text{Monat} + \text{LV}$

```
##
## t test of coefficients:
##
##          Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.3781988  0.1446106  2.6153 0.0093673 **
## log(NABEL)   0.9689603  0.0212975 45.4963 < 2.2e-16 ***
## trend       -0.0019986  0.0010447 -1.9130 0.0567007 .
## WochenendeTRUE -0.0701114  0.0142980 -4.9036 1.547e-06 ***
## MonatAugust  0.0235857  0.1129421  0.2088 0.8347232
## MonatDezember 0.2684219  0.2449721  1.0957 0.2740811
## MonatFebruar -0.2237088  0.0640268 -3.4940 0.0005478 ***
## MonatJänner  -0.3610690  0.1010976 -3.5715 0.0004135 ***
## MonatJuli    -0.0243163  0.0981172 -0.2478 0.8044369
## MonatJuni    -0.1250073  0.0896977 -1.3937 0.1644582
## MonatMai     -0.0287193  0.0502100 -0.5720 0.5677631
## MonatMärz    -0.1460931  0.0611774 -2.3880 0.0175599 *
## MonatNovember 0.2380776  0.2161288  1.1016 0.2715416
## MonatOktober 0.1707353  0.1634080  1.0448 0.2969407
## MonatSeptember 0.0884536  0.1613350  0.5483 0.5839225
## PV          0.0199432  0.0030478  6.5435 2.606e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## [1] 0.9559571
```



```
##
## t test of coefficients:
##
##      Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.3364614  0.1438529  2.3389 0.019998 *
## log(NABEL)   0.9692754  0.0218241 44.4130 < 2.2e-16 ***
## trend        -0.0017470  0.0010413 -1.6777 0.094460 .
## WochenendeTRUE -0.0703724  0.0143251 -4.9125 1.483e-06 ***
## MonatAugust  0.0118557  0.1099937  0.1078 0.914238
## MonatDezember 0.2289166  0.2430082  0.9420 0.346947
## MonatFebruar -0.1893250  0.0656209 -2.8851 0.004197 **
## MonatJänner  -0.3313644  0.1075828 -3.0801 0.002262 **
## MonatJuli    -0.0318829  0.0999783 -0.3189 0.750027
## MonatJuni    -0.1173794  0.0867624 -1.3529 0.177115
## MonatMai     -0.0169176  0.0487626 -0.3469 0.728882
## MonatMärz    -0.1105504  0.0600325 -1.8415 0.066537 .
## MonatNovember 0.2036000  0.2126445  0.9575 0.339105
## MonatOktober 0.1351588  0.1599140  0.8452 0.398677
## MonatSeptember 0.0787386  0.1548209  0.5086 0.611423
## LV          0.0191711  0.0044264  4.3311 2.028e-05 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

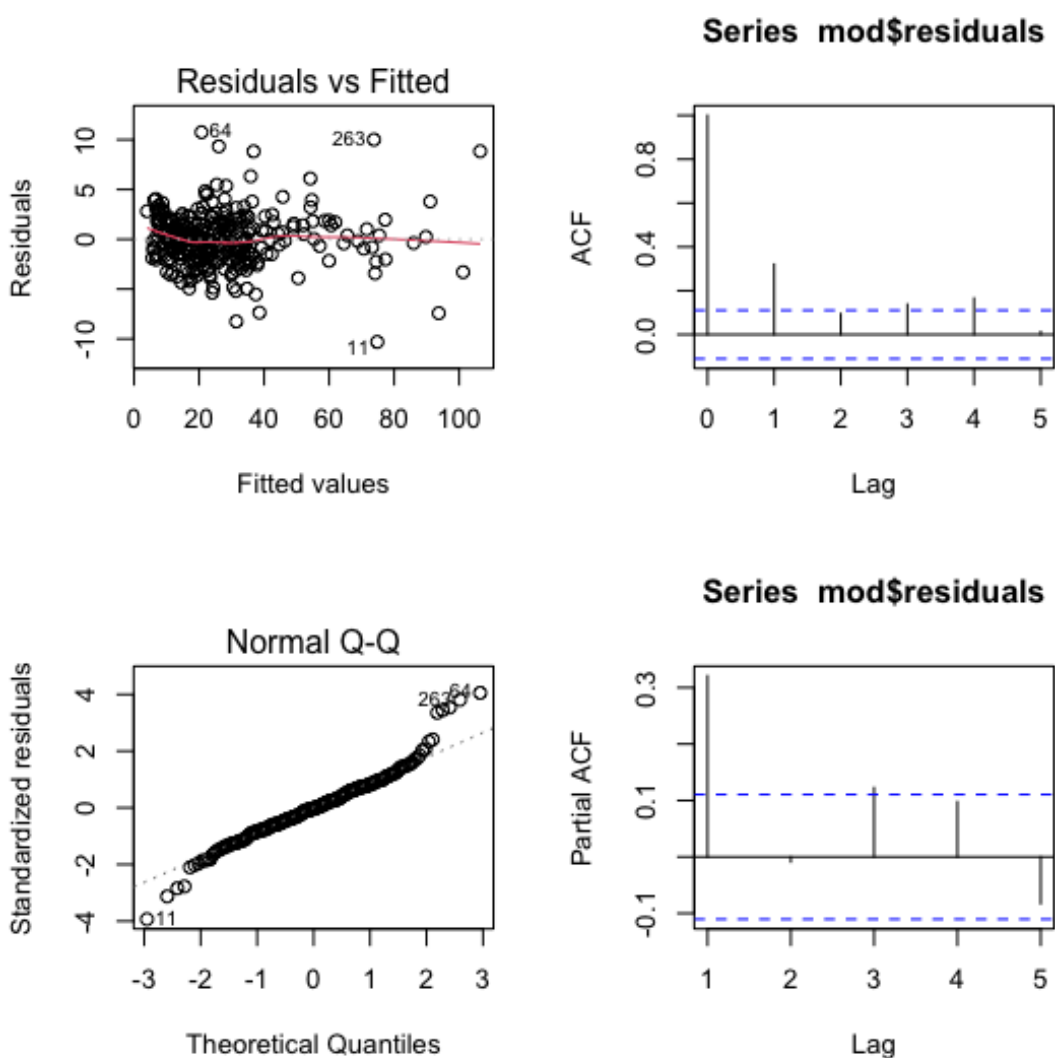
## [1] 0.9556622
```

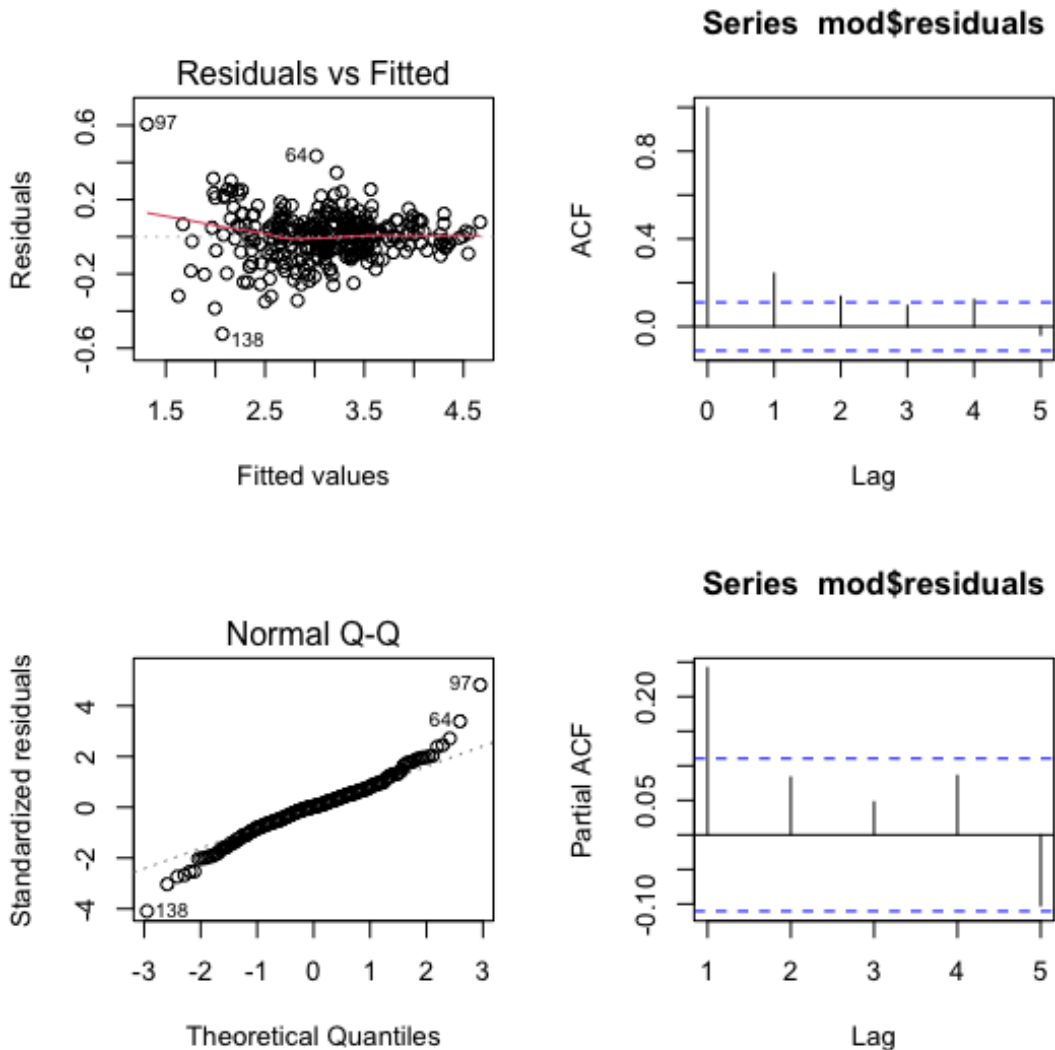


*Residuenanalyse zur Rechtfertigung der Inferenz (exemplarisch für Verkehrsaufkommen mittels LV gemessen)*

