

Brigitte PFEILER

**Entwicklung von statistischen
Modellen zur Abschätzung solarer
Energie-Erträge**

DIPLOMARBEIT

**zur Erlangung des akademischen Grades einer
Diplom-Ingenieurin**

Diplomstudium Technische Mathematik



Graz University of Technology

Technische Universität Graz

Betreuer:

Univ.-Prof. Dipl.-Ing. Dr. techn. Ernst STADLOBER

Institut für Statistik

Graz, im November 2011

EIDESSTATTLICHE ERKLÄRUNG

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt, und die den benutzten Quellen wörtlich und inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Graz, am
.....
(Unterschrift)

STATUTORY DECLARATION

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly marked all material which has been quoted either literally or by content from the used sources.

.....
date
.....
(signature)

Kurzfassung

Das Ziel dieser Diplomarbeit ist die Entwicklung von statistischen Modellen, mit deren Hilfe der Energie-Output einer Solaranlage sehr einfach abgeschätzt werden kann. Diese Modelle benutzen Wetterdaten sowie bestehende Messdaten von sechs Solaranlagen aus der Grazer Region im Zeitraum 2002 bis 2008, die von der Firma S.O.L.I.D. GmbH zur Verfügung gestellt wurden. Für die Zielgrößen Tagesertrag und Monatsertrag der Wärmemenge [kWh/m²] konnten multiple Regressionsmodelle mit den erklärenden Variablen Globalstrahlung, Diffusstrahlung und Lufttemperatur gefunden werden. Basierend auf einem sorgfältig ausgewählten Datenpool liefern diese Modelle adäquate Ergebnisse mit Anpassungsgüten von 81% bis 90%. Für die Anwendung in der Praxis wurden die Modelle als Excel-Prognosefiles zur Verfügung gestellt.

Abstract

In this diploma thesis we develop statistical models, which deliver simple estimates of the energy-outputs of a solar plant. These models use meteorological data as well as existing measurements of six solar plants from the region of Graz, observed in the period 2002 to 2008. Those data were made available by the company S.O.L.I.D. GmbH. For the target variables daily yield and monthly yield of the heat quantity [kWh/m²] we found multiple regression models with the explanatory variables global radiation, diffuse radiation and air temperature. Based on a carefully selected pool of data, these models provide adequate results with goodness of fits of 81% to 90%. For application in practice, the models were implemented as Excel forecast files.

Danksagung

An dieser Stelle möchte ich mich besonders bei Herrn Univ.-Prof. Dipl.-Ing. Dr. Ernst Stadlober bedanken, der mich während meiner Diplomarbeit betreut und umfangreich unterstützt hat.

Mein Dank gilt auch der Firma S.O.L.I.D. GmbH für die Ermöglichung der Durchführung dieser Diplomarbeit durch die Überlassung des Datenmaterials, das als Grundlage für diese Arbeit diente.

Danke sagen möchte ich außerdem Herrn DI Philip Ohnewein von der Firma S.O.L.I.D. GmbH und Herrn DI Robert Hausner von der Firma AEE INTEC, die mir von ihnen in VBA programmierte Excel-Files zur Verfügung gestellt haben.

Ein herzliches Dankeschön geht auch an Verena Feirer, die mich bei der Fertigstellung dieser Arbeit tatkräftig unterstützt hat.

Nicht zuletzt möchte ich mich bei meinen Eltern bedanken, die mir das Studium erst ermöglicht haben und da im Speziellen bei meinem im April 2011 verstorbenen Vater.



Inhaltsverzeichnis

1	Einleitung	1
2	Eckdaten und Datenmaterial	3
2.1	Anlage- und Wetterdaten	3
2.2	Aufzeichnungstage der einzelnen Solaranlagen	5
2.3	Aufbereitung der Tagesvariablen	8
3	Explorative Analyse der Zielvariablen WMZ und WMZ/m²	13
3.1	Tageserträge pro Monat in kWh und kWh/m ²	13
3.2	Vergleich der sechs Grazer Solaranlagen	26
4	Einfluss der Regressoren auf die Zielvariable WMZ/m²	29
4.1	Korrelationen und lineare Regressionen der Tageswerte	29
4.2	Analyse der Anlagedaten	50
5	Prognosen für den Tagesertrag pro m²	53
5.1	Lineare Regressionsmodelle mit berechneten & vorhersagbaren Variablen	53
5.1.1	Modelle mit geneigter Strahlung auf Kollektorebene (berechnet)	53
5.1.2	Modelle mit horizontaler Strahlung (gemessen)	56
5.2	Lineare Regressionsmodelle ohne Kenntnis der Anlagedaten	59
5.2.1	Modelle mit geneigter Strahlung auf Kollektorebene (berechnet)	59
5.2.2	Modelle mit horizontaler Strahlung (gemessen)	61
6	Prognosen für den Monatsertrag pro m²	65
6.1	Lineare Regressionsmodelle mit berechneten & vorhersagbaren Variablen	65
6.1.1	Modelle mit geneigter Strahlung auf Kollektorebene (berechnet)	66
6.1.2	Modelle mit horizontaler Strahlung (gemessen)	68
6.2	Lineare Regressionsmodelle ohne Kenntnis der Anlagedaten	71

6.2.1	Modelle mit geneigter Strahlung auf Kollektorebene (berechnet)	72
6.2.2	Modelle mit horizontaler Strahlung (gemessen)	74
7	Mathematische und Statistische Grundlagen	77
7.1	Daten, Kenngrößen und Graphiken	77
7.1.1	Datenaspekt	77
7.1.2	Statistische Kenngrößen	79
7.1.3	Explorative Graphiken	81
7.2	Stetige Verteilungen	84
7.2.1	Dichte und Verteilungsfunktion	84
7.2.2	Normalverteilung	85
7.3	Parameterschätzung	86
7.3.1	Punktschätzung	86
7.3.2	Intervallschätzung	88
7.4	Testverteilungen	89
7.4.1	Chi-Quadrat-Verteilung	89
7.4.2	Student-t-Verteilung	90
7.4.3	Fishers F-Verteilung	90
7.5	Lineare Regression	91
7.5.1	Einfache lineare Regression	91
7.5.2	Multiple lineare Regression	93
7.5.3	Graphische Modelldiagnostik	102
7.5.4	Methoden der linearen Regressionsanalyse im SPSS	104
	Zusammenfassung	107
	A Excel-Prognosefiles	109
	Literaturverzeichnis	115

Tabellenverzeichnis

2.1	Eckdaten der Solaranlagen	3
2.2	Aufzeichnungsmonate	5
2.3	Korrekte Aufzeichnungstage	6
2.4	„Verwendbare“ (links) & „Nicht Verwendbare“ (rechts) Aufzeichnungstage	7
3.1	Summe der Erträge in kWh pro Monat	13
3.2	Anzahl der Aufzeichnungstage pro Monat & Jahr (Stadion Liebenau) .	15
3.3	Tageserträge pro Monat in kWh/m ² (Stadion Liebenau)	16
3.4	Anzahl der Aufzeichnungstage pro Monat & Jahr (Grottenhofstraße) .	17
3.5	Tageserträge pro Monat in kWh/m ² (Grottenhofstraße)	18
3.6	Anzahl der Aufzeichnungstage pro Monat & Jahr (Eggenberger Allee) .	19
3.7	Tageserträge pro Monat in kWh/m ² (Eggenberger Allee)	20
3.8	Anzahl der Aufzeichnungstage pro Monat & Jahr (Schwarzen Weg) . .	21
3.9	Tageserträge pro Monat in kWh/m ² (Schwarzen Weg)	21
3.10	Anzahl der Aufzeichnungstage pro Monat & Jahr (Theodor-Körner-Str.)	22
3.11	Tageserträge pro Monat in kWh/m ² (Theodor-Körner-Straße)	23
3.12	Anzahl der Aufzeichnungstage pro Monat & Jahr (Judendorf)	24
3.13	Tageserträge pro Monat in kWh/m ² (Judendorf)	25
3.14	Tageserträge pro m ²	26
3.15	Erträge pro m ² & Viertelstunde	27
3.16	Eingeschaltete Viertelstunden pro Tag (onoff=1)	28
3.17	Anzahl der Tage mit Ausreißern	28
4.1	Korrelationen der Anlage Stadion Liebenau	30
4.2	Lineare Regression (Stadion Liebenau)	32
4.3	Korrelationen der Anlage Grottenhof	32
4.4	Lineare Regression (Grottenhof)	34

4.5	Korrelationen der Anlage Eggenberger Allee	35
4.6	Lineare Regression (Eggenberg)	37
4.7	Korrelationen der Anlage Schwarzer Weg	38
4.8	Lineare Regression (Schwarzer Weg)	40
4.9	Korrelationen der Anlage Theodor-Körner-Straße	41
4.10	Lineare Regression (Theodor-Körner-Straße)	42
4.11	Korrelationen der Anlage Judendorf	43
4.12	Lineare Regression (Judendorf)	45
4.13	Korrelationen aller Solaranlagen	46
4.14	Lineare Regression (Alle Solaranlagen)	48
4.15	Zusammenfassung der linearen Regressionen aller Solaranlagen	49
5.1	Lineare Regressionsansätze für die Prognosen aller Solaranlagen	53
5.2	Prognosemodell mit berechneten Regressoren für alle Solaranlagen	54
5.3	Prognosemodell mit vorhersagbaren Regressoren für alle Solaranlagen	56
5.4	Prognosemodell ohne geographische Koordinaten der Solaranlagen	58
5.5	Prognosemodell mit den Regressoren Saison, Ediff, Eglob und Lute	59
5.6	Prognosemodell mit den Regressoren Saison, Globalstrahlung, Diffusanteil und Lufttemperatur	61
6.1	Zusammenfassung der linearen Regressionen aller Solaranlagen	65
6.2	Lineare Regressionsansätze für die Prognosen aller Solaranlagen	66
6.3	Prognosemodell mit berechneten Regressoren für alle Solaranlagen	67
6.4	Prognosemodell mit vorhersagbaren Regressoren für alle Solaranlagen	69
6.5	Prognosemodell ohne geographische Koordinaten der Solaranlagen	71
6.6	Prognosemodell mit den Regressoren Saison, Ediff, Eglob und Lute	72
6.7	Prognosemodell mit den Regressoren Globalstrahlung, Diffusanteil und Lufttemperatur	74
7.1	Symbolisierung des Signifikanzniveaus in SPSS	99

Abbildungsverzeichnis

2.1	Solaranlagen Stadion Liebenau (links) und Judendorf (rechts)	4
2.2	Solaranlagen Grottenhofstraße (links) und Eggenberger Allee (rechts) .	4
3.1	Säulendiagramme der Ertragssummen in kWh pro Monat	14
3.2	Fehlerbalken & Boxplots der Tageserträge pro Monat in kWh/m ² (Stadion Liebenau)	16
3.3	Fehlerbalken & Boxplots der Tageserträge pro Monat in kWh/m ² (Grottenhofstraße)	18
3.4	Fehlerbalken & Boxplots der Tageserträge pro Monat in kWh/m ² (Eggenberger Allee)	20
3.5	Fehlerbalken & Boxplots der Tageserträge pro Monat in kWh/m ² (Schwarzen Weg)	22
3.6	Fehlerbalken & Boxplots der Tageserträge pro Monat in kWh/m ² (Theodor-Körner-Straße)	23
3.7	Fehlerbalken & Boxplots der Tageserträge pro Monat in kWh/m ² (Judendorf)	25
3.8	Fehlerbalken & Boxplots der Tageserträge pro m ²	27
4.1	Streudiagramme der Regressoren mit WMZ/m ² (Stadion Liebenau) . .	31
4.2	Streudiagramme der Regressoren mit WMZ/m ² (Grottenhof)	33
4.3	Streudiagramme der Regressoren mit WMZ/m ² (Eggenberg)	36
4.4	Streudiagramme der Regressoren mit WMZ/m ² (Schwarzer Weg) . . .	39
4.5	Streudiagramme der Regressoren mit WMZ/m ² (Theodor-Körner-Str.)	42
4.6	Streudiagramme der Regressoren mit WMZ/m ² (Judendorf)	44
4.7	Streudiagramme der Regressoren mit WMZ/m ² (Alle Solaranlagen) . .	47
4.8	Fehlerbalken & Boxplots der Tageserträge/m ² bzgl. der Saison	48
4.9	Boxplots der Tageserträge/m ² bzgl. der Anlagedaten	51