Oben:  $PM10 \sim NABEL + trend + Wochenende + Monat + LV$  (abhängige Variable in Levels);

unten:  $\log(PM10) \sim \log(NABEL) + trend + Wochenende + Monat + LV$  (abhängige Variable und Referenzwert in Logs)





*Robustheit der NABEL-Spezifikation (Gleichung 6-1) unter Berücksichtigung eines AR(1) für PM10 (exemplarische Spezifikationen)*

Oben:  $PM10 \sim NABEL + L1.PM10 + trend + Wochenende + Monat + PV$ ; unten:  $\log(PM10) \sim \log(NABEL) + \log(L1.PM10) + trend + Wochenende + Monat + PV$

```
##  
## t test of coefficients:  
##  
##           Estimate Std. Error t value Pr(>|t|)  
## (Intercept)  4.158254  3.237994  1.2842 0.2001583
```



```
## NABEL      0.964592  0.020777 46.4259 < 2.2e-16 ***
## L1.PM10    0.028723  0.017647  1.6277 0.1047524
## trend     -0.025799  0.023479 -1.0988 0.2728157
## WochenendeTRUE -1.389398  0.362417 -3.8337 0.0001568 ***
## MonatAugust -1.909580  2.567448 -0.7438 0.4576578
## MonatDezember 2.286426  5.639732  0.4054 0.6854908
## MonatFebruar -3.700507  1.395345 -2.6520 0.0084688 **
## MonatJänner  -9.042346  2.687187 -3.3650 0.0008753 ***
## MonatJuli    -2.254958  1.951205 -1.1557 0.2488247
## MonatJuni    -3.100086  1.415196 -2.1906 0.0293285 *
## MonatMai     -1.925166  0.555287 -3.4670 0.0006113 ***
## MonatMärz    -2.516785  1.234797 -2.0382 0.0424919 *
## MonatNovember 1.691247  4.586826  0.3687 0.7126234
## MonatOktober  0.878887  3.489760  0.2518 0.8013484
## MonatSeptember -1.113289  3.823394 -0.2912 0.7711362
## PV          0.544272  0.138171  3.9391 0.0001039 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## [1] 0.9787572

##
## t test of coefficients:
##
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.31499107  0.14109673  2.2324 0.0263966 *
## log(NABEL)   0.96195770  0.01899886 50.6324 < 2.2e-16 ***
## log(L1.PM10) 0.03023623  0.00931302  3.2467 0.0013134 **
## trend       -0.00209745  0.00094511 -2.2193 0.0272905 *
```



```
## WochenendeTRUE -0.06867893 0.01471514 -4.6672 4.790e-06 ***
## MonatAugust 0.02674453 0.10043741 0.2663 0.7902241
## MonatDezember 0.28348500 0.22156722 1.2795 0.2018236
## MonatFebruar -0.23299508 0.05319586 -4.3799 1.694e-05 ***
## MonatJänner -0.36775933 0.10297087 -3.5715 0.0004194 ***
## MonatJuli -0.02106429 0.08421991 -0.2501 0.8026899
## MonatJuni -0.12958350 0.07167478 -1.8079 0.0717168 .
## MonatMai -0.02443122 0.04522186 -0.5403 0.5894633
## MonatMärz -0.15476722 0.05234369 -2.9568 0.0033809 **
## MonatNovember 0.26214054 0.20107084 1.3037 0.1934266
## MonatOktober 0.18158271 0.13953286 1.3014 0.1942319
## MonatSeptember 0.11116924 0.14911134 0.7455 0.4565836
## PV 0.02105223 0.00320447 6.5696 2.543e-10 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## [1] 0.9574859
```

#### **4. Airborne Particle Number Profiles, Particle Mass Distributions and Particle Bound PAH Concentrations Within the City Environment of Basel: an Assessment as Part of the Briska Project (Kapitel 7)**

##### **4.1 Zusammenfassung und Würdigung**

In Kapitel 7 wird die räumliche Variabilität der Partikelanzahl und der partikelgebundenen polyzyklischen aromatischen Kohlenwasserstoffe (pPAH) durch einen qualitativen Vergleich von Punktmessungen an zwei Verkehrsstandorten und einem typischen Wohnstandort analysiert (siehe S. 39).

Obwohl sich die Forschungsfrage auf die Auswirkung unterschiedlicher Verkehrsintensitäten bezieht, konnten laut Autoren keine Verkehrsdaten zeitgleich mit den Feinstaub-





Daten erhoben werden (siehe S. 76). Es werden dann zwar Daten von einer anderen Messung herangezogen, jedoch sind diese für den am stärksten belasteten Messort nicht verfügbar (vgl. S. 76-77).

Bezüglich der statistischen Bearbeitung der Fragestellung greifen die Autoren ausschließlich auf eine deskriptive Analyse zurück, wobei die Berechnung eines statistischen Modells aus Sicht der Daten möglich gewesen wäre und wohl auch zu aussagekräftigeren Ergebnissen geführt hätte.

In der Interpretation der Ergebnisse wird zwar eingeräumt, dass unterschiedliche Feinstaubkonzentrationen mit menschlichem Verhalten in Verbindung stehen könnten, jedoch wird ebenso auf meteorologische Bedingungen verwiesen, die ebenso dafür verantwortlich sein könnten, jedoch wurde für diese nicht weiter kontrolliert.

Eine detaillierte Überprüfung der empirischen Ergebnisse ist nicht möglich, da die entsprechenden Daten nicht vorhanden sind.

#### **4.2 Übersicht**

1. Tabelle 7-1, S. 69; Inhalt: Deskriptive Zusammenfassung des Datensatzes; Anmerkungen: Keine vollständigen Originaldaten vorhanden;
2. Tabelle 7-2, S. 71; Inhalt: Erklärungen zu Messgerät
3. Abbildung 7-1, S. 72; Inhalt: Partikelnummer – und -größe; Anmerkungen: Luzernring Konfidenzbänder, Standardabweichung?
4. Abbildung 7-2, S. 73; Inhalt: Partikelnummer – und -größe Wasgenring; Anmerkungen: Konfidenzbänder, Standardabweichung?
5. Abbildung 7-3, S. 74; Inhalt: Partikelnummer – und -größe Sevogelstrasse; Anmerkungen: Konfidenzbänder, Standardabweichung?

In den Tagesprofilen zeigt sich eine stark positive Autokorrelation. Laut Beschreibung im



Text scheint es sich hierbei um berechnete Autokorrelationen erster Ordnung zu handeln (vgl. S. 74). Wie auch die Grafiken 7-1 bis 7-3 nahelegen, scheint es sich hier aber um eine zeitliche Korrelation höhere Ordnung zu handeln. Es wäre diesbezüglich methodisch hilfreich bzw. erboten, hätten die Autoren die (partielle) Autokorrelationsfunktion auch an höheren Lags analysiert.

6. Abbildung 7-4, S. 75; Inhalt: Feinstaubbelastung Luzernerring; Anmerkungen: Konfidenzbänder bzw. Standardfehler? (vgl. Im Gegensatz dazu Abb. 7-6)
7. Tabelle 7-3, S. 75; Inhalt: Durchschnittliche Schadstoffkonzentration; Anmerkungen: Warum kein statistischer Test bezüglich Unterschieden in den Konzentrationen der Schadstoffe an den drei Standorten?

Tabelle 7-3 enthält Informationen über tagesdurchschnittlich gemessene Schadstoffkonzentrationen für die drei unterschiedlichen Messorte. Es ist nicht klar, ob bei der Berechnung der in Klammern angegebenen Standardabweichung die zeitliche Korrelation der Daten berücksichtigt wurde. Grundsätzlich wäre es auf Basis dieser Daten möglich gewesen, statistisch für Unterschiede in den dort genannten Partikelkonzentrationen zu testen.

8. Abbildung 7-5, S. 76; Inhalt: Feinstaubbelastung Sevoglstrasse
9. Abbildung 7-6, S. 77; Inhalt: Feinstaubbelastung Wasgenring

In den Abbildungen 7-5 und 7-6 sind das Verkehrsaufkommen über den Tag und die Konzentrationen ultrafeiner Partikel zusammen dargestellt. Aus den beiden Grafiken zeigt sich, dass das Verkehrsaufkommen vorher steigt und dann erst die Partikelkonzentration. Für Personenverkehr wird dies als nur schwache Korrelation interpretiert (siehe S. 77 und 78), jedoch – wie auch die grafische Inspektion nahelegt – könnte dies daran liegen, dass die



positive Abhängigkeit durch höhere Lag-Strukturen getrieben ist, d.h. der Effekt des Personenverkehrs auf die Feinstaubbelastung könnte schlichtweg zeitverzögert registrieren, sodass das Beachten ausschließlich kontemporärer Korrelationen zu einer Unterschätzung des Zusammenhangs führen könnte. Es ist jedenfalls aus statistischer Sicht nicht nachvollziehbar, warum die Hypothese einer zeitverzögerten Registration des Verkehrsaufkommens auf die Partikelkonzentration nicht untersucht wurde.

10. Abbildung 7-7, S. 79; Inhalt: Mittlere Partikelmassenverteilungen; Anmerkungen: Ohne Originaldaten nicht überprüfbar.

11. Tabelle 7-4, S. 80; Inhalt: Mittlere Partikelmassenverteilungen; Anmerkungen: Ohne Originaldaten nicht überprüfbar.

12. Tabelle 7-5, S. 80; Inhalt: Mittlere Partikelmassenverteilungen im Akkumulationsmodus; Anmerkungen: Ohne Originaldaten nicht überprüfbar; Konfidenzbänder, Standardabweichung bzw. Interquartilsabstand wie bei Tabelle 7-4?

### 4.3 Berechnungen

Informationen zu den Tagesprofilen sind im übermittelten Datensatz nicht vorhanden. Auch bei den Daten aus dem oben unter 2.2 bezeichnetem Datensatz 2 scheint es sich nicht um die gleichen Daten wie in Kapitel 7 zu handeln.

#### A) Deskriptive Statistik Datensatz 2

```
## [1] "Messort" "Datum" "SO2" "NO" "NO2" "NOx 1)"
## [7] "O3" "CO" "Russ" "PAH" "PM 2.5" "PM 10"
## [13] "TSP" "TSP/NAB" "PM10/NAB" "PM2.5/NAB" "PM10/StJ" "NABEL"

## Messort Datum PM10 PAH
## Length:73 Min. :1997-08-06 Min. : 3.70 Min. : 8.00
## Class :character 1st Qu.:1997-08-25 1st Qu.: 17.50 1st Qu.: 22.60
## Mode :character Median :1997-10-11 Median : 27.70 Median : 35.80
```



```
##          Mean :1997-10-13  Mean : 30.36  Mean :45.82
##          3rd Qu.:1997-11-28  3rd Qu.: 34.50  3rd Qu.: 57.50
##          Max.  :1997-12-31  Max.  :115.30  Max.  :136.70
##  NABEL          Verkehr Wochenende  Gradienthoch
## Min.   : 5.10  Sevogelstrasse:26  Mode :logical  Mode :logical
## 1st Qu.:18.90  Wasgenring :21  FALSE:51     FALSE:42
## Median :27.10  Luzernerring :26  TRUE :22     TRUE :31
## Mean   : 31.17
## 3rd Qu.: 35.70
## Max.   :107.50
##  Monat          L1.PM10          L2.PM10          trend
## Length:73      total    :34167  total    :34167  Min.   :209.0
## Class :character  between_id :6858  between_id :6858  1st Qu.:228.0
## Mode  :character  between_time:34167  between_time:34167  Median :275.0
##
##                                     Mean :277.4
##                                     3rd Qu.:323.0
##                                     Max.  :356.0
##  ratioPM10
## Min.   :0.7068
## 1st Qu.:0.8829
## Median :0.9495
## Mean   :0.9514
## 3rd Qu.:1.0100
## Max.   :1.3774
## # A tibble: 6 × 13
##  Messort  Datum    PM10  PAH NABEL Verkehr Woche...1 Gradi...2 Monat L1.PM10
##  <chr>    <date>  <dbl> <dbl> <dbl> <ord>  <lgl>  <lgl>  <chr> <pseri>
```



```
## 1 Luzernerri... 1997-08-06 22.9 51.8 21.4 Luzern... FALSE FALSE Augu... 31.2
## 2 Luzernerri... 1997-08-07 31.2 89.4 28.3 Luzern... FALSE FALSE Augu... 25.2
## 3 Luzernerri... 1997-08-08 25.2 57.5 25.6 Luzern... FALSE FALSE Augu... 20.7
## 4 Luzernerri... 1997-08-09 20.7 34.4 21.8 Luzern... TRUE FALSE Augu... 19.7
## 5 Luzernerri... 1997-08-10 19.7 23.7 20.4 Luzern... TRUE FALSE Augu... 30.1
## 6 Luzernerri... 1997-08-11 30.1 83.9 27.6 Luzern... FALSE FALSE Augu... 33.7
## # ... with 3 more variables: L2.PM10 <pseries>, trend <dbl>, ratioPM10 <dbl>, and
## # abbreviated variable names 1Wochenende, 2Gradienthoch
```

Verglichen mit den Werten Tabelle 7-1, S. 69, stimmen die Mittelwerte für PAH zwar überein, nicht aber diejenigen für PM10 (die Werte in Tabelle 7-1 sind geringer):

```
summary(df$PM10[df$Messort=="Luzernerring"])
##   Min. 1st Qu.  Median    Mean 3rd Qu.   Max.
##  8.90  23.48  31.20  42.00  56.62 115.30

summary(df$PM10[df$Messort=="Wasgenring"])
##   Min. 1st Qu.  Median    Mean 3rd Qu.   Max.
##  3.70  10.80  22.10  19.58  27.70  34.90

summary(df$PM10[df$Messort=="Sevogelstrasse"])
##   Min. 1st Qu.  Median    Mean 3rd Qu.   Max.
##  8.90  16.15  27.70  27.42  34.38  64.10

summary(df$PAH[df$Messort=="Luzernerring"])
##   Min. 1st Qu.  Median    Mean 3rd Qu.   Max.
## 18.60  43.02  77.30  73.63  96.72 136.70

summary(df$PAH[df$Messort=="Wasgenring"])
```



```
## Min. 1st Qu. Median Mean 3rd Qu. Max.
```

```
## 8.00 20.40 35.80 32.83 42.60 59.00
```

```
summary(df$PAH[df$Messort=="Sevogelstrasse"])
```

```
## Min. 1st Qu. Median Mean 3rd Qu. Max.
```

```
## 9.00 20.27 22.90 28.51 35.12 66.40
```

## **B) Zeitreihen-Korrelation von PM10 (und TSP)**

Das vom Autor übermittelte File NABEL98.XLS enthält tägliche Messwerte der NABEL-Station. Auch wenn nicht mit den Tagesdaten aus Kapitel 7 unmittelbar vergleichbar, können diese Daten dazu verwendet werden um PM10 und TSP hinsichtlich ihrer Zeitreihenstruktur zu untersuchen. Es wird im Folgenden der in Abschnitt 2.2 als Datensatz 3 bezeichnete Datensatz analysiert. An verschiedenen Stellen der empirischen Überlegungen in der Dissertation ist die Zeitreihenstruktur insbesondere für PM10 bezüglich der Inferenz von Relevanz. (Sind die Residuen einer Regression autokorreliert, so führt die unterlassene Berücksichtigung dieser Tatsache im Rahmen der Inferenz zu verzerrten Standardfehlern bzw. Konfidenzintervallen.)

### *Deskriptive Eigenschaften Datensatz 3*

7 (8) der Werte für PM4 (TSP) entsprachen 9999 und wurden aus dem Datensatz entfernt

```
## TSP PM10 Datum
```

```
## Min. : 5.00 Min. : 2.40 Min. :1998-01-01
```

```
## 1st Qu.: 16.70 1st Qu.: 13.60 1st Qu.:1998-03-31
```

```
## Median : 25.70 Median : 20.00 Median :1998-07-05
```

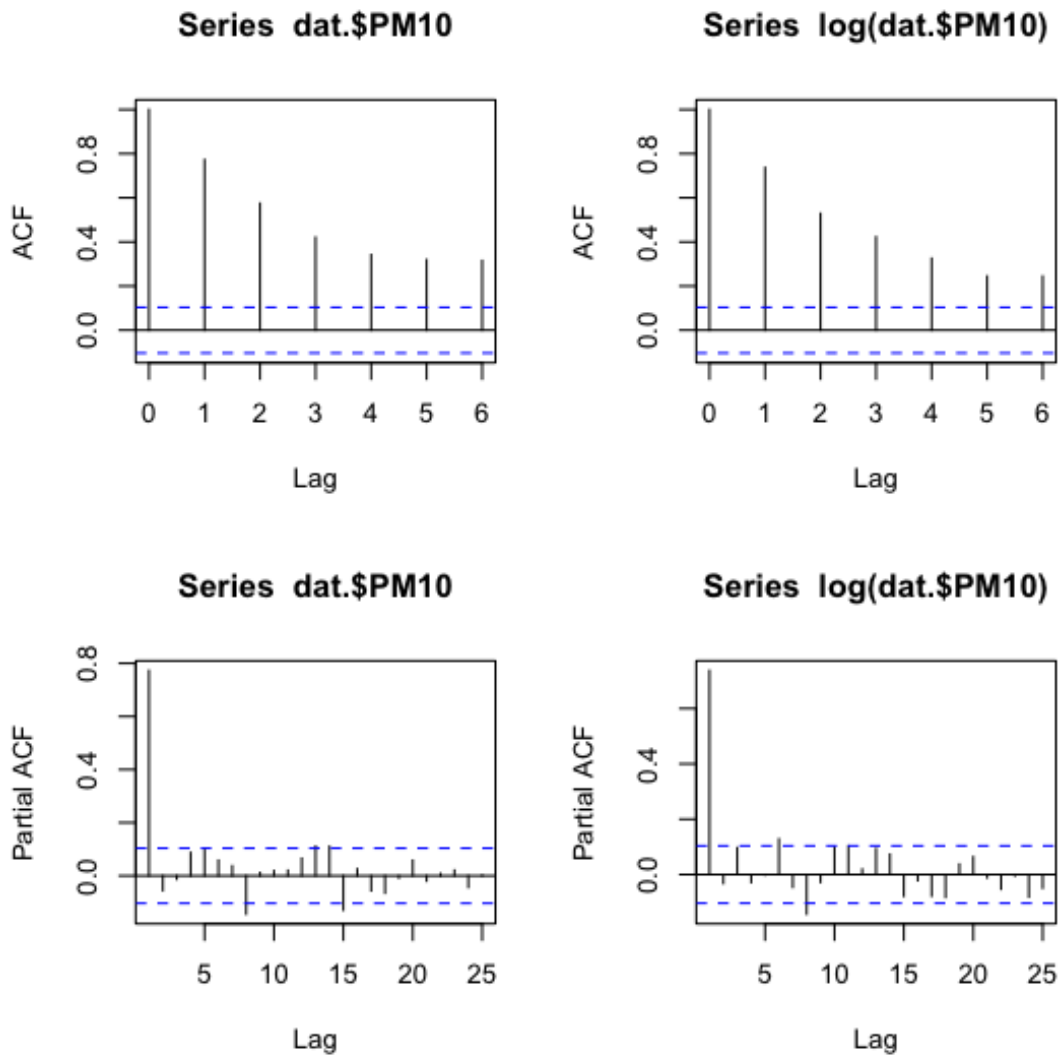
```
## Mean : 30.04 Mean : 24.01 Mean :1998-07-02
```

```
## 3rd Qu.: 39.30 3rd Qu.: 31.00 3rd Qu.:1998-10-02
```

```
## Max. :122.80 Max. :111.70 Max. :1998-12-30
```