5.1	Q-Q-Diagramm der standardisierten Residuen und Residuenplot	55
5.2	Liniendiagramme WMZ/m^2 geschätzt- WMZ/m^2	56
5.3	Q-Q-Plot der standardisierten Residuen und Residuenplot	57
5.4	Liniendiagramme WMZ/m^2 geschätzt- WMZ/m^2	58
5.5	Histogramm und Q-Q-Plot der standardisierten Residuen	60
5.6	Residuenplot und Liniendiagramm WMZ/m^2 geschätzt- WMZ/m^2 . . .	61
5.7	Histogramm und Q-Q-Plot der standardisierten Residuen	62
5.8	Residuenplot und Liniendiagramm WMZ/m^2 geschätzt- WMZ/m^2 . . .	63
5.9	Partielle Regressionsdiagramme der Regressoren und WMZ/m^2	64
6.1	Histogramm und Q-Q-Plot der standardisierten Residuen	67
6.2	Residuenplot und Liniendiagramm WMZ/m^2 geschätzt- WMZ/m^2 . . .	68
6.3	Histogramm und Q-Q-Plot der standardisierten Residuen	70
6.4	Residuenplot und Liniendiagramm WMZ/m^2 geschätzt- WMZ/m^2 . . .	70
6.5	Histogramm und Q-Q-Plot der standardisierten Residuen	73
6.6	Residuenplot und Liniendiagramm WMZ/m^2 geschätzt- WMZ/m^2 . . .	73
6.7	Histogramm und Q-Q-Plot der standardisierten Residuen	75
6.8	Residuenplot und Liniendiagramm WMZ/m^2 geschätzt- WMZ/m^2 . . .	75
6.9	Partielle Regressionsdiagramme der Regressoren und WMZ/m^2	76
7.1	Definition des Boxplots	82
7.2	Quadratsummenzerlegung	97

1 Einleitung

Diese Diplomarbeit wurde in Zusammenarbeit mit folgenden Firmen fertiggestellt:

- **S.O.L.I.D.** Gesellschaft für Solarinstallation und Design m.b.H ist ein steirisches Solartechnikunternehmen, das sich auf sämtliche Anwendungsgebiete für thermische Großsolaranlagen spezialisiert hat. Zusätzlich arbeitet die Firma an Forschungs- und Entwicklungsprojekten (F&E) mit, sowohl auf nationaler als auch internationaler EU-Ebene.¹
- **AEE INTEC** (Arbeitsgemeinschaft Erneuerbarer Energie - Institut für Nachhaltige Technologien) wurde als unabhängiger gemeinnütziger Verein zur Förderung des sinnvollen Einsatzes erneuerbarer Energien und der rationellen, nachhaltigen Energienutzung gegründet. Das Institut beschäftigt sich mit der Erforschung der naturwissenschaftlich-technischen Grundlagen der Solarenergienutzung, mit der Entwicklung von energieeffizienten Energieversorgungssystemen für Gebäude sowie mit nachhaltigen Techniken für Wasserversorgung und Abwasserentsorgung.²

Die Messdaten wurden mittels `Microsoft Office Excel 2007` bearbeitet und mit dem Programm `SPSS Version 15.0` (Windows XP, Windows 2000, Windows NT) „The Statistical Package for Social Sciences“ statistisch analysiert. Die abschließenden Prognosemodelle wurden in `Microsoft Office Excel 2007` mit Makros realisiert.

¹<http://www.solid.at>

²<http://www.aee-intec.at>

2 Eckdaten und Datenmaterial

2.1 Anlage- und Wetterdaten

Für die Analyse von Solaranlagen wurden von der Firma S.O.L.I.D. GmbH durch Herrn DI Philip Ohnewein Anlage- und Wetterdaten zur Verfügung gestellt. Die Anlagedaten stammen von den in Tabelle 2.1 dargestellten fünf SOLID-Solaranlagen in Graz, sowie einer in Judendorf-Straßengel, die folgende Eckdaten besitzen.

Tabelle 2.1: Eckdaten der Solaranlagen

	S O L A R A N L A G E					
	Stadion Liebenau	Grottenhof- straße	Eggenberger Allee	Schwarzer Weg	Theodor-Körner- Straße	Judendorf
Gesamt Netto- Absorberfläche [m ²]	1.242,7	444,3	104,6	84,3	201,2	87,7
geographische Breite [°]	47,05	47,05	47,07	47,03	47,09	47,11
geographische Länge [°]	15,45	15,4	15,4	15,41	15,43	15,32
Ausrichtungswinkel [°]	-21	0	3	-17	-19,5	12
Neigungswinkel Kollektor[°]	31	25	27	27	30	39
Leitungslänge außen [m]	125	32	45	25	0	0
Glykolgehalt [%]	33	37	33	32	32,5	33
Wärmekapazität [J/(m ² K)]	6.040	6.810	6.040	6.040	6.040	6.040

Die mit Abstand größte **Netto-Absorberfläche** weist die Anlage am Dach des **Liebenauer Stadions** mit 1.242,7 m² auf, gefolgt von den Anlagen in der **Grottenhofstraße** mit 444,3 m² und in der **Theodor-Körner-Straße** mit 201,2 m². Vergleichbar kleine Flächen besitzen die Anlagen in der **Eggenberger Allee** mit 104,6 m², in **Judendorf** mit 87,7 m² und am **Schwarzen Weg** mit 84,3 m². Die Abbildungen 2.1 und 2.2 auf der nächsten Seite veranschaulichen diese Größenunterschiede.

Die in Tabelle 2.1 enthaltenen **geographischen Koordinaten** der einzelnen Solaranlagen zeigen in der Breite einen Unterschied von 0,08° und in der Länge von 0,13°. Auf diese Nähe der einzelnen Solaranlagen zueinander wird bei der Berechnung der Prognosemodelle zu achten sein. Hingegen könnten die unterschiedlichen **Ausrichtungswinkel**, **Neigungswinkel**, **äußeren Leitungslängen**, **Glykolgehalte**, sowie die in



Abbildung 2.1: Solaranlagen Stadion Liebenau (links) und Judendorf (rechts)

der Grottenhofstraße von den anderen Anlagen verschiedene **Wärmekapazität** von statistischem Interesse sein, was in Kapitel 4.2 auf Seite 50 näher untersucht wird.

Die Anlagen in Abbildung 2.1 werden zur Energieeinspeisung ins jeweilige Fernwärmenetz in Graz und Judendorf verwendet, während die Anlagen in Abbildung 2.2 der Warmwasserbereitung und Heizungsunterstützung der in den dazugehörigen Häusern befindlichen Wohneinheiten bzw. Geschäftsflächen dienen.



Abbildung 2.2: Solaranlagen Grottenhofstraße (links) und Eggenberger Allee (rechts)

Zusätzlich zu den Eckdaten aus Tabelle 2.1 auf der vorherigen Seite enthielten die Daten für jede Solaranlage die **Messparameter** WMZ in kWh , *primärseitige Vor- und Rücklauftemperatur* in $^{\circ}C$ und „onoff“ mit den Werten 0 und 1 . Die Abkürzung WMZ steht für *Wärmemengenzähler* und gibt über den Solaren Ertrag in kWh Auskunft. Die Variable „onoff“ beschreibt den Pumpenstatus der Sekundärpumpe und bedeutet, dass nur bei Wert 1 Solarer Ertrag geliefert werden kann.

Weiters gab es noch Aufzeichnungsdaten der Solaranlage in der **Lange Gasse**, jedoch war der Zeitraum von 02.09.2005 bis 15.01.2006 mit 4,5 Monaten für eine statistische Analyse nicht ausreichend.

Darüber hinaus wurden von der Solaranlage in der **Krottendorfer Straße** Aufzeichnungsdaten für den Zeitraum März 2007 bis Dezember 2008 geliefert, die sich aber im Laufe der Datenaufbereitung als zu fehlerhaft für eine fundierte statistische Analyse herausgestellt haben.

Die **Wetterdaten** auf Tagesebene wurden von der **ZAMG** (Zentralanstalt für Meteorologie und Geodynamik) für den Standort **Universität Graz** zur Verfügung gestellt und bestehen aus *Globalstrahlung [J/cm²]*, *Diffusstrahlung [J/cm²]*, *Sonnenscheindauer [min]*, *Lufttemperatur [°C]*, *Windgeschwindigkeit [m/s]*, *Windrichtung [°]* und *Luftfeuchtigkeit [%]* für den Zeitraum **Jänner 2002 bis Dezember 2008**.

Im Folgenden findet sich eine kurze Übersicht zu den betrachteten Stationen inklusive der dazugehörigen Daten der korrekten und fehlenden Aufzeichnungstage, festgestellte Aufzeichnungsfehler und eine Beschreibung der notwendigen Aufbereitung der Daten zur anschließenden statistischen Analyse.

2.2 Aufzeichnungstage der einzelnen Solaranlagen

Einen ersten Überblick in welchen Monaten (1-12) und Jahren (2002-2008) je Solaranlage Daten aufgezeichnet wurden gibt Tabelle 2.2.

Tabelle 2.2: Aufzeichnungsmonate

	J A H R						
	2002	2003	2004	2005	2006	2007	2008
Stadion Liebenau	7-11	1-2, 6-12	2 - 12	1-3, 5-11	3-12	1-11	1-12
Grottenhofstraße						6-12	1-12
Eggenberger Allee				1-12	1		
Schwarzer Weg				1-12	1		
Theodor-Körner-Straße				1-12	1		
Judendorf	1-12	1-12	1-12	1-12	1-12	10-12	1-9

Gemeinsam mit Tabelle 2.3 kann man sehen, dass die Solaranlage am Dach des **Liebenauer Stadions** in den **Jahren 2002 bis 2008** insgesamt **1.131 korrekte Aufzeichnungstage** lieferte, was bei insgesamt 2.102 Tagen in 69 Monaten einen **Anteil** von ca. **54%** ergibt. In der **Grottenhofstraße** waren mit **447** von 580 Tagen in 19 Monaten ca. **77%** für die weitere Analyse verwendbar, während sich die Aufzeichnungsqualität in der **Eggenberger Allee** mit fast **90%**, sowie am **Schwarzer Weg** und in der **Theodor-Körner-Straße** mit **100%** als hervorragend herausstellte. Von der Anlage in **Judendorf** konnten mit **2.080** von 2.192 Tagen in 72 Monaten ebenfalls fast **95%** in die weiteren Analysen einfließen.