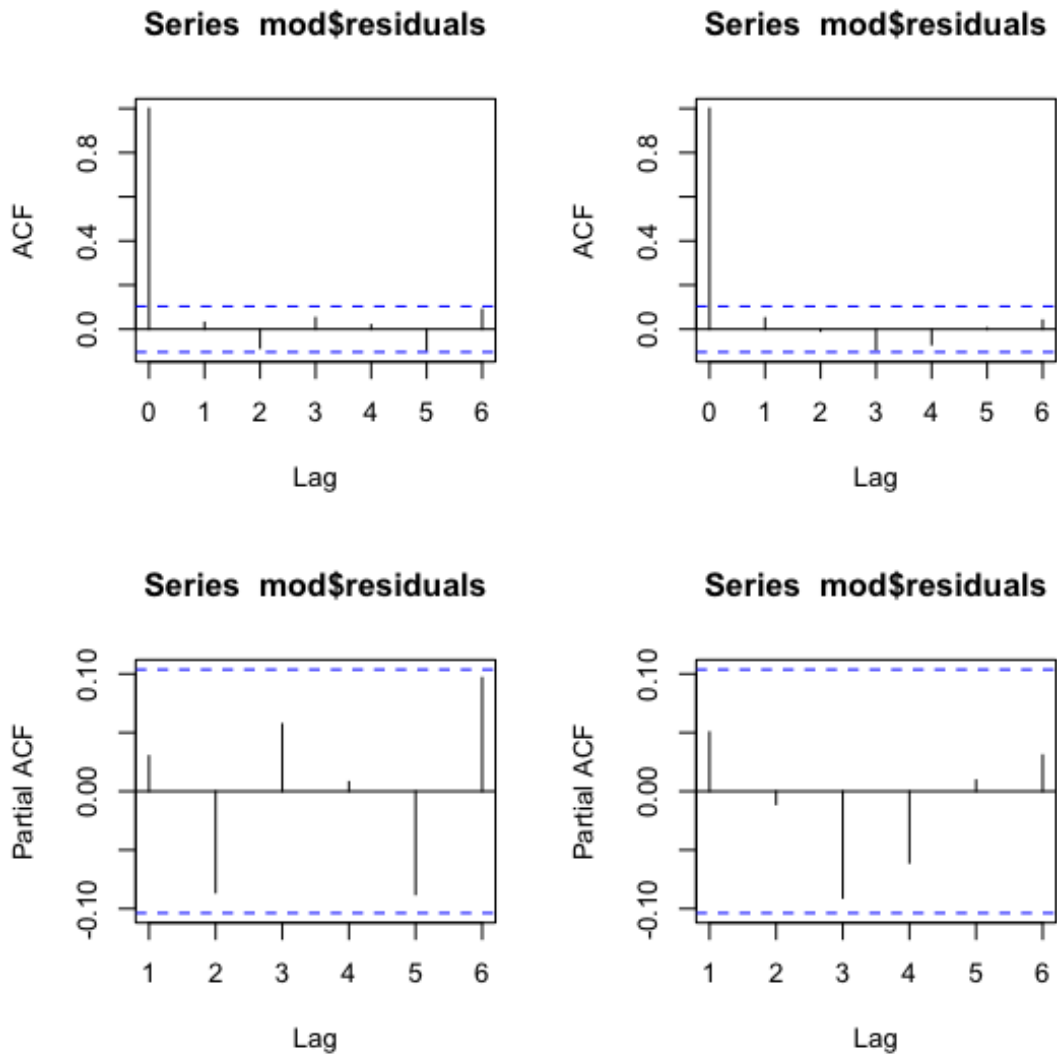


Eine grafische Inspektion der Autokorrelationsfunktion und der partiellen Korrelation zeigt für PM10 in Levels als auch in Logs einen starken Hinweis auf einen AR(1) Prozess:





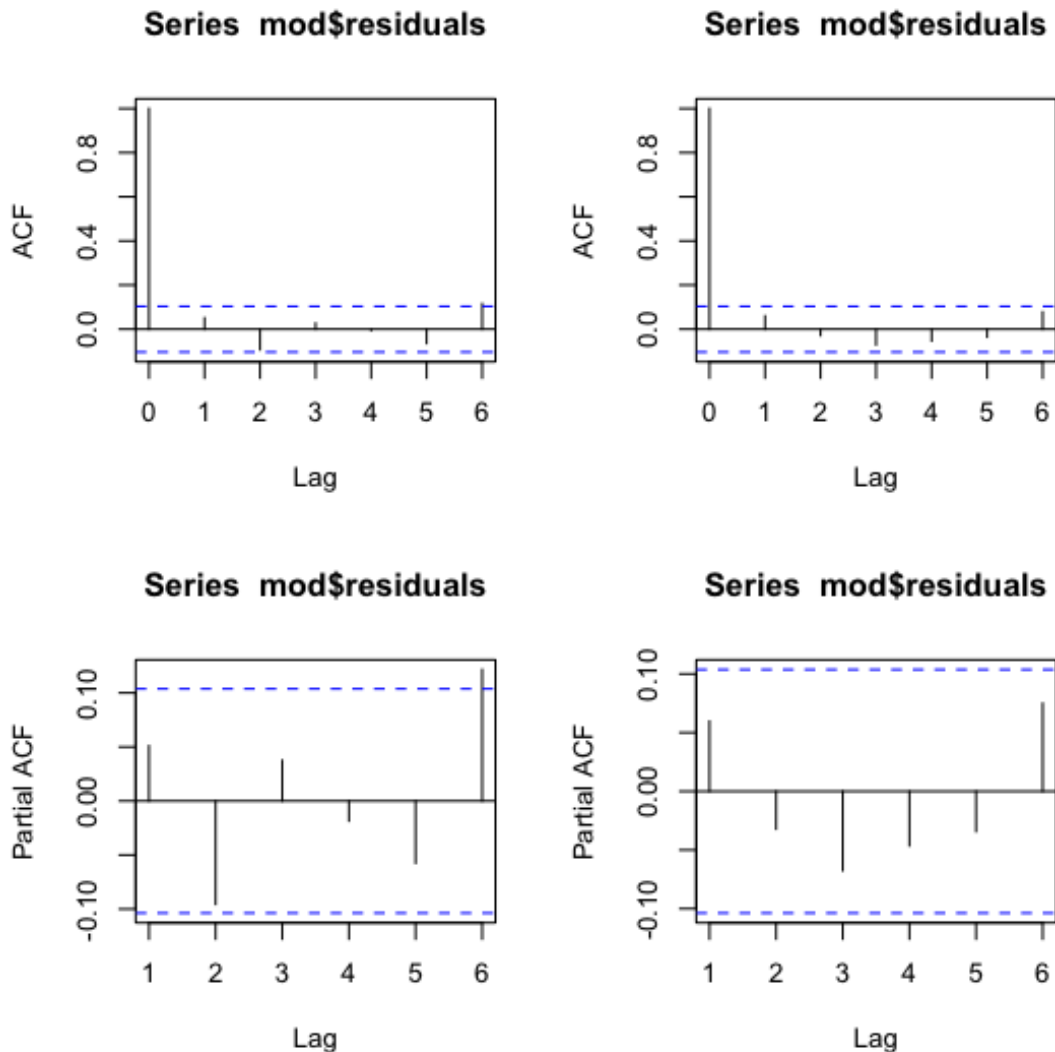
Die Residuen aus der Schätzung eines AR(1)-Modells sind zeitlich unkorreliert, wie unsere Abbildung links für das Modell in Logs und rechts für das Modell in Levels zeigt:







Auch TSP lässt sich mittels eines AR(1) in Logs und in Levels gut modellieren, wie die Residuen der entsprechenden AR(1)-Modelle (links Logs, rechts Levels) zeigen:



## 5. Temporal and Spatial Variation of the Chemical Composition of PM10 at Urban and Rural Sites in the Basel Area, Switzerland (Kapitel 8)

### 5.1 Zusammenfassung und Würdigung

In Kapitel 8 wurden von 41 Tagen Analysen der PM10-Filter desselben Tages, die gleichzeitig an der festen städtischen Messstelle "C" und an einer der temporären Messstellen gemessen wurden, durchgeführt. Dass hier keine umfassendere Datensammlung möglich war, wird durch finanzielle Beschränkungen erklärt (siehe S. 38).



Weiters seien diese Daten auch zu begrenzt gewesen, um die Unterschiede in der chemischen Zusammensetzung zwischen den sechs temporären Messstellen zu analysieren (ibid.). Eine hohe Korrelation zwischen einer zeitgleich gemessenen Konzentration an der fixen Messstation und einer der temporären Messstationen wurde dabei als Indikator für eine homogene kleinräumige Variabilität innerhalb der Stadt Basel interpretiert (siehe S. 38). Aus den Messungen 1998/99 wurde eine begrenzte Anzahl von Stoffen chemisch analysiert ( $\text{SO}_4$ ,  $\text{NO}_3$ ,  $\text{NH}_4$ ,  $\text{Cl}$ ,  $\text{Pb}$ ,  $\text{Zn}$ ,  $\text{Cd}$ ) und daraus monatliche Mittelwerte ermittelt, was einen direkten Vergleich zwischen den vier Standorten erlaubte (siehe S. 39).

Die Autoren folgern eine gleichmäßige Verteilung der  $\text{PM}_{10}$ -Messkonzentration in Basel sowie, dass Unterschiede innerhalb der städtischen Umgebung von Basel sowie zwischen städtischen und ländlichen Standorten primär verkehrsbedingt sind (vgl. S. 103).

Qualitativ kann die Schlussfolgerung, dass die  $\text{PM}_{10}$ -Messungen der lokalen Stationen sehr gut durch die Station NABEL erklärt werden können, aus einer Analyse von Datensatz 2 repliziert werden. Wie bereits für die Rekapitulation von Kapitel 6 der Dissertation, zeigt sich auch hier ein stark signifikant positiver Zusammenhang zwischen gemessenem  $\text{PM}_{10}$  und dem durchschnittlichen Verkehrsaufkommen, wenn man für meteorologische Bedingungen mittels Proxy-Variablen kontrolliert.

## **5.2 Übersicht**

1. Tabelle 8-1, S. 87; Inhalt: Tabelle 6-1
2. Tabelle 8-2, S. 90; Inhalt: Chemische Analyse
3. Tabelle 8-3, S. 92; Inhalt: Chemische Zusammensetzung der Partikel über Jahreszeiten; Anmerkungen: Standardabweichung, Konfidenzintervall, statistischer Test auf Unterschiede in den Mittelwerten?
4. Abbildung 8-1, S. 94; Inhalt:  $\text{PM}_{10}$  und  $\text{PM}_4$  Sommer vs. Winter; Anmerkungen: Standardabweichung, Konfidenzintervall, statistischer Test auf Unterschiede zwischen Sommer und Winter?



5. Abbildung 8-2, S. 95; Inhalt: Verhältnis PM10 Arbeitstage vs. Wochenenden
6. Abbildung 8-3, S. 97; Inhalt: Pearson Korrelation für temporäre vs. fixe Messstation C; Anmerkungen: Mit welcher temporären Messstation wurde hier verglichen? Konfidenzband?
7. Tabelle 8-4, S. 99; Inhalt: Jährliche Durchschnittskonzentration für unterschiedliche Stationen; Anmerkungen: Statistischer Test auf Unterschiede? Standardfehler? Konfidenzintervall?
8. Abbildung 8-4, S. 101; Inhalt: Monatliche Durchschnittswerte Stadt/Land; Anmerkungen: Konfidenzbänder? Inferenz?
9. Abbildung 8-5, S. 102; Inhalt: Luftbelastung durch Lagerfeuer; Anmerkungen: Konfidenzbänder? Inferenz?

### **5.3 Berechnungen**

Ausgehend von Datensatz 2 können wiederum die gleichen empirischen Überlegungen wie zu Kapitel 6 angestellt werden. Um für meteorologische Bedingungen zu kontrollieren, wird wieder ein Set von Proxy-Variablen herangezogen (siehe Abschnitt 2.2). Um für das Verkehrsaufkommen zu kontrollieren, wird eine ordinale Variable ("Verkehr" unten) für die Verkehrsbelastung der drei Messorte laut Tabelle 8-1 auf S. 94 konstruiert. Weitere Variablen wurden in Abschnitt 2.2 beschrieben. Die Inferenz beruht auf Cluster-robusten Standardfehlern auf Ebene der Messstationen.

Die nachstehenden Regressionsmodelle legen im Ergebnis wiederum nahe (vgl. Abschnitt 3.3), dass die PM10-Messungen der lokalen Stationen sehr gut durch die Station NABEL erklärt werden. Weiters zeigt sich ein stark signifikant positiver Zusammenhang zwischen gemessenem PM10 und steigendem Verkehrsaufkommen, wenn man für meteorologische Bedingungen mittels der in Abschnitt 2.2 beschriebenen Proxy-Variablen kontrolliert.



Modell mit PM10 lokale Messstation und NABEL in Logs (unterschiedliche Variablen als Proxies für einen hohen versus niedrigen Temperaturgradienten)

Oben:  $\log(\text{PM10}) \sim \log(\text{NABEL}) + \text{trend} + \text{Wochenende} + \text{Gradienthoch} + \text{Verkehr}$ ; unten:  
 $\log(\text{PM10}) \sim \log(\text{NABEL}) + \text{trend} + \text{Wochenende} + \text{Monat} + \text{Verkehr}$

```
##
## t test of coefficients:
##
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.86292095  0.21140423 -4.0819 0.0001230 ***
## log(NABEL)   1.05259899  0.02784468 37.8025 < 2.2e-16 ***
## trend        0.00273737  0.00078218  3.4997 0.0008397 ***
## WochenendeTRUE -0.07436706  0.02881673 -2.5807 0.0120911 *
## GradienthochTRUE -0.25581857  0.07360780 -3.4754 0.0009065 ***
## Verkehr.L     0.08255459  0.02137832  3.8616 0.0002591 ***
## Verkehr.Q     0.10613769  0.03590211  2.9563 0.0043149 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## [1] 0.9795287

##
## t test of coefficients:
##
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.39592089  0.31920319 -1.2403 0.2194488
## log(NABEL)   1.03725853  0.03005010 34.5176 < 2.2e-16 ***
## trend        0.00078085  0.00135837  0.5748 0.5674460
## WochenendeTRUE -0.08224034  0.02699724 -3.0463 0.0033819 **
## MonatDezember -0.02812559  0.14909596 -0.1886 0.8509807
```



```
## MonatNovember -0.04364483 0.12110140 -0.3604 0.7197543
## MonatOktober 0.15859185 0.07192289 2.2050 0.0311105 *
## MonatSeptember -0.08949834 0.03062803 -2.9221 0.0048223 **
## Verkehr.L 0.06157944 0.02346827 2.6239 0.0108911 *
## Verkehr.Q 0.12358546 0.03343359 3.6964 0.0004602 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## [1] 0.980846
```

*Modell mit abhängiger Variable Verhältnis PM10 lokale Messstation und NABEL*

Oben:  $\log(\text{ratioPM10}) \sim \log(\text{L1.PM10}) + \text{trend} + \text{Wochenende} + \text{Monat} + \text{Verkehr}$ ; unten:  
 $\log(\text{ratioPM10}) \sim \log(\text{L1.PM10}) + \text{trend} + \text{Wochenende} + \text{Monat} + \text{Verkehr}$

```
##
## t test of coefficients:
##
##      Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.20223864 0.32161321 -0.6288 0.531701
## trend      0.00043361 0.00144967 0.2991 0.765824
## WochenendeTRUE -0.09335108 0.03086159 -3.0248 0.003579 **
## MonatDezember 0.01961074 0.16632723 0.1179 0.906513
## MonatNovember 0.00294598 0.12451949 0.0237 0.981198
## MonatOktober 0.18673484 0.09065499 2.0598 0.043486 *
## MonatSeptember -0.09649368 0.03115905 -3.0968 0.002903 **
## Verkehr.L 0.06469934 0.02180295 2.9675 0.004219 **
## Verkehr.Q 0.13982036 0.03115967 4.4872 3.071e-05 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



```
## [1] 0.4186147

##
## t test of coefficients:
##
##          Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.14344188  0.31691670 -0.4526  0.652513
## log(L1.PM10)  0.00248727  0.02342912  0.1062  0.915821
## trend        0.00012409  0.00139439  0.0890  0.929394
## WochenendeTRUE -0.09394820  0.03398274 -2.7646  0.007629 **
## MonatDezember  0.05182370  0.16145181  0.3210  0.749375
## MonatNovember  0.03892963  0.12168638  0.3199  0.750180
## MonatOktober   0.20731879  0.08568772  2.4195  0.018699 *
## Verkehr.L      0.06637630  0.02290831  2.8975  0.005300 **
## Verkehr.Q      0.14647158  0.03147445  4.6537  1.939e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## [1] 0.4402686
```

## **6. Spatial Variability of Different Fractions of Particulate Matter Within an Urban Environment and Between Urban and Rural Sites (Kapitel 9)**

### **6.1 Zusammenfassung und Würdigung**

Im Ergebnis ist festzuhalten, dass die entscheidende Annahme für die Berechnungen im Appendix zu Kapitel 9 (S. 114-115) nur für einen sehr begrenzten Bereich der Daten erfüllt ist (Taylorreihen-Approximation erster Ordnung der Logarithmus-Funktion). Eine auf Basis der aus Kapitel 9 und 10 verfügbaren Informationen über empirische Werte hier angestellte Simulation (siehe Abschnitt 6.3 B) unten) zeigt, dass nur rund 10-40% der zu



modellierenden Daten die erforderlichen Eigenschaften erfüllen, unter welchen die von den Autoren verwendete Approximation zu nur kleinen Näherungsfehlern führt.

Dies wäre auch eine plausible Erklärung dafür, warum – wie auch von den Autoren diskutiert – die mit ihrem Modell prognostizierten Werte in Kapitel 9 und 10 wesentlich größer sind als die aus der früheren Literatur zitierten Vergleichswerte. In Anbetracht der Berechnungen unten ist es offensichtlich, dass der vorgeschlagene epidemiologische Ansatz zu einer erheblich verzerrten Schätzung der erwarteten Krebsfälle führen muss, da die verwendete lineare Approximation die ER für die meisten Daten überschätzt.

## 6.2 Übersicht

1. Tabelle 9-1, S. 109; Inhalt: Risiko amerikanische Studien; Anmerkungen: In Fußnote wurde der Durchschnitt falsch berechnet (Adventists)

Datenquellen zu Tabelle 9-1: Zeile 1 Tabelle 9-1: „Six-cities“

(<https://www.nejm.org/doi/full/10.1056/nejm199312093292401>) und Zeile 2 Tabelle 9-1:

„Adventists“ ([https://www.atsjour-](https://www.atsjournal.org/doi/epdf/10.1164/ajrccm.159.2.9806020?role=tab)

[nals.org/doi/epdf/10.1164/ajrccm.159.2.9806020?role=tab](https://www.atsjournal.org/doi/epdf/10.1164/ajrccm.159.2.9806020?role=tab); Abbey et. al 1999)

In Abbey et. al. 1999 nahmen 4060 Frauen und 2278 Männer an der Studie teil (siehe Abbey et. al. 1999 Table 1). Als Anmerkung zu Tabelle 9-1 in Rösli wird gesagt, dass die RR

für Abbey et. al. 1999 durch Mittelwertbildung zwischen Frauen und Männern ermittelt

wurde (tatsächlich  $(3.36 + 1.33) * 0.5 = 2.35$  wie in Tabelle 9-1); korrekt wäre hier aber

ein gewichtetes Mittel  $1.33 * 4060 / (4060 + 2278) + 3.36 * 2278 / (4060 + 2278) =$

2.0596. Außerdem wurde das angegebene Konfidenzintervall ebenfalls durch Mittelung

der Werte für Männer und Frauen aus Tabelle 3 in Abbey et. al. 1999 ermittelt:

$(1.57 + 0.6) / 2 = 1.085$  ca. 1.09 und  $(7.19 + 2.96) / 2 = 5.075$  ca. 5.08 wie in Tabelle 9-1 in

Rösli angegeben, was nicht korrekt ist.

Zum Vergleich: In der obigen Tabelle von Abbey et. al. 1999 kann man die Varianzen der



Koeffizienten aus den berichteten 95% Gauß'schen Konfidenzintervallen berechnen:

$(3.36/1.96)^2 = 2.94$  für Männer und  $(1.33/1.96)^2 = 0.46$  für Frauen. Unter der Annahme, dass die Schätzungen zwischen Männern und Frauen unkorreliert sind, beträgt die Varianz ihrer gewichteten Summe  $0.46 * (4060/(4060 + 2278))^2 + 2.94 * (2278/(4060 + 2278))^2 = 0.568554$  und die entsprechende Standardabweichung etwa 0,75. Ein zweiseitiges 95%iges Gauß'sches Konfidenzintervall beträgt somit  $2.0596 \pm 1.96 * 0.75$  ca. [0.59;3.53].

2. Tabelle 9-2, S. 111; Inhalt: Geschätztes relatives Risiko und Excess Rate; Anmerkungen: Die als Konfidenzintervalle bezeichneten Größen sind keine Konfidenzintervalle

Die Werte aus Tabelle 9-2 konnten bis auf Rundungsdiskrepanzen repliziert werden (siehe für die Ergebnisse Abschnitt 6.3, A unten).

Anzumerken ist diesbezüglich, dass die als Konfidenzintervalle (95%-CI bezeichnet) bezeichneten Größen für ein Konfidenzintervall falsch berechnet wurden. Für ein Gauß'sches Konfidenzintervall z.B., hätte zuerst die Varianz des Punktschätzers berechnet werden müssen um mit dieser im Abschluss ein Konfidenzintervall auf Basis der Normalverteilung zu berechnen; aus Sicht der Autoren wäre es allerdings möglich gewesen, diese "Konfidenzintervalle" schlichtweg als Risikoschätzungen an den Konfidenzgrenzen zu bezeichnen.

**Herleitung von Formel (9-4) aus (9-3) wie im Appendix zu Kapitel 9 der Dissertation (S. 114-15) gezeigt**

Formel (9-8) auf S. 114 ist von der Form  $X = \log(1+A)$  und kann durch A approximiert werden (Taylorreihen-Approximation erster Ordnung, vgl. z.B. <https://en.wikipedia.org/wiki/Logarithm>). Im Allgemeinen sagt man, dass diese Näherung für Werte von A mit einem Absolutwert kleiner als 1 geeignet ist.

In Abschnitt 6.3, B) unten wurde im Rahmen einer Simulation für den Wertebereich der





Daten in der Dissertation im Detail untersucht, inwieweit die modellierten Daten die entscheidende Voraussetzung dafür erfüllen, dass die Approximation zu verlässlichen Ergebnissen führt (die nur einen vernachlässigbaren Annäherungsfehler aufweisen).

### 6.3 Berechnungen

#### A) Replikation Tabelle 9-2