Tabelle 2.3: Korrekte Aufzeichnungstage

	Ertrag = 0		Ertrag > 0		SUMME	
	Anzahl	Prozent	Anzahl	Prozent		
Stadion Liebenau	339	30,0 %	792	70,0 %	1.131	Tage
Grottenhofstraße	78	17,4 %	369	82,6 %	447	Tage
Eggenberger Allee	63	16,1 %	329	83,9 %	392	Tage
Schwarzer Weg	92	23,2 %	304	76,8 %	396	Tage
Theodor-Körner-Straße	65	16,4 %	331	83,6 %	396	Tage
Judendorf	381	18,3 %	1.699	81,7 %	2.080	Tage
Gesamt	1.018	21,0 %	3.824	79,0 %	4.842	Tage

Für die statistischen Analysen dieser Arbeit wurden im Folgenden nur Aufzeichnungstage mit Ertrag > 0 berücksichtigt.

Demnach standen von der Solaranlage am Dach des **Liebenauer Stadions** in den Jahren 2002 und 2003 je 16%, 2004 43%, 2005 47%, 2006 30%, 2007 24% und 2008 41% verwendbare Tage zur Verfügung. In der **Grottenhofstraße** waren es im Jahr 2007 36% und 2008 65%, von den Anlagen in der **Eggenberger Allee**, am **Schwarzer Weg** und in der **Theodor-Körner-Straße** konnten im Jahr 2005 zwischen 78% und 85%, sowie 2006 ca. 5% verwendet werden.

Die Aufstellungen in Tabelle 2.4 geben einen genauen Überblick über die Datenlage der einzelnen Solaranlagen. Die Tabellen auf der linken Seite, welche in Kapitel 3.1 näher betrachtet werden, zeigen die Anzahl der im Weiteren berücksichtigten Aufzeichnungstage nach Monaten und Jahren pro Anlage, während die Tabellen rechts die Anzahl der fehlenden beziehungsweise fehlerhaften Tage und jenen mit Ertrag = 0 auflisten.

2.2. AUFZEICHNUNGSTAGE DER EINZELNEN SOLARANLAGEN

Tabelle 2.4: „Verwendbare“ (links) & „Nicht Verwendbare“ (rechts) Aufzeichnungstage

Stadion Liebenau

	2002	2003	2004	2005	2006	2007	2008	Gesamt		2002	2003	2004	2005	2006	2007	2008	Gesamt	
Jänner		2		9		10	4	25	Jänner	29			22			21	27	99
Februar		12	6	13		2	8	41	Februar	16	23	15			26	21	101	
März			16	21	15	4	14	70	März			15	10	16	27	17	85	
April			10		9	4	17	40	April			20		21	26	13	80	
Mai				27	15	21	9	18	90	Mai			4	16	10	22	13	65
Juni			12	27	20	9	11	8	87	Juni	18	3	10	21	19	22	93	
Juli	7	10	20	21	22	12	23	115	Juli	24	21	11	10	9	19	8	102	
August	16		13	25	11	10	17	92	August	15		18	6	20	21	14	94	
September	19	7	14	16	9	11	13	89	September	11	23	16	14	21	19	17	121	
Oktober	11	7	15	22	1	11	14	81	Oktober	20	24	16	9	30	20	17	136	
November	6	7	8	7	8	4	11	51	November	24	23	22	23	22	26	19	159	
Dezember			3	1	4		3	11	Dezember			28	30	27		28	113	
Gesamt	59	57	159	170	109	88	150	792	Gesamt	94	154	176	165	197	246	216	1.248	
	16%	16%	43%	47%	30%	24%	41%											

Grottenhofstraße

	2002	2003	2004	2005	2006	2007	2008	Gesamt		2002	2003	2004	2005	2006	2007	2008	Gesamt
Jänner							20	20	Jänner							11	11
Februar							24	24	Februar							5	5
März							22	22	März							9	9
April							27	27	April							3	3
Mai									Mai								
Juni						23	14	37	Juni						7	16	23
Juli						27	30	57	Juli						4	1	5
August						18	29	47	August						13	2	15
September						25	27	52	September						5	3	8
Oktober						10	25	35	Oktober						21	6	27
November						19	16	35	November						11	14	25
Dezember						10	3	13	Dezember						21	28	49
Gesamt						132	237	369	Gesamt						82	98	180
						36%	65%										

Eggenberger Allee

	2002	2003	2004	2005	2006	2007	2008	Gesamt		2002	2003	2004	2005	2006	2007	2008	Gesamt
Jänner				27	21			48	Jänner				4	10			14
Februar				23				23	Februar				5				5
März				30				30	März				1				1
April				27				27	April				3				3
Mai				30				30	Mai				1				1
Juni				30				30	Juni				0				0
Juli				30				30	Juli				1				1
August				25				25	August				6				6
September				26				26	September				4				4
Oktober				26				26	Oktober				5				5
November				15				15	November				15				15
Dezember				19				19	Dezember				12				12
Gesamt				308	21			329	Gesamt				57	10			67
				84%	6%												

Schwarzer Weg

	2002	2003	2004	2005	2006	2007	2008	Gesamt		2002	2003	2004	2005	2006	2007	2008	Gesamt
Jänner				24	20			44	Jänner				7	11			18
Februar				19				19	Februar				9				9
März				24				24	März				7				7
April				24				24	April				6				6
Mai				30				30	Mai				1				1
Juni				24				24	Juni				6				6
Juli				28				28	Juli				3				3
August				28				28	August				3				3
September				25				25	September				5				5
Oktober				26				26	Oktober				5				5
November				13				13	November				17				17
Dezember				19				19	Dezember				12				12
Gesamt				284	20			304	Gesamt				81	11			92
				78%	5%												

Theodor-Körner-Straße

	2002	2003	2004	2005	2006	2007	2008	Gesamt		2002	2003	2004	2005	2006	2007	2008	Gesamt
Jänner				23	21			44	Jänner				8	10			18
Februar				19				19	Februar				9				9
März				29				29	März				2				2
April				27				27	April				3				3
Mai				30				30	Mai				1				1
Juni				30				30	Juni				0				0
Juli				29				29	Juli				2				2
August				29				29	August				2				2
September				30				30	September				0				0
Oktober				28				28	Oktober				3				3
November				14				14	November				16				16
Dezember				22				22	Dezember				9				9
Gesamt				310	21			331	Gesamt				55	10			65
				85%	6%												

Judendorf

	2002	2003	2004	2005	2006	2007	2008	Gesamt		2002	2003	2004	2005	2006	2007	2008	Gesamt
Jänner	21	17	23	26	20		20	127	Jänner	10	14	8	5	11		11	59
Februar	23	23	21	22	22		23	134	Februar	5	5	8	6	6		6	36
März	26	29	22	28	24		26	155	März	5	2	9	3	7		5	31
April	21	30	22	28	26		29	156	April	9	0	8	2	4		1	24
Mai	27	30	29	30	28		30	174	Mai	4	1	2	1	3		1	12
Juni	29	28	30	26	28		14	155	Juni	1	2	0	4	2		16	25
Juli	30	31	29	19	31		31	171	Juli	1	0	2	12	0		0	15
August	30	30	29	23	23		29	164	August	1	1	2	8	8		2	22
September	27	28	29	25	10		18	137	September	3	2	1	5	20		12	43
Oktober	24	26	22	25	13	11		121	Oktober	7	5	9	6	18	20		65
November	20	16	15	15	25	21		112	November	10	14	15	15	5	9		68
Dezember	9	19	14	14	18	19		93	Dezember	22	12	17	17	13	12		93
Gesamt	287	307	285	281	268	51	220	1.699	Gesamt	78	58	81	84	97	41	54	493
	79%	84%	78%	77%	73%	14%	60%										

2.3 Aufbereitung der Tagesvariablen

Für die statistische Analyse der sechs zur Verfügung stehenden Solaranlagen im Raum Graz mussten die vorhandenen Daten in Tageswerte umgerechnet werden. Diese Datenaufbereitung hat einen Großteil der Zeit, die für diese Arbeit benötigt wurde, beansprucht!

- WMZ [Ertrag in kWh]: $WMZ^{24Uhr} - WMZ^{0Uhr}$ bzw. $\sum WMZ$

Der Messparameter *WMZ* (*Wärmemengenzähler*), der über den Solaren Ertrag in *kWh* Auskunft gibt, wurde auf zwei unterschiedliche Arten aufbereitet.

1. Für die Stationen Stadion Liebenau, Grottenhofstraße und Judendorf wurden die Differenzen zwischen den Werten um 24:00 und 0:00 Uhr desselben Tages berechnet.

2. Für die Stationen Eggenberger Allee, Schwarzer Weg und Theodor-Körner-Straße konnten die Tagessummen der Werte verwendet werden.

- WMZ/m² [kWh/m²]:

Der Parameter WMZ/m^2 gibt den Tagesertrag in kWh pro m^2 an, d.h. es wurden die Tageserträge durch die Gesamt Netto-Absorberfläche (= Kollektorfläche) jeder einzelnen Anlage dividiert. Dieser Variablen kommt eine wichtige Bedeutung zu, da dadurch die Anlagen vergleichbar werden.

- dT1 [°C]: $\frac{1}{2} (PVL + PRL) onoff$

Die Variable $dT1$ entspricht dem arithmetischen Tagesmittelwert zwischen den primärseitigen Vor- und Rücklauftemperaturen in °C, gewichtet mittels der Variable „onoff“ (0: kein Ertrag | 1: Ertrag möglich).

- 1/4h-on = DL [Tageslänge]: $\sum \frac{1}{4}h$

Die Variable DL gibt die Summe der Viertelstunden an, in denen die Variable „onoff“ den Wert 1 (= Anlage läuft) am Tag annimmt.

- dT2 [°C]:

Die Variable $dT2$, die in °C angegeben ist, wurde aus der Temperaturdifferenz zwischen der Außentemperatur zum Zeitpunkt des Sonnenaufgangs und der Variable $dT1$ zum Zeitpunkt des letzten Ausschaltens der Solaranlage (= letzter Zeitpunkt mit „onoff“ = 1) desselben Tages berechnet.

- eta0_dir, eta0_diff, c1 & c2: $x (PVL - PRL) onoff$

Die Variablen $eta0_dir$, $eta0_diff$, $c1$ und $c2$ stellen die Tagesmittelwerte der zeitabhängigen Kollektorkennwerte aus der „Kollektorregression“ von Herrn DI Robert Hausner (Firma AEE INTEC, Gleisdorf) dar, gewichtet mit der Differenz der primären Vor- und Rücklauftemperatur multipliziert mit der Variable „onoff“.

- Direktstrahlung [W/m²]:

Die Variable *Direktstrahlung* bezieht sich auf die Tagessumme der direkten Sonneneinstrahlung auf die Kollektorfläche in W/m^2 ($= \frac{10.000}{3.600} J/cm^2$).

- Edir [W/m^2]: Die Variable *Edir* gibt die Tagessumme der Direktstrahlung auf die geneigte Kollektorebene in W/m^2 an, die mit Hilfe des Excel-Files „Strahlungsprozessor“ von Herrn DI Philip Ohnewein berechnet wurde, gewichtet mittels der Variable „onoff“.

- Diffusstrahlung [W/m^2]:

Die Variable *Diffusstrahlung* bezieht sich auf die Tagessumme der diffusen Sonneneinstrahlung auf die Kollektorfläche in W/m^2 . Bei der Berechnung der Prognosen in Kapitel 5 und 6 wurde anstelle der Diffusstrahlung in W/m^2 der *Diffusanteil* an der Globalstrahlung in % verwendet.

- Ediff [W/m^2]:

Die Variable *Ediff* gibt die Tagessumme der Diffusstrahlung auf die geneigte Kollektorebene in W/m^2 an, die mit Hilfe des Excel-Files „Strahlungsprozessor“ von Herrn DI Philip Ohnewein berechnet wurde, gewichtet mittels der Variable „onoff“.

- Globalstrahlung [W/m^2]:

Die Variable *Globalstrahlung* bezieht sich auf die Tagessumme der globalen Sonneneinstrahlung auf die Kollektorfläche, also die Summe der Variablen (*Direktstrahlung* und *Diffusstrahlung*) in W/m^2 .

- Eglob [W/m^2]:

Die Variable *Eglob* gibt die Tagessumme der Globalstrahlung auf die geneigte Kollektorebene, also die Summe der Variablen *Edir* und *Ediff*, in W/m^2 an.

- Sonne [min]:

Die Variable *Sonne* bezieht sich auf die Tagessumme der Sonnenscheindauer in *Minuten*.

- Lute [$^{\circ}\text{C}$]:

Die Variable *Lute* bezieht sich auf den Tagesmittelwert der Lufttemperatur in $^{\circ}\text{C}$.

- Wige [m/s]:

Die Variable *Wige* bezieht sich auf den Tagesmittelwert der Windgeschwindigkeit in *m/s*.

- Lufe [%]:

Die Variable *Lufe* bezieht sich auf den Tagesmittelwert der relativen Luftfeuchtigkeit in *%*.

3 Explorative Analyse der Zielvariablen WMZ und WMZ/m²

3.1 Tageserträge pro Monat in kWh und kWh/m²

Einen ersten Überblick über die Summen und Verteilungen der Monatserträge in kWh der sechs Solaranlagen, sowie die Anzahl der Tage mit Erträgen pro Monat liefern die Tabelle 3.1 und Abbildung 3.1.

Tabelle 3.1: Summe der Erträge in kWh pro Monat

Stadion Liebenau			Grottenhofstraße			Eggenberger Allee		
<i>WMZ [kWh] 1.242,7 m²</i>			<i>WMZ [kWh] 444,3 m²</i>			<i>WMZ [kWh] 104,6 m²</i>		
Monat	Summe	N	Monat	Summe	N	Monat	Summe	N
Jänner	7.800	25	Jänner	3.929	20	Jänner	3.479	48
Februar	31.500	41	Februar	12.139	24	Februar	1.924	23
März	108.300	70	März	10.064	22	März	5.125	30
April	92.300	40	April	17.009	27	April	6.273	27
Mai	229.300	90	Mai			Mai	7.056	30
Juni	227.900	87	Juni	32.985	37	Juni	5.746	30
Juli	335.100	115	Juli	54.456	57	Juli	5.125	30
August	210.300	92	August	42.391	47	August	5.044	25
September	186.300	89	September	32.106	52	September	6.114	26
Oktober	93.100	81	Oktober	19.124	35	Oktober	4.400	26
November	21.900	51	November	10.185	35	November	1.058	15
Dezember	1.700	11	Dezember	2.174	13	Dezember	1.234	19
Insgesamt	1.545.500	792	Insgesamt	236.562	369	Insgesamt	52.575	329
<i>MW</i>	<i>1.951,4</i>		<i>MW</i>	<i>641,1</i>		<i>MW</i>	<i>159,8</i>	

KAPITEL 3. EXPLORATIVE ANALYSE DER ZIELVARIABLEN WMZ UND WMZ/M²

Schwarzer Weg <i>WMZ [kWh] 84,3 m²</i>			Theodor-Körner-Straße <i>WMZ [kWh] 201,2 m²</i>			Judendorf <i>WMZ [kWh] 87,7 m²</i>		
Monat	Summe	N	Monat	Summe	N	Monat	Summe	N
Jänner	1.785	44	Jänner	3.414	44	Jänner	6.206	127
Februar	966	19	Februar	1.461	19	Februar	10.047	134
März	3.330	24	März	7.002	29	März	17.122	155
April	5.011	24	April	10.321	27	April	16.308	156
Mai	6.423	30	Mai	11.213	30	Mai	28.856	174
Juni	5.517	24	Juni	11.598	30	Juni	27.239	155
Juli	5.633	28	Juli	10.538	29	Juli	30.893	171
August	4.657	28	August	8.938	29	August	25.108	164
September	4.758	25	September	9.087	30	September	23.635	137
Oktober	3.278	26	Oktober	8.091	28	Oktober	12.801	121
November	837	13	November	2.326	14	November	6.014	112
Dezember	957	19	Dezember	1.689	22	Dezember	4.322	93
Insgesamt	43.152	304	Insgesamt	85.677	331	Insgesamt	208.551	1.699
<i>MW</i>	<i>141,9</i>		<i>MW</i>	<i>258,8</i>		<i>MW</i>	<i>122,7</i>	

Aufgrund der unterschiedlich großen Netto-Absorberflächen sind die Werte in Tabelle 3.1 nicht miteinander vergleichbar. Ebenso verhält es sich bei den Graphiken in Abbildung 3.1, bei denen die y-Achsen unterschiedlich skaliert sind. Daher werden diese Diagramme im Folgenden bei der ersten Analyse der Anlagen getrennt betrachtet.

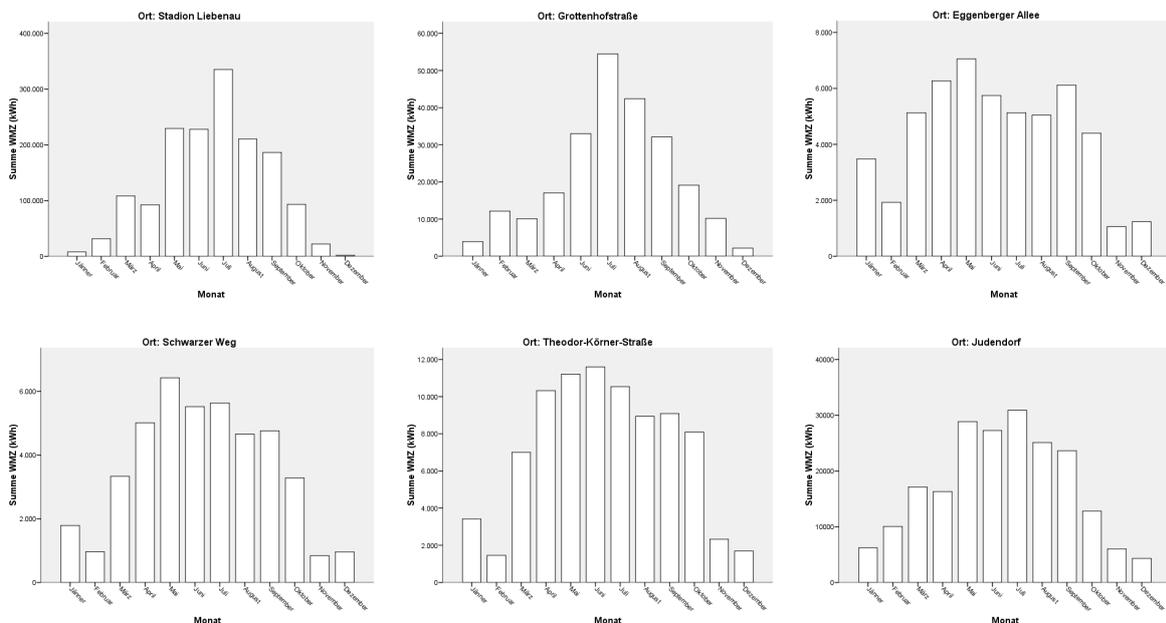


Abbildung 3.1: Säulendiagramme der Ertragssummen in kWh pro Monat

STADION LIEBENAU

Die obere linke Graphik in Abbildung 3.1 auf der vorherigen Seite zeigt, dass es am Dach des Liebenauer Stadions erwartungsgemäß den meisten Ertrag in den Sommermonaten Mai bis September mit dem Höhepunkt im Juli mit 335.100 kWh gab. Die Werte in Tabelle 3.2 zeigen allerdings auch, dass es einige Jahre mit keinem Ertrag oder Aufzeichnungsausfällen ganzer Monate, bzw. mehrerer Tage pro Monat gab, die fast ausschließlich im Winter vorkamen. Die häufigsten Ertragstage hatte das Jahr 2005, gefolgt von den Jahren 2004 und 2008. Wenige Tage mit Ertrag gab es in den Jahren 2002 und 2003.

Tabelle 3.2: Anzahl der Aufzeichnungstage pro Monat & Jahr (Stadion Liebenau)

	J A H R							Gesamt
	2002	2003	2004	2005	2006	2007	2008	
Jänner		2		9		10	4	25
Februar		12	6	13		2	8	41
März			16	21	15	4	14	70
April			10		9	4	17	40
Mai			27	15	21	9	18	90
Juni		12	27	20	9	11	8	87
Juli	7	10	20	21	22	12	23	115
August	16	0	13	25	11	10	17	92
September	19	7	14	16	9	11	13	89
Oktober	11	7	15	22	1	11	14	81
November	6	7	8	7	8	4	11	51
Dezember			3	1	4		3	11
Gesamt	59	57	159	170	109	88	150	792

Tabelle 3.3 enthält die wichtigsten Eckdaten, wie etwa Mittelwerte, Mediane und Standardabweichungen der Tageserträge/m² auf Monatsbasis, die auch für Abbildung 3.2 verwendet werden. Diese Abbildung enthält links die Fehlerbalken (vgl. Seite 82) und rechts die Boxplots (vgl. Seite 81) der Tageserträge pro Monat in kWh/m² für die Solaranlage am Dach des Liebenauer Stadions.

Die Fehlerbalken zeigen die 95%-Konfidenzintervalle um die angegebenen Mittelwerte der Tageserträge/m², während die Boxplots die Verteilungen um die Mediane der Tageserträge/m² abbilden. Die Box enthält jeweils die zentralen 50% der Daten (= interquartiler Bereich).