```
# FUNKTIONEN ZUR BERECHNUNG
```

```
RR <- function(dC, RRtot) exp(10*log(RRtot)/dC) # Gleichung (9-1)
```

```
NRR <- function(RR,Itar, P, Cav) (RR-1)*Itar*P*Cav/10 # Gleichung (9-2)
```

```
ER <- function(I0, RR) I0*log(RR) # Gleichung (9-4)
```

```
NER <- function(ER, P, Cav) P*Cav*ER/10 # Gleichung (9-5)
```

```
# WERTE FÜR PARAMETER LAUT ARBEIT
```

```
Ci <- seq(from=10, to =35, length=500)
```

```
Cav <- 17.5
```

```
Iav <- 50/100000
```

```
I01 <- 106/100000
```

```
I02 <- 42/100000
```

```
# POINT ESTIMATES
```

```
RR.1 <- RR(28.3,1.37)
```



```
RR.2 <- RR(24.1,2.35) #RR(24.1,2.01)
```

```
RR.1
```

```
## [1] 1.117664
```

```
RR.2
```

```
## [1] 1.425509
```

```
ER.1 <- ER(I01,RR.1)
```

```
ER.1*100000
```

```
## [1] 11.7915
```

```
ER.2 <- ER(I02,RR.2)
```

```
ER.2*100000
```

```
## [1] 14.89023
```

```
NRR.1 <- NRR(RR.1, lav, 0.7, Cav)
```

```
NRR.1*100000
```

```
## [1] 7.206903
```

```
NRR.2 <- NRR(RR.2, lav, 0.7, Cav)
```

```
NRR.2*100000
```

```
## [1] 26.06245
```

```
NER.1 <- NER(ER.1, 0.7, Cav)
```

```
NER.1*100000
```

```
## [1] 14.44458
```



```
NER.2 <- NER(ER.2, 0.7, Cav)
```

```
NER.2*100000
```

```
## [1] 18.24053
```

```
# "CI"
```

```
RR.1u <- RR(28.3,2.31)
```

```
RR.2u <- RR(24.1,5.08)
```

```
RR.1u
```

```
## [1] 1.344265
```

```
RR.2u
```

```
## [1] 1.962861
```

```
ER.1u <- ER(I01,RR.1u)
```

```
ER.1u*100000
```

```
## [1] 31.3598
```

```
ER.2u <- ER(I02,RR.2u)
```

```
ER.2u*100000
```

```
## [1] 28.32493
```

```
NRR.1u <- NRR(RR.1u, lav, 0.7, Cav)
```

```
NRR.1u*100000
```

```
## [1] 21.08621
```

```
NRR.2u <- NRR(RR.2u, lav, 0.7, Cav)
```

```
NRR.2u*100000
```



## [1] 58.97523

NER.1u <- NER(ER.1u, 0.7, Cav)

NER.1u\*100000

## [1] 38.41576

NER.2u <- NER(ER.2u, 0.7, Cav)

NER.2u\*100000

## [1] 34.69804

RR.1l <- RR(28.3,0.81)

RR.2l <- RR(24.1,1.09)

RR.1l

## [1] 0.9282449

RR.2l

## [1] 1.036405

ER.1l <- ER(I01,RR.1l)

ER.1l\*100000

## [1] -7.892731

ER.2l <- ER(I02,RR.2l)

ER.2l\*100000

## [1] 1.501852

NRR.1l <- NRR(RR.1l, lav, 0.7, Cav)

NRR.1l\*100000



```
## [1] -4.395003  
  
NRR.2l <- NRR(RR.2l, lav, 0.7, Cav)  
NRR.2l*100000  
  
## [1] 2.229831  
  
NER.1l <- NER(ER.1l, 0.7, Cav)  
NER.1l*100000  
  
## [1] -9.668596  
  
NER.2l <- NER(ER.2l, 0.7, Cav)  
NER.2l*100000  
  
## [1] 1.839769
```

## **B) Simulation der Taylorreihen-Approximation**

Dazu wird ausgehend von den in der Arbeit erwähnten empirischen Werten die im Appendix zu Kapitel 9 (S. 114-115) verwendete Approximation der Form  $\log(1 + A) \approx A$  simuliert. Für ER werden die oben in diesem Abschnitt unter Punkt A) berechneten Werte eingesetzt.

Zum Zweck der hier angestellten Simulation werden folgende Definitionen bzw. empirischen Werte verwendet:

Formel (9-8) ist von der Form  $X = \log(1+A)$  mit dem Term A wie folgt definiert:

$$A = ER \frac{C_i - C_{av}}{I_{av}}$$

Um eine Vorstellung von der Bandbreite der Werte zu vermitteln, die A angesichts der Datenanwendung annehmen kann, werden die folgenden Zahlen eingefügt:

- $C_{av} = 17.5 \mu\text{g}/\text{m}^3$  (siehe S. 110)



- $I_{av}$  ist die Gesamtkrebsinzidenz in der Zielbevölkerung (50 Fälle pro 100000 Personenjahre für die Region Basel, vgl. S. 109)
- Abbildung 10-1 auf S. 121 zeigt eine Schätzung plausibler PM10-Expositionswerte, wie sie in den Daten beobachtet wurden, die für  $C_i$  in den obigen Ausdruck für A eingesetzt werden können. Die Werte (mit einer relevanten Wahrscheinlichkeitsmasse) reichen von 10 bis 35. Aus Abbildung 10-1 ist auch ersichtlich, dass die Daten linksschief sind und etwa 90% der beobachteten Werte sind größer gleich 20 und etwa 50% der beobachteten Werte sind größer oder gleich 25.

#### *Parameterwerte und Funktion zur Berechnung von A*

```
Ci <- seq(from=10, to =35, length=500)
Cav <- 17.5
lav <- 50/100000

I01 <- 106/100000
I02 <- 42/100000

# FUKTION ZUR BERECHNUNG VON A

A <- function(ER, Ci,Cav, lav) ER*(Ci-Cav)/lav
```

#### *Ergebnisse der Simulationsstudie*

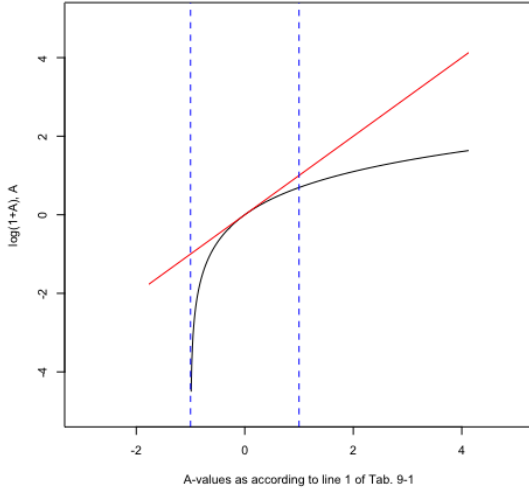
Die folgende Abbildung fasst die Ergebnisse dieser Simulation zusammen. Ausgehend von den Werten in Tabelle 9-1 auf S. 109 für RR (point estimates unten) und das dafür gegebene 95%-Konfidenzintervall (upper confidence limit und lower confidence limit unten), wird der Wertebereich von  $\log(1+A)$  als schwarzer Funktionsgraph dargestellt. Die rote Gerade stellt die Approximation von  $\log(1+A)$  durch A dar. Die blauen vertikalen Linien



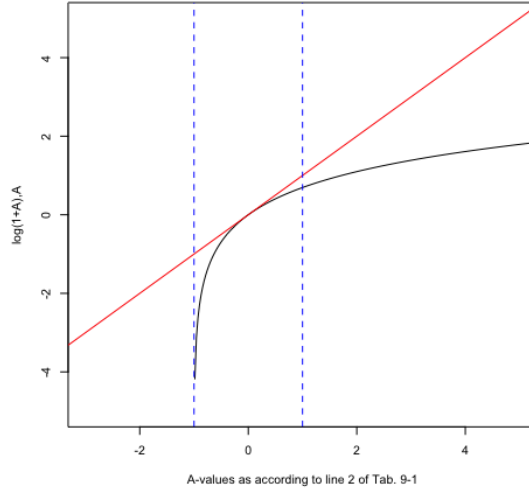
stellen den Bereich dar, innerhalb dessen der absolute Wert von  $A$  innerhalb von 1 liegt, sodass von einer guten Qualität der Approximation ausgegangen werden kann. In den Titeln der Grafiken sind auch die Werte für  $C_i$  angegeben, sodass  $A$  innerhalb des blau indizierten Wertebereichs liegt. Dass der (nahezu) volle Wertebereich der Daten durch die Approximation gut abgebildet wird, ist nur im Bereich der Werte für die untere Konfidenzgrenze gegeben. Im Bereich der Punktschätzer aus Tabelle 9-1 ist die Approximation für Werte von  $C_i$  im Wertebereich 13-22 bzw. 14-21 erfüllt. Vergleicht man dies mit der Wahrscheinlichkeitsmasse in den Kategorien in Abbildung 10-1 auf S. 120, so liegen schätzungsweise zwischen 60-90% der Daten außerhalb des blauen Bereichs, d.h. können nicht zuverlässig durch die Approximation in Form der roten Geraden abgebildet werden.



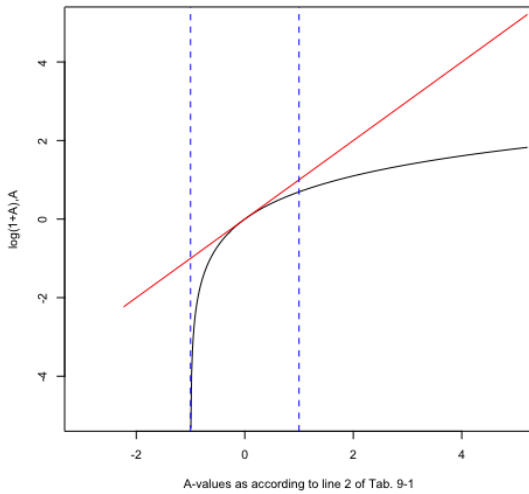
**Point estimates; CI-values: 13 to 22**



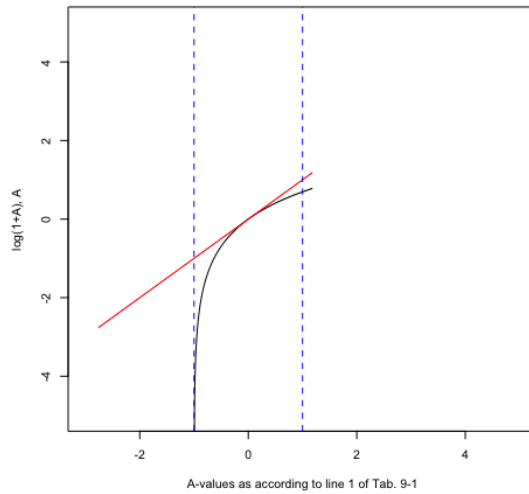
**Upper confidence limit; CI-values: 16 to 19**



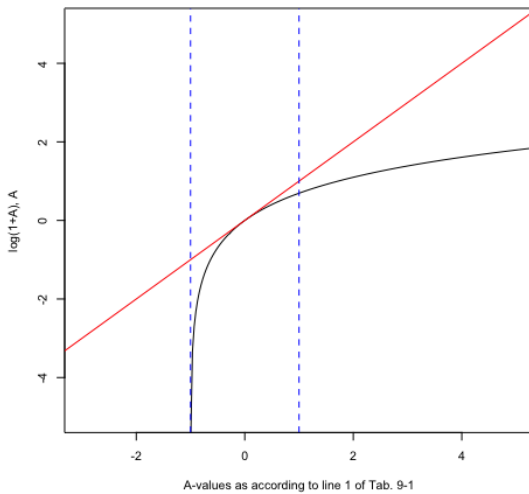
**Point estimates; CI-values: 14 to 21**



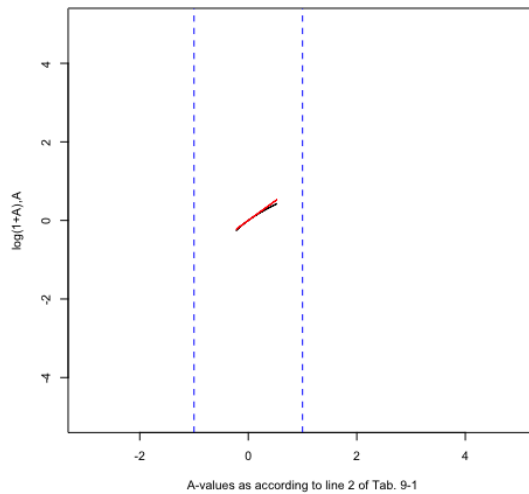
**Lower confidence limit; CI-values: 11 to 24**



**Upper confidence limit; CI-values: 16 to 19**



**Lower confidence limit; CI-values: 10 to 35**







## **7. Assessment of Cancer Risk Attributable to Air Pollution Using an Epidemiology Based Method (Kapitel 10)**

### **7.1 Zusammenfassung und Würdigung**

Kapitel 10 enthält ein auf den Berechnungen zu Kapitel 9 aufbauendes Datenbeispiel. Wie bereits zu Kapitel 9 angemerkt, erfüllen die Daten die wesentliche Annahme für den von den Autoren vorgeschlagenen *epidemiologiebasierten Ansatz* nicht. Somit ist davon auszugehen, dass die diesbezüglich in diesem Kapitel präsentierten Berechnungen ebenfalls verzerrt sind. Wie im vorangegangenen Kapitel wurde die Inferenz falsch berechnet bzw. eine Größe fälschlich als Konfidenzintervall bezeichnet. Weiters ergaben sich Schwierigkeiten bei der Replikation bzw. der Validierung des Designs der Meta-Analyse (poolability of studies).

### **7.2 Übersicht**

1. Tabelle 10-1, S. 120; Inhalt: Inzidenzen aus anderen Studien; Anmerkungen: Imputation des Wertes für Cancer Society zweifelhaft bzw. nicht weiter erklärt; Wie wurden Männer und Frauen für Adventists aggregiert (vgl. Tab. 9-1)?

Im Text auf S. 119 steht geschrieben, dass ein gepoolter Schätzer auf Basis einer Gewichtung mit der inversen Varianz berechnet wurde. Es werden dabei zwei Referenzen zitiert (Petitti, D. B. (1994a): *Meta-analysis, decision analysis, and cost-effectiveness analysis*, Oxford University Press, New York. und <https://www.sciencedirect.com/science/article/abs/pii/0197245686900462?via%3Dihub>). Auch auf Basis dieser konnten die Berechnungen in der Arbeit nicht nachvollzogen bzw. repliziert werden.

2. Tabelle 10-2, S. 122; Inhalt: Risikoschätzungen Substanzen; Anmerkungen: Geometrische Mittelwerte können repliziert werden



3. Tabelle 10-3, S. 124; Inhalt: Durchschnittskonzentrationen nach niedriger/mittlerer/hocher Exposition; Anmerkungen: Bevölkerungsgewichte zur Ermittlung der letzten Spalte in Text gegeben

4. Tabelle 10-4, S. 125; Inhalt: Schätzungen der Excess Rate; Anmerkungen: Konfidenzintervalle wegen negativer Werte nicht plausibel; Ermittlung des pooled value unklar

Zur Ermittlung des *pooled value* in Tabelle 10-4 steht auf S. 125 unten, dass es sich hierbei um einen gewichteten Durchschnitt (weighted average) handelt bzw. wie diese zu ermitteln sind.

Die in Tabelle 10-4 gegebenen Schätzungen für ER konnten in etwa repliziert werden. Da nicht ganz klar ist, wie RR aus Tabelle 10-1 (S. 120) für das gesamte Adventist-Sample ermittelt wurde, wurde hier von der gleichen Art der Durchschnittsbildung wie in Tabelle 9-1 auf S. 109 ausgegangen (siehe auch die Anmerkung dazu oben in Abschnitt 6.2). Zur Berechnung der ER wurde auf Formel (9-4) auf S. 110 zurückgegriffen. Siehe dazu die Ergebnisse in Abschnitt 7.3 Punkt C) unten;

5. Tabelle 10-5, S. 126; Inhalt: Ergebnis der epidemiologischen und der auf dem Einheitsrisiko basierenden Krebsrisikobewertung; Anmerkungen: Berechnung Konfidenzintervalle unklar; negative Werte für untere Konfidenzgrenze nicht plausibel



### 7.3 Berechnungen

#### A) Geometrische Mittel aus Tabelle 10-2 (exemplarisch)

$$(400 \cdot 150 \cdot 430 \cdot 330)^{(1/4)}$$

## [1] 303.762

$$(0.9 \cdot 0.57 \cdot 0.41 \cdot 2.9)^{(1/4)}$$

## [1] 0.8837406

$$(28 \cdot 17)^{(1/2)}$$

## [1] 21.81742

$$(38 \cdot 48 \cdot 26)^{(1/3)}$$

## [1] 36.19646

$$(7000 \cdot 8700)^{(1/2)}$$

## [1] 7803.845

$$(2.3 \cdot 0.53)^{(1/2)}$$

## [1] 1.104083

$$(0.043 \cdot 0.2)^{(1/2)}$$

## [1] 0.09273618

#### B) Population weighted average aus Tabelle 10-3 (exemplarisch)

Der Wert für PM10 ist aus der Tabelle nur sehr ungenau replizierbar, da die Werte in Intervallen gegeben sind.