Tabelle 3.3: Tageserträge pro Monat in kWh/m² (Stadion Liebenau)

	<i>WMZ/m² [kWh/m²]</i>						
	Summe	Mittelwert	Median	N	Standard- abweichung	Min	Max
Jänner	6,28	0,25	0,16	25	0,19	0,08	0,72
Februar	25,35	0,62	0,56	41	0,36	0,08	1,45
März	87,15	1,24	1,13	70	0,79	0,08	2,74
April	74,27	1,86	1,85	40	1,03	0,24	3,46
Mai	184,52	2,05	2,09	90	1,08	0,08	3,86
Juni	183,39	2,11	2,17	87	1,14	0,16	3,86
Juli	269,65	2,34	2,49	115	0,93	0,16	3,78
August	169,23	1,84	2,01	92	1,00	0,08	3,30
September	149,92	1,68	1,69	89	0,77	0,08	2,82
Oktober	74,92	0,92	0,89	81	0,48	0,08	1,93
November	17,62	0,35	0,24	51	0,25	0,08	0,89
Dezember	1,37	0,12	0,08	11	0,06	0,08	0,24
Insgesamt	1.243,66	1,57	1,53	792	1,08	0,08	3,86

Die Vorteile des Boxplots liegen darin, dass man einen schnellen Überblick über Charakteristika wie Median und Streubreite erhält und erkennen kann, ob Schiefe im Zentrum oder außerhalb vorliegt. Er ist ideal einsetzbar beim Vergleich von mehreren Stichproben oder als Zusatzinformation am Rand von anderen Plot-Darstellungen.³

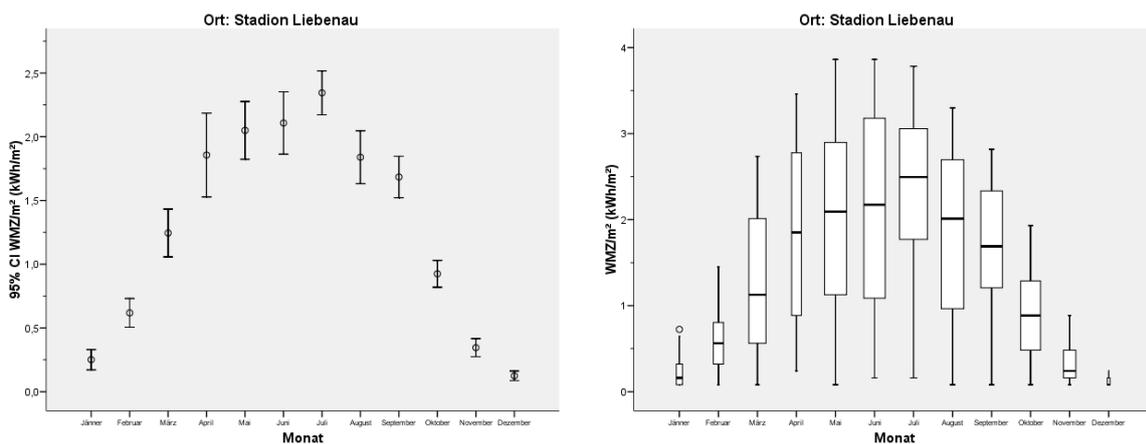


Abbildung 3.2: Fehlerbalken & Boxplots der Tageserträge pro Monat in kWh/m² (Stadion Liebenau)

³[12] Stadlober & Schauer, 2011, S. 34-35

Es ist deutlich erkennbar, dass der mittlere Tagesertrag/m² von Jänner bis in die Monate Mai bis Juli ansteigt und danach bis Dezember wieder abfällt. Dabei liegen die Werte in den Sommermonaten April bis September zwischen 1,7 und 2,3 kWh/m², in den Wintermonaten Oktober bis März zwischen 0,1 und 1,2 kWh/m² deutlich darunter. Die Standardabweichungen weisen denselben Trend wie die Mittelwerte auf: Hohe Werte in den Sommersaisonen, niedrige Werte in den Wintersaisonen. Dieser Trend ist auch bei den anderen fünf Solaranlagen ersichtlich, weswegen in der Folge generell nur auf systematische Abweichungen von diesem Muster näher eingegangen wird.

Der Vergleich der Mittelwerte gegen Mediane zeigt ähnliche Werte mit einem Ausreißer nach oben im Boxplot am 30.01.2007 mit 0,7 kWh/m².

GROTTENHOFSTRAÙE

Die obere mittlere Graphik in Abbildung 3.1 auf Seite 14 zeigt für die Solaranlage in der Grottenhofstraße einen großen Unterschied zwischen den Monaten Juni bis September und den restlichen Monaten. Zum Teil ist das durch das Fehlen von Ertragstagen in den Monaten Jänner bis April 2007 zu erklären (siehe Tabelle 3.4), jedoch zeigt dies auch deutlich den zu erwartenden besseren Ertrag im Sommer.

Tabelle 3.4: Anzahl der Aufzeichnungstage pro Monat & Jahr (Grottenhofstraße)

	J A H R		Gesamt
	2007	2008	
Jänner		20	20
Februar		24	24
März		22	22
April		27	27
Mai			
Juni	23	14	37
Juli	27	30	57
August	18	29	47
September	25	27	52
Oktober	10	25	35
November	19	16	35
Dezember	10	3	13
Gesamt	132	237	369

In Tabelle 3.5 ist erkennbar, dass auch in der Grottenhofstraße die Monate April bis

Tabelle 3.5: Tageserträge pro Monat in kWh/m² (Grottenhofstraße)

	<i>WMZ/m² [kWh/m²]</i>						
	Summe	Mittelwert	Median	N	Standard- abweichung	Min	Max
Jänner	8,84	0,44	0,31	20	0,37	0,00	1,25
Februar	27,32	1,14	1,27	24	0,54	0,00	1,86
März	22,65	1,03	0,78	22	0,66	0,16	2,61
April	38,28	1,42	1,09	27	0,90	0,00	2,98
Mai							
Juni	74,24	2,01	2,26	37	0,90	0,04	3,13
Juli	122,57	2,15	2,33	57	0,84	0,13	3,20
August	95,41	2,03	2,41	47	0,87	0,01	3,09
September	72,26	1,39	1,53	52	0,79	0,00	2,63
Oktober	43,04	1,23	1,40	35	0,67	0,00	2,17
November	22,92	0,65	0,67	35	0,45	0,00	1,66
Dezember	4,89	0,38	0,42	13	0,26	0,00	0,72
Insgesamt	532,44	1,44	1,37	369	0,93	0,00	3,20

September mit 1,4 - 2,2 kWh/m² höhere mittlere Tageserträge/m² liefern als die Wintermonate Oktober bis März mit 0,4 - 1,2 kWh/m².

In Abbildung 3.3 sind rechts im Boxplot vier Ausreißer nach unten an den Tagen 01.02.2008, 03.02.2008, 06.02.2008 und 15.02.2008 zu erkennen. Ansonsten weist der

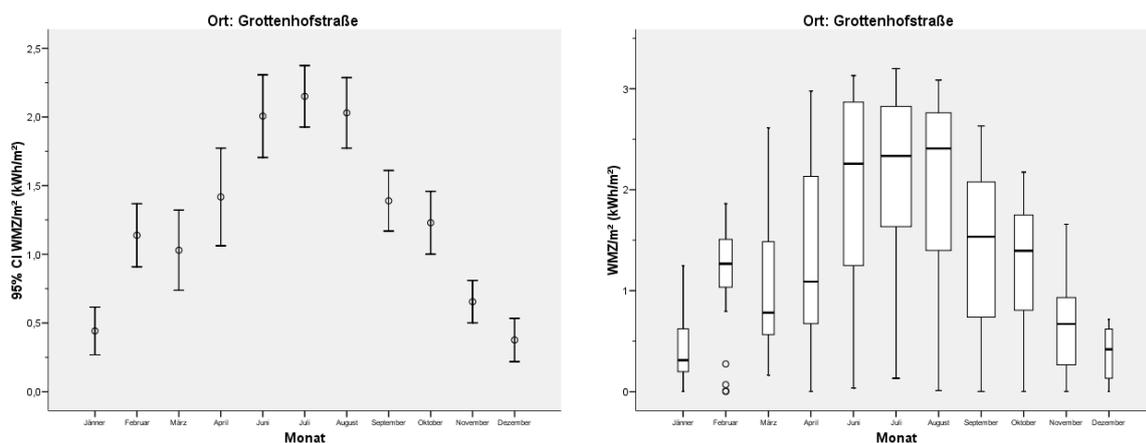


Abbildung 3.3: Fehlerbalken & Boxplots der Tageserträge pro Monat in kWh/m² (Grottenhofstraße)

Februar 2008 höhere mittlere Tageserträge/m² als die Monate Jänner und März 2008 auf, was trotz Berücksichtigung dieser Ausreißer auch in der Graphik der Fehlerbalken ersichtlich ist.

EGGENBERGER ALLEE

Die Solaranlage in der Eggenberger Allee scheint von März bis Oktober ähnlich ertragsreich zu sein, wie die obere rechte Graphik in Abbildung 3.1 auf Seite 14 zeigt. Allerdings gibt es hier nur 13 Aufzeichnungsmonate (Jänner 2005 bis Jänner 2006), siehe Tabelle 3.6. Die Anlage am Dach des Liebenauer Stadions hat schon gezeigt, dass 2005 das Jahr mit den meisten Ertragstagen war.

Tabelle 3.6: Anzahl der Aufzeichnungstage pro Monat & Jahr (Eggenberger Allee)

	J A H R		Gesamt
	2005	2006	
Jänner	27	21	48
Februar	23		23
März	30		30
April	27		27
Mai	30		30
Juni	30		30
Juli	30		30
August	25		25
September	26		26
Oktober	26		26
November	15		15
Dezember	19		19
Gesamt	308	21	329

Der auffällig höhere Ertrag in kWh (siehe Tabelle 3.1 auf Seite 13 rechts) im Jänner beruht darauf, dass es sowohl Aufzeichnungen im Jänner 2005 als auch 2006 gab, was auch die Summe der Tageserträge in kWh/m² in Tabelle 3.7 zeigt.

Die Anlage in der Eggenberger Allee weist einen untypischen Jahresverlauf sowohl im Fehlerbalken, als auch im Boxplot der Abbildung 3.4 auf. Die Monate Juni bis August 2005 brachten mit 1,6 - 1,9 kWh/m² weniger mittlere Tageserträge/m² als die Monate April, Mai und September 2005 mit über 2,2 kWh/m². Auffällig ist auch das ähnlich tiefe Niveau der Monate Jänner, Februar, November und Dezember.

Tabelle 3.7: Tageserträge pro Monat in kWh/m² (Eggenberger Allee)

	<i>WMZ/m² [kWh/m²]</i>						
	Summe	Mittelwert	Median	N	Standard- abweichung	Min	Max
Jänner	33,26	0,69	0,60	48	0,42	0,01	1,45
Februar	18,39	0,80	0,77	23	0,54	0,04	1,83
März	49,00	1,63	1,64	30	0,84	0,09	3,01
April	59,97	2,22	2,59	27	1,14	0,02	3,37
Mai	67,45	2,25	2,20	30	0,86	0,76	3,55
Juni	54,93	1,83	1,91	30	0,42	1,01	2,77
Juli	49,00	1,63	1,77	30	1,04	0,01	3,71
August	48,22	1,93	2,00	25	1,21	0,02	3,61
September	58,45	2,25	2,48	26	1,05	0,02	3,39
Oktober	42,06	1,62	1,77	26	0,77	0,11	3,07
November	10,11	0,67	0,63	15	0,57	0,01	1,67
Dezember	11,79	0,62	0,64	19	0,43	0,07	1,23
Insgesamt	502,63	1,53	1,40	329	1,01	0,01	3,71

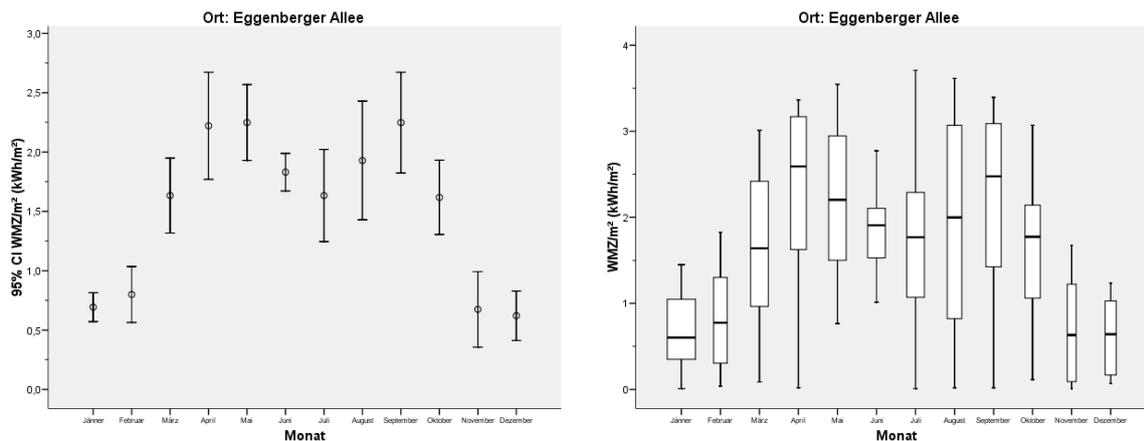


Abbildung 3.4: Fehlerbalken & Boxplots der Tageserträge pro Monat in kWh/m² (Eggenberger Allee)

SCHWARZER WEG

Die untere linke Graphik in Abbildung 3.1 auf Seite 14 für die Solaranlage am Schwarzen Weg zeigt ein ähnliches Bild wie zuvor in der Eggenberger Allee mit guten Erträgen in den Monaten März bis Oktober. Auch hier wurden von Jänner 2005 bis Jänner 2006 Daten zur Analyse zur Verfügung gestellt (siehe Tabelle 3.8), was wieder den höheren

3.1. TAGESERTRÄGE PRO MONAT IN KWH UND KWH/M²

Ertrag in kWh im Jänner erklärt.

Tabelle 3.8: Anzahl der Aufzeichnungstage pro Monat & Jahr (Schwarzen Weg)

	J A H R		Gesamt
	2005	2006	
	Jänner	24	
Februar	19		19
März	24		24
April	24		24
Mai	30		30
Juni	24		24
Juli	28		28
August	28		28
September	25		25
Oktober	26		26
November	13		13
Dezember	19		19
Gesamt	284	20	304

Tabelle 3.9: Tageserträge pro Monat in kWh/m² (Schwarzen Weg)

	<i>WMZ/m² [kWh/m²]</i>						
	Summe	Mittelwert	Median	N	Standard- abweichung	Min	Max
Jänner	21,17	0,48	0,50	44	0,31	0,01	1,27
Februar	11,46	0,60	0,68	19	0,36	0,09	1,35
März	39,50	1,65	1,64	24	0,85	0,01	2,88
April	59,44	2,48	2,84	24	0,99	0,18	3,56
Mai	76,19	2,54	2,95	30	1,21	0,49	4,10
Juni	65,44	2,73	2,95	24	0,97	0,02	3,75
Juli	66,82	2,39	2,49	28	1,02	0,05	3,83
August	55,24	1,97	2,18	28	1,16	0,07	3,52
September	56,44	2,26	2,73	25	0,96	0,11	3,38
Oktober	38,88	1,50	1,58	26	0,76	0,07	2,87
November	9,93	0,76	0,81	13	0,40	0,12	1,47
Dezember	11,35	0,60	0,62	19	0,40	0,01	1,23
Insgesamt	511,89	1,68	1,47	304	1,18	0,01	4,10

Die Solaranlage am Schwarzen Weg zeigt wieder, dass die Monate April bis September mit 2 - 2,7 kWh/m² höhere mittlere Tageserträge/m² liefern als die Wintermonate Oktober bis März mit 0,5 - 1,6 kWh/m² (siehe Tabelle 3.9). Allerdings gibt es einen

leichten Ertragsabfall in den Monaten Juli und August 2005. Die Boxplots in Abbildung 3.5 zeigen einen Ausreißer nach unten am 06.06.2005.

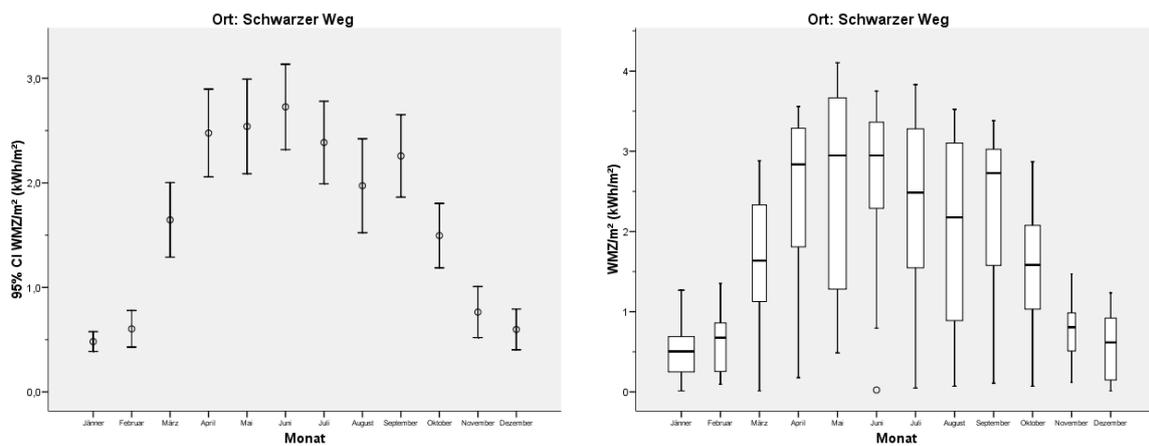


Abbildung 3.5: Fehlerbalken & Boxplots der Tageserträge pro Monat in kWh/m² (Schwarzen Weg)

THEODOR-KÖRNER-STRAÙE

Ein ähnliches Ergebnis wie in den Solaranlagen in der Eggenberger Allee und am Schwarzen Weg liefert die Solaranlage in der Theodor-Körner-StraÙe mit den besten Er-

Tabelle 3.10: Anzahl der Aufzeichnungstage pro Monat & Jahr (Theodor-Körner-Str.)

	J A H R		Gesamt
	2005	2006	
Jänner	23	21	44
Februar	19		19
März	29		29
April	27		27
Mai	30		30
Juni	30		30
Juli	29		29
August	29		29
September	30		30
Oktober	28		28
November	14		14
Dezember	22		22
Gesamt	310	21	331

3.1. TAGESERTRÄGE PRO MONAT IN KWH UND KWH/M²

tragsmonaten März bis Oktober 2005 (siehe untere mittlere Graphik in Abbildung 3.1 auf Seite 14). Auch hier passt der Jänner zu den Monaten Februar, November und Dezember, wenn man bedenkt, dass es 23 Ertragstage im Jänner 2005 und 21 im Jahr 2006 gab, siehe Tabelle 3.10.

Tabelle 3.11: Tageserträge pro Monat in kWh/m² (Theodor-Körner-Straße)

	Summe	Mittelwert	Median	N	Standard- abweichung	Min	Max
Jänner	16,97	0,39	0,37	44	0,26	0,01	0,98
Februar	7,26	0,38	0,28	19	0,36	0,00	1,15
März	34,80	1,20	1,20	29	0,73	0,04	2,43
April	51,30	1,90	2,49	27	1,09	0,00	3,13
Mai	55,73	1,86	2,02	30	0,93	0,34	3,26
Juni	57,64	1,92	1,99	30	0,76	0,72	3,46
Juli	52,37	1,81	1,83	29	0,91	0,02	3,19
August	44,42	1,53	1,56	29	0,99	0,01	3,20
September	45,16	1,51	1,42	30	1,06	0,00	2,99
Oktober	40,21	1,44	1,60	28	0,85	0,00	2,76
November	11,56	0,83	0,90	14	0,57	0,01	1,67
Dezember	8,39	0,38	0,37	22	0,29	0,00	0,78
Insgesamt	425,83	1,29	1,10	331	0,98	0,00	3,46

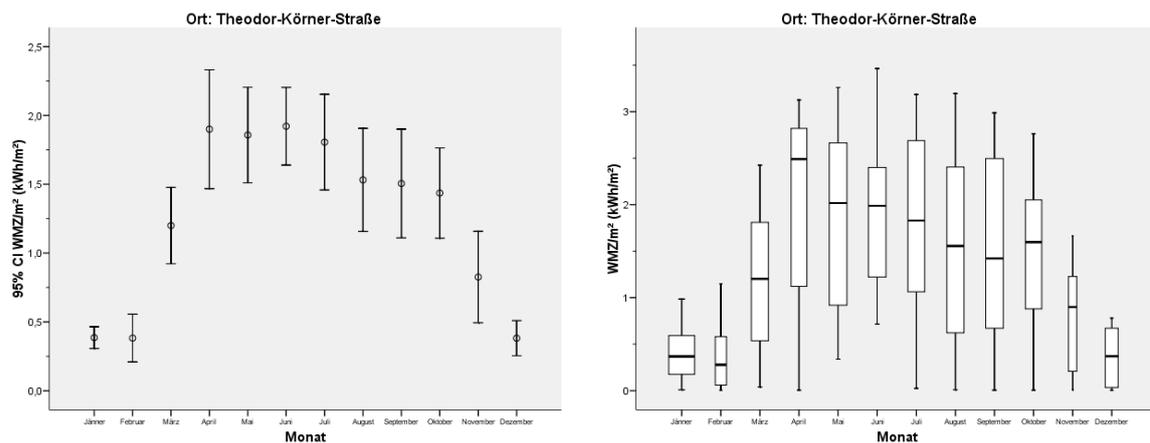


Abbildung 3.6: Fehlerbalken & Boxplots der Tageserträge pro Monat in kWh/m² (Theodor-Körner-Straße)

Auch in der Theodor-Körner-Straße ist deutlich erkennbar, dass die Monate April bis

September mit 1,5 - 1,9 kWh/m² höhere mittlere Tageserträge/m² liefern als die Wintermonate Oktober bis März mit 0,4 - 1,4 kWh/m². Diese Solaranlage weist allerdings höhere Mediane der Tageserträge/m² in den Monaten April und Oktober 2005 auf als in den entsprechenden Vor- und Nachmonaten.

JUDENDORF

Für die Analyse der Solaranlage in Judendorf standen mit 1.699 Ertragstagen in den Jahren 2002 bis 2008 die meisten Aufzeichnungen zur Verfügung (siehe Tabelle 3.12). Mit Ausnahme der Jahre 2007 (Jänner bis September) und 2008 (Oktober bis Dezember) sind die Aufzeichnungen vollständig, und fasst man diese beiden Jahre als ein Jahr zusammen, dann stehen insgesamt sechs Jahre zur Analyse zur Verfügung. Die untere rechte Graphik in Abbildung 3.1 auf Seite 14 zeigt den Anstieg der Summe der Erträge in kWh von Jänner bis in die Sommermonate und danach den Abfall bis in den Dezember.

Tabelle 3.12: Anzahl der Aufzeichnungstage pro Monat & Jahr (Judendorf)

	J A H R							Gesamt
	2002	2003	2004	2005	2006	2007	2008	
Jänner	21	17	23	26	20		20	127
Februar	23	23	21	22	22		23	134
März	26	29	22	28	24		26	155
April	21	30	22	28	26		29	156
Mai	27	30	29	30	28		30	174
Juni	29	28	30	26	28		14	155
Juli	30	31	29	19	31		31	171
August	30	30	29	23	23		29	164
September	27	28	29	25	10		18	137
Oktober	24	26	22	25	13	11		121
November	20	16	15	15	25	21		112
Dezember	9	19	14	14	18	19		93
Gesamt	287	307	285	281	268	51	220	1.699

Die Solaranlage in Judendorf zeigt, dass die Monate Mai bis September mit 1,7 - 2,1 kWh/m² höhere mittlere Tageserträge/m² liefern als die Wintermonate Oktober bis April mit 0,5 - 1,3 kWh/m². Hier gibt es einen Ertragsabfall in den Monaten April und August, die sowohl im Mittelwert, als auch im Median niedrigere Tageserträge/m² aufweisen als die entsprechenden Vor- und Nachmonate.