```
w.low <- 50000/451000
```

```
w.av <-160000/451000
```

```
w.high <- 240000/451000
```

```
0.3*w.low + 0.5*w.av + 0.7*w.high # ARSENE
```

```
## [1] 0.5831486
```

```
0.13*w.low + 0.38*w.av + 0.6*w.high # 1,3-BUTADIENE
```

```
## [1] 0.4685144
```



### C) Replikation Ergebnisse Tabelle 10-4 aus 10-1

```
ER <- function(I0, RR) I0*log(RR) # (9-4)
```

```
I0.1 <- 106/100000
```

```
I0.2 <- 106/100000
```

```
I0.3 <- 61/100000
```

```
I0.4 <- 22/100000
```

```
I0.3. <- (22+61)/2/100000
```

```
r.1 <- 1.12
```

```
r.2 <- 1.01
```

```
r.3 <- 1.65
```

```
r.4 <- 1.13
```

```
r.3. <- (r.3 + r.4)/2
```

```
ER.1 <- ER(I0.1, r.1)*10000 # 1  $\mu\text{m}^3$ 
```

```
ER.2 <- ER(I0.2, r.2)*10000
```

```
ER.3 <- ER(I0.3., r.3.)*10000
```

```
round(ER.1, digits = 2)
```

```
## [1] 1.2
```

```
round(ER.2, digits = 2)
```

```
## [1] 0.11
```



```
round(ER.3, digits = 2)
```

```
## [1] 1.37
```

## **8. Zusammenfassende Beurteilung**

### **8.1 Rohdaten**

Auf die Datenanfrage hin, hat Prof. Rösli entsprechend guter wissenschaftlicher Praxis unmittelbar reagiert. Es wurde jedoch ein unvollständiger (Roh-)Datensatz übermittelt (keine meteorologischen Variablen, keine Daten zu Kapitel 5). Aus diesen über 20 Jahre alten Daten-Fragmenten konnten jedoch Behelfs-Datensätze zur qualitativen Untersuchung der wesentlichen empirischen Hypothesen aus den Kapiteln 6 bis 8 konstruiert werden. Die Ergebnisse aus den Kapiteln 9 und 10 konnten größtenteils aus den in der Arbeit gegebenen Informationen nachgerechnet werden.

### **8.2 Methodologie**

Insgesamt ziehen sich Fehler und Ungenauigkeiten bei der Berechnung der Inferenz über statistisch geschätzte Parameter durch die gesamte Arbeit. Insbesondere Konfidenzintervalle werden mathematisch falsch berechnet (Kapitel 9 und 10) oder beruhen auf einer Verteilungsannahme, die in den übermittelten Datenfragmenten nicht erfüllt ist (Kapitel 6 und 8 gehen von unabhängig und identisch normal-verteilten Residuen aus, während die Analysen der Behelfs-Datensätze Autokorrelation der PM10-Tages-Zeitreihe nahe legt sowie heteroskedastische und autokorrelierte Regressionsresiduen der Spezifikation in Kapitel 6 aus den Replikationsübungen anzunehmen sind). Ebenso ist die statistische Modellierung teilweise zweifelhaft (fehlende Regressionsspezifikation, in der meteorologische Variablen und Verkehrsaufkommen als simultane Erklärungsgründe zugelassen sind in den Kapiteln 6 und 8) und/oder Ergebnisse bzw. Berechnungen sind intransparent dargestellt (fehlende Nennung erklärender Variablen, der exakten Stichprobengröße; keine



Robustheits- und Residuenanalysen; nicht nachvollziehbare Berechnung der gepoolten Ergebnisse in Kapitel 10). Darüber hinaus liegt den Berechnungen in den Kapiteln 9 und 10 eine Taylorreihen-Approximation zu Grunde, die für rund 60-90% der Daten zu keiner guten Annäherung führt.

### **8.3 Interpretationen und Schlussfolgerungen**

Die Schlussfolgerung einer stark positiven Korrelation zwischen Referenz- und mobiler Messstation (Kapitel 6 bis 8) bleibt auf Basis der empirischen Erkenntnisse aus den Replikationsübungen mit den Behelfs-Datensätzen erhalten, jedoch wird anhand der Behelfs-Datensätze ein stark signifikant positiver Effekt des Verkehrsaufkommens auf die PM10-Belastung nahegelegt. Ein solcher wird in der Dissertation nur wage bestätigt bzw. werden Unterschiede (qualitativ bzw. narrativ) eher mit meteorologischen Phänomenen in Zusammenhang gebracht. Die – wie von den Autoren selbst eingeräumt – mögliche Überschätzung des Krebsrisikos für die Kapitel 9 und 10 mittels des in Kapitel 9 entwickelten epidemiologiebasierten Ansatzes könnte naheliegender Weise durch einen Bias aufgrund der oben bezeichneten – für den Wertebereich der Daten zu einer schlechten Annäherung führenden – Taylorreihen-Approximation erklärt werden. Anzumerken ist diesbezüglich, dass die Autoren die Ergebnisse stets sehr vorsichtig und damit angemessen interpretieren.

### **8.4 Wertung hinsichtlich möglichen wissenschaftlichen Fehlverhaltens**

Bei den Fehlern und Ungenauigkeiten bei der Berechnung statistischer Kennzahlen und der doch unbeholfen wirkenden Methodenauswahl dürfte es sich in der Würdigung um leichte bis gröbere Fahrlässigkeiten handeln, welche jedoch gehäuft auftreten und die inhaltlichen Ergebnisse der Kapitel 6 bis 8 teilweise berühren. Dabei sollte jedoch bedacht werden, dass die Ansprüche an die Ausgefeiltheit einer statistischen bzw. formalen Analyse über Disziplinen variieren und die Kapitel 6 bis 10 schließlich alle in internationalen Fachzeitschriften mit Peer-Review publiziert wurden. Dass im Rahmen der Analyse der



Kapitel 9 und 10 übersehen wurde, dass die Daten eine wesentliche Annahme für die angestellten Berechnungen nicht erfüllen sowie auch die negativen „Konfidenzintervalle“ nicht plausibel sind, entspricht keiner äußerst gewissenhaften Arbeitsweise, insgesamt allerdings entsteht der Eindruck, dass dies tatsächlich übersehen bzw. nicht bedacht wurde. Obwohl an manchen Stellen der Dissertation der subjektive Eindruck entstehen könnte, dass einer expliziten empirischen Untersuchung des Zusammenhangs zwischen Verkehrsaufkommen und PM10- bzw. Schadstoff-Konzentration – aus welchen Gründen auch immer – bewusst ausgewichen worden sein könnte, ist **eine offensichtliche Manipulation der Datengrundlage oder der berichteten Ergebnisse nicht zu erkennen**. Dennoch könnte den Autoren eine **gewisse Selektivität** dahingehend unterstellt werden, dass eine explizite empirische Untersuchung des Zusammenhangs von Verkehrsaufkommen (für meteorologische Bedingungen kontrollierend) und PM10-Konzentration nicht durchgeführt wurde. Dies könnte jedoch schlichtweg auch durch finanzielle Limitationen bei der Datenerhebung bzw. durch andere nicht von den Autoren zu verantwortende Einschränkungen erklärbar sein.





## **Prüfwerk:**

RÖÖSLI, Martin (2001): Spatial Variability of Air Pollutants in the Basel Area and Carcinogenic and Non-Carcinogenic Health Risk. Universität Basel, Dissertation. xii und 175 Seiten, [https://edoc.unibas.ch/16/1/DissB\\_5893.pdf](https://edoc.unibas.ch/16/1/DissB_5893.pdf)

## **Erklärung zur Irrtumswahrscheinlichkeit und Unbefangenheit**

Plagiats-, Autorschafts- und Titelüberprüfungen sind immer mit einer Rest-Irrtumswahrscheinlichkeit behaftet. Der Sachverständige hat bei Befunderhebung und Gutachten-Erstellung so sorgfältig und objektiv wie möglich gearbeitet.

Der Gutachter ist Sachverständiger für akademische Redlichkeit mit den Schwerpunkten Plagiats- und Titelüberprüfung. Er ist Autor des Buchs „*Das Google-Copy-Paste-Syndrom. Wie Netzplagiate Ausbildung und Wissen gefährden*“ (Heise, 2. Auflage, 2008) und hat bislang mehr als 200 Plagiatsfälle aus Wissenschaft, Politik und Journalismus dokumentiert. Seit 2007 ist er professionell mit der Erstellung von Gutachten und dem damit verknüpften Einsatz spezialisierter Software beschäftigt und hat seitdem hunderte Sprachwerke überprüft. Das Spektrum der zu begutachtenden Texte reichte dabei von Dissertationen bis zu Patentschriften, von Projektexposes bis zu Gerichtsgutachten.

Die Gutachter sind mit dem Verfasser des hier überprüften Sprachwerks nicht bekannt. Sie haben dieses Gutachten nach objektiven Parametern und wissenschaftlichen Standards sowie unabhängig von möglichen persönlichen, politischen oder wirtschaftlichen Einflussfaktoren erstellt. Ihr Prüfmaßstab folgt der Lehrbuchliteratur sowie der einschlägigen Judikatur.



Es handelt sich bei diesem Gutachten um das Ergebnis einer privaten Beauftragung.

**Das Gutachterteam:**

**Mag. Mag. Dr. ETH Sabrina Dorn, MSc ETH:** Check der empirischen Teile

**Birgit Kaiserlehner, MA:** *Turnitin*-Auswertung

**Doz. Dr. Stefan Weber:** Qualitätskontrolle, Endredaktion

Mit freundlichen Grüßen

  **DOZ. DR. STEFAN WEBER**  
Sachverständiger für Plagiatsprüfung  
Schopperstraße 10 +43 664 13 13 444  
5020 Salzburg weber@plagiatsgutachten.de  
Österreich <http://plagiatsgutachten.de>

Doz. Dr. Stefan Weber

Salzburg, 17.04.23