3.1. TAGESERTRÄGE PRO MONAT IN KWH UND KWH/M²

Tabelle 3.13: Tageserträge pro Monat in kWh/m² (Judendorf)

	<i>WMZ/m² [kWh/m²]</i>						
	Summe	Mittelwert	Median	N	Standard- abweichung	Min	Max
Jänner	70,76	0,56	0,49	127	0,39	0,01	1,52
Februar	114,56	0,85	0,83	134	0,57	0,01	2,28
März	195,23	1,26	1,25	155	0,75	0,02	2,84
April	185,95	1,19	1,20	156	0,86	0,01	3,23
Mai	329,03	1,89	1,94	174	0,91	0,01	3,48
Juni	310,59	2,00	2,11	155	0,91	0,01	3,42
Juli	352,26	2,06	2,22	171	0,83	0,01	3,27
August	286,29	1,75	1,89	164	0,90	0,02	3,38
September	269,50	1,97	2,19	137	0,93	0,01	3,17
Oktober	145,96	1,21	1,19	121	0,69	0,01	2,82
November	68,57	0,61	0,47	112	0,53	0,01	2,03
Dezember	49,28	0,53	0,60	93	0,33	0,01	1,05
Insgesamt	2.378,00	1,40	1,28	1.699	0,95	0,01	3,48

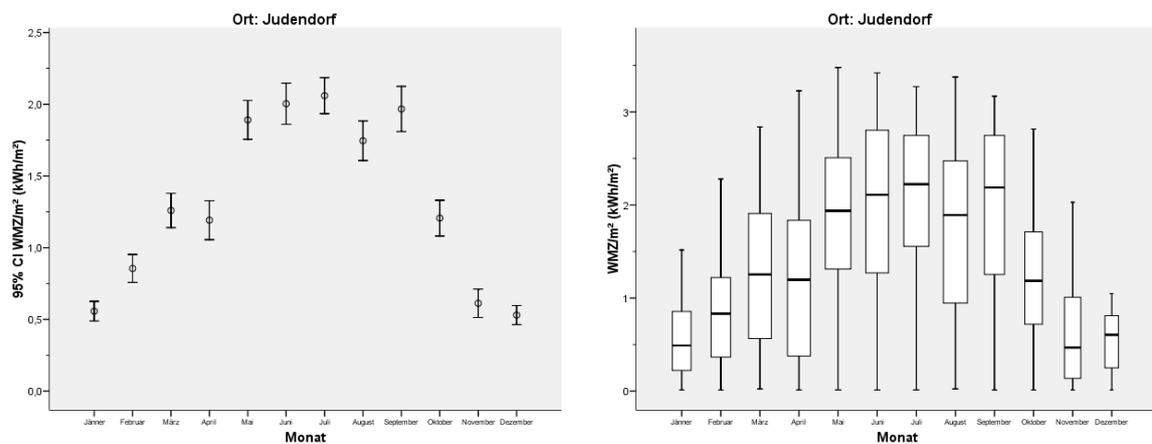


Abbildung 3.7: Fehlerbalken & Boxplots der Tageserträge pro Monat in kWh/m² (Judendorf)

3.2 Vergleich der sechs Grazer Solaranlagen

Tabelle 3.14 zeigt, dass im Mittel die Solaranlage am **Schwarzen Weg** den höchsten Tagesertrag mit 1,7 kWh/m² erzeugt, gefolgt von den Solaranlagen am Dach des **Liebenauer Stadions** (1,6 kWh/m²) und in der **Eggenberger Allee** (1,5 kWh/m²). Den niedrigsten mittleren Tagesertrag liefert die Solaranlage in der **Theodor-Körner-Straße** mit nur 1,3 kWh/m². Den höchsten Tagesertrag liefert auch die Solaranlage am Schwarzen Weg mit 4,1 kWh/m². Insgesamt ist ein mittlerer Tagesertrag von 1,46 kWh/m² für alle Solaranlagen in und um Graz zu erwarten.

Tabelle 3.14: Tageserträge pro m²

	WMZ/m ² [kWh/m ²]						
	Summe	Mittelwert	Median	N	Standard- abweichung	Min	Max
Stadion Liebenau	1.243,66	1,57	1,53	792	1,08	0,08	3,86
Grottenhof- straße	532,44	1,44	1,37	369	0,93	0,00	3,20
Eggenberger Allee	502,63	1,53	1,40	329	1,01	0,01	3,71
Schwarzer Weg	511,89	1,68	1,47	304	1,18	0,01	4,10
Theodor- Körner- Straße	425,83	1,29	1,10	331	0,98	0,00	3,46
Judendorf	2.378,00	1,40	1,28	1.699	0,95	0,01	3,48
Insgesamt	5.594,45	1,46	1,35	3.824	1,01	0,00	4,10

Einen ersten Überblick über die Verteilung der **Tageserträge/m²** pro Solaranlage liefern die Fehlerbalken und Boxplots in Abbildung 3.8. Hier ist besonders hervorzuheben, dass die generelle Struktur der Verteilungen der unterschiedlichen Solaranlagen, wie sie in den Boxplots gezeigt werden, sehr ähnlich ist.

Tabelle 3.15 zeigt die Erträge pro m² pro Viertelstunde, aufgespalten nach den sechs Solaranlagen. Dabei liefert im Mittel die Solaranlage am Dach des **Liebenauer Stadions** den meisten Ertrag pro eingeschalteter Viertelstunde (onoff=1) mit 0,08 kWh/m² auf. Die restlichen Solaranlagen weisen annähernd gleiche Werte von 0,05 - 0,06 kWh/m².

3.2. VERGLEICH DER SECHS GRAZER SOLARANLAGEN

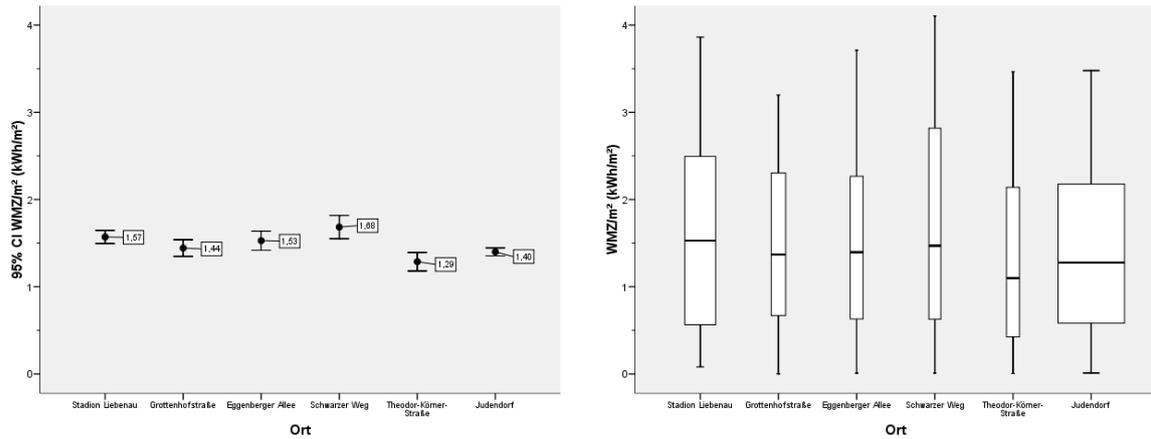


Abbildung 3.8: Fehlerbalken & Boxplots der Tageserträge pro m²

Den höchsten Ertrag liefert wiederum die Solaranlage am Dach des Liebenauer Stadions mit 0,38 kWh/m² pro Viertelstunde, gefolgt von der Solaranlage in der Grottenhofstraße mit 0,24 kWh/m².

Tabelle 3.15: Erträge pro m² & Viertelstunde

$$WMZ/(on * m^2) [kWh/(1/4h * m^2)]$$

	Summe	Mittelwert	Median	N	Standard- abweichung	Min	Max
Stadion Liebenau	62,06	0,08	0,08	792	0,05	0,00	0,38
Grottenhofstraße	17,92	0,05	0,05	368	0,02	0,00	0,24
Eggenberger Allee	18,09	0,05	0,05	329	0,02	0,01	0,10
Schwarzer Weg	19,25	0,06	0,06	304	0,02	0,01	0,13
Theodor-Körner-Straße	15,33	0,05	0,05	331	0,02	0,00	0,09
Judendorf	106,57	0,06	0,07	1.699	0,02	0,00	0,14
Insgesamt	239,21	0,06	0,06	3.823	0,03	0,00	0,38

Tabelle 3.16 stellt die Anzahl der Viertelstunden der Solaranlagen dar, in denen sie zumindest einmal in Betrieb waren. Dabei erkennt man, dass die Anlage in der **Grottenhofstraße** im Mittel mit 26 Viertelstunden (= 6,5 Stunden) pro Tag am häufigsten

in Betrieb war und Ertrag lieferte. Im Mittel am seltensten lief die Anlage am Dach des **Liebenauer Stadions** mit nur 19 Viertelstunden (= 4,75 Stunden).

Tabelle 3.16: Eingeschaltete Viertelstunden pro Tag (onoff=1)

	<i>1/4h-on [Anzahl]</i>						
	Summe	Mittelwert	Median	N	Standard- abweichung	Min	Max
Stadion Liebenau	15.057	19,01	19	792	7,97	1	37
Grottenhofstraße	9.659	26,18	28	369	10,68	1	43
Eggenberger Allee	7.999	24,31	26	329	10,99	1	43
Schwarzer Weg	7.183	23,63	26	304	11,73	1	41
Theodor-Körner-Straße	7.802	23,57	25	331	12,91	1	60
Judendorf	33.961	19,99	20	1.699	10,26	1	44
Insgesamt	81.661	21,35	22	3.824	10,60	1	60

Im Zuge der weiteren Analyse der Daten wurden an insgesamt 112 Tagen extrem ungewöhnliche Werte bei den berechneten Variablen *eta0_dir*, *eta0_diff*, *c1* und *c2* festgestellt. Tabelle 3.17 zeigt die Anzahl der Tage mit Ausreißern pro Solaranlage, die für die statistischen Analysen der folgenden Kapitel nicht berücksichtigt wurden.

Tabelle 3.17: Anzahl der Tage mit Ausreißern

	Aufzeichnungstage	Ausreißer	Analysstage
Stadion Liebenau	792	17	775 Tage
Grottenhofstraße	369	19	350 Tage
Eggenberger Allee	329	12	317 Tage
Schwarzer Weg	304	7	297 Tage
Theodor-Körner-Straße	331	18	313 Tage
Judendorf	1.699	39	1.660 Tage
Gesamt	3.824	112	3.712 Tage

Die verbliebenen Extremwerte (z.B. bei Ediff [W/m²] im nächsten Kapitel) haben keinen relevanten Einfluss auf die Parameterschätzer und wurden daher nicht entfernt.

4 Einfluss der Regressoren auf die Zielvariable WMZ/m²

4.1 Korrelationen und lineare Regressionen der Tageswerte

Der Korrelationskoeffizient $r_{X,Y}$ von Pearson (vgl. Seite 80) ist ein Maß für den *Grad des linearen Zusammenhangs* zwischen zwei *metrischen* Merkmalen X und Y . Er kann Werte zwischen -1 (Gegenlauf) und $+1$ (Gleichlauf) annehmen. Bei einem Wert von $+1$ (bzw. -1) besteht ein vollständig positiver (bzw. negativer) linearer Zusammenhang zwischen den betrachteten Merkmalen. Ist $r = 0$, so bezeichnet man die Merkmale als unkorreliert.

Als grobe *Faustregel* für die folgenden Interpretationen kann angegeben werden:⁴

Falls $0 \leq |r| < 0,5$ gilt, dann spricht man von einer *schwachen Korrelation*.

Falls $0,5 \leq |r| < 0,8$ gilt, dann spricht man von einer *mittleren Korrelation*.

Falls $0,8 \leq |r| \leq 1$ gilt, dann spricht man von einer **starken Korrelation**.

Die Variablen in den Korrelationstabellen der einzelnen Solaranlagen auf den nächsten Seiten sind nach der Prüfgröße T (vgl. Seite 98f) der jeweiligen Regressionstabellen von links nach rechts gereiht. Die Werte für die Korrelation $|r|$ lassen sich erst im Zusammenhang mit **Streudiagrammen** (vgl. Seite 83) richtig interpretieren. Die Abbildungen 4.1 - 4.7 zeigen die **Scatterplots** der einflussreichsten Regressoren gegen die Tageserträge in kWh/m² für jede Solaranlage.

⁴[12] Stadlober & Schauer, 2011, S. 71-73

STADION LIEBENAU

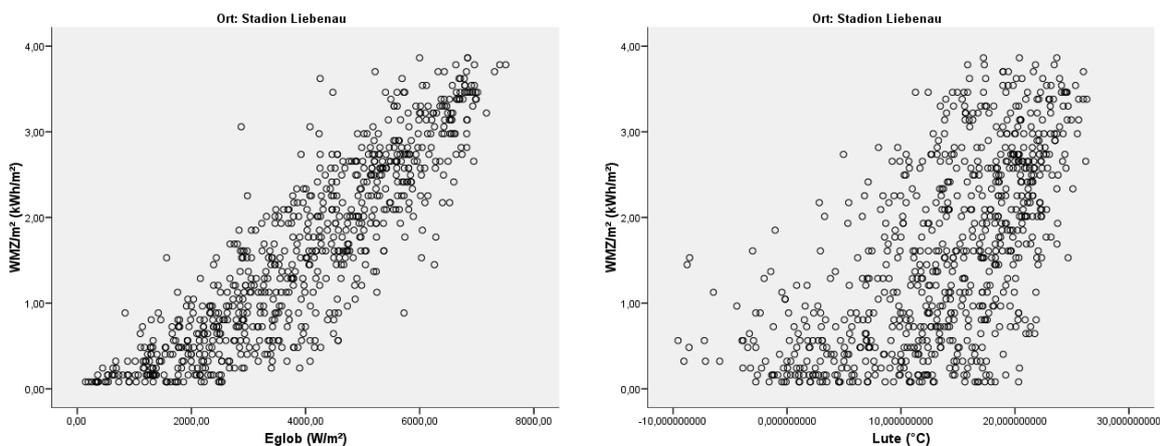
Tabelle 4.1 und Abbildung 4.1 zeigen, dass bei der Solaranlage am Dach des Liebenauer Stadions die Variablen **Eglob** (= Globalstrahlung geneigt) und **Sonne** (= Sonnenscheindauer) **starke Korrelationen** zum **Tagesertrag/m²** besitzen. Immerhin noch **mittlere Zusammenhänge** weisen die Variablen **Lute** (= Lufttemperatur), **Saison** (0=Winter / 1=Sommer) und **1/4h-on** (= Anzahl der Viertelstunden, die Ertrag liefern) auf. Die Variablen **Ediff** (= Diffusstrahlung geneigt), **c1** (= berechneter zeitabhängiger Kollektorkennwert) und **eta0_dir** (= berechneter zeitabhängiger Kollektorkennwert) zeigen eine eher *schwache Korrelation* zum Tagesertrag/m².

Tabelle 4.1: Korrelationen der Anlage Stadion Liebenau

	WMZ/m ² [kWh/m ²]							
	Eglob [W/m ²]	Lute [°C]	Saison	Ediff [W/m ²]	Sonne [min]	c1	1/4h-on	eta0_dir
Korrelation	,884**	,614**	,555**	,444**	,805**	,452**	,711**	,174**
Signifikanz	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
N	754	775	775	775	775	775	775	775

** Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

Bei der Solaranlage am Dach des Liebenauer Stadions wird der Tagesertrag/m² zu **91%** vom Modell mit **acht Regressoren** erklärt (siehe Tabelle 4.2). Den stärksten Einfluss haben die **Globalstrahlung geneigt** und die **Lufttemperatur** mit den T-Werten 21, 73 und 15, 86, gefolgt von der Saison, der Diffusstrahlung geneigt, der Sonnenscheindauer, dem berechneten Kollektorkennwert c1 und der Anzahl der Viertelstunden, an denen die Anlage in Betrieb ist (7, 29 ≥ |T| ≥ 2, 46). Den geringsten Einfluss hat mit



4.1. KORRELATIONEN UND LINEARE REGRESSIONEN DER TAGESWERTE

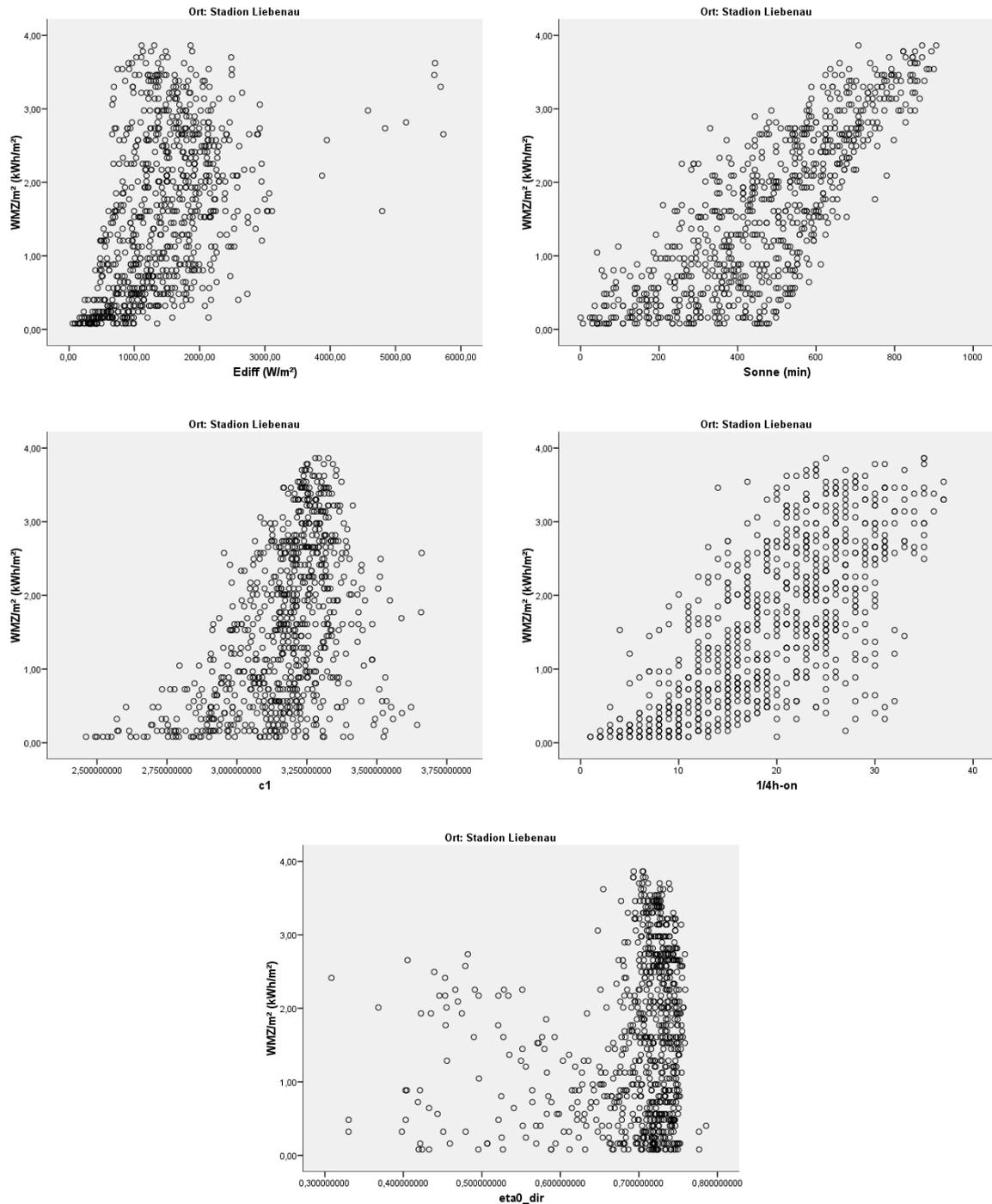


Abbildung 4.1: Streudiagramme der Regressoren mit WMZ/m² (Stadion Liebenau)

$T = -1,83$ und einer Signifikanz von $0,068$ der berechnete Kollektorkennwert eta0_dir , der aber dennoch eine kleine Verbesserung des Modells bewirkt.

Tabelle 4.2: Lineare Regression (Stadion Liebenau)

Modellzusammenfassung^(b,c)

R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers	Durbin-Watson-Statistik
,955(a)	0,912	0,911	0,318	1,583

a. Einflussvariablen: (Konstante), eta0_dir, 1/4h-on, Lute [°C], c1, Ediff [W/m²], Sonne [min], Saison, Eglob [W/m²]
 b. Abhängige Variable: WMZ/m² [kWh/m²]; n = 754
 c. Ort = Stadion Liebenau

Koeffizienten^(a)

	Nicht standardisierte Koeffizienten		Standardisierte Koeff.		95%-Konfidenzintervall für B	
	B	Standardfehler	Beta	T	Untergrenze	Obergrenze
(Konstante)	0,20027	0,286		0,70	0,484	
Eglob [W/m ²]	0,00046	0,000	0,754	21,73	< 10 ⁻³	0,00042 0,00050
Lute [°C]	0,04078	0,003	0,282	15,86	< 10 ⁻³	0,03574 0,04583
Saison	0,30641	0,042	0,138	7,29	< 10 ⁻³	0,22388 0,38894
Ediff [W/m ²]	-0,00014	0,000	-0,086	-5,67	< 10 ⁻³	-0,00019 -0,00009
Sonne [min]	0,00059	0,000	0,110	4,45	< 10 ⁻³	0,00033 0,00085
c1	-0,27941	0,092	-0,046	-3,04	0,002	-0,45999 -0,09883
1/4h-on	-0,00704	0,003	-0,051	-2,46	0,014	-0,01265 -0,00142
eta0_dir	-0,31710	0,174	-0,023	-1,83	0,068	-0,65782 0,02361

a. Abhängige Variable: WMZ/m² [kWh/m²]
 b. Ort = Stadion Liebenau

GROTTENHOFSTRAÙE

Die Solaranlage in der Grottenhofstraße zeigt **starke Zusammenhänge** des **Tagesertrages/m²** zu den Variablen **Eglob**, **c2** (=berechneter zeitabhängiger Kollektor-

Tabelle 4.3: Korrelationen der Anlage Grottenhof

WMZ/m² [kWh/m²]

	Eglob [W/m ²]	c2	1/4h-on	dT1 [°C]	Sonne [min]	Saison
Korrelation	,962**	-,822**	,876**	,847**	,877**	,505**
Signifikanz	0,000	0,000	0,000	0,000	0,000	0,000
N	341	350	350	350	350	350

** Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

4.1. KORRELATIONEN UND LINEARE REGRESSIONEN DER TAGESWERTE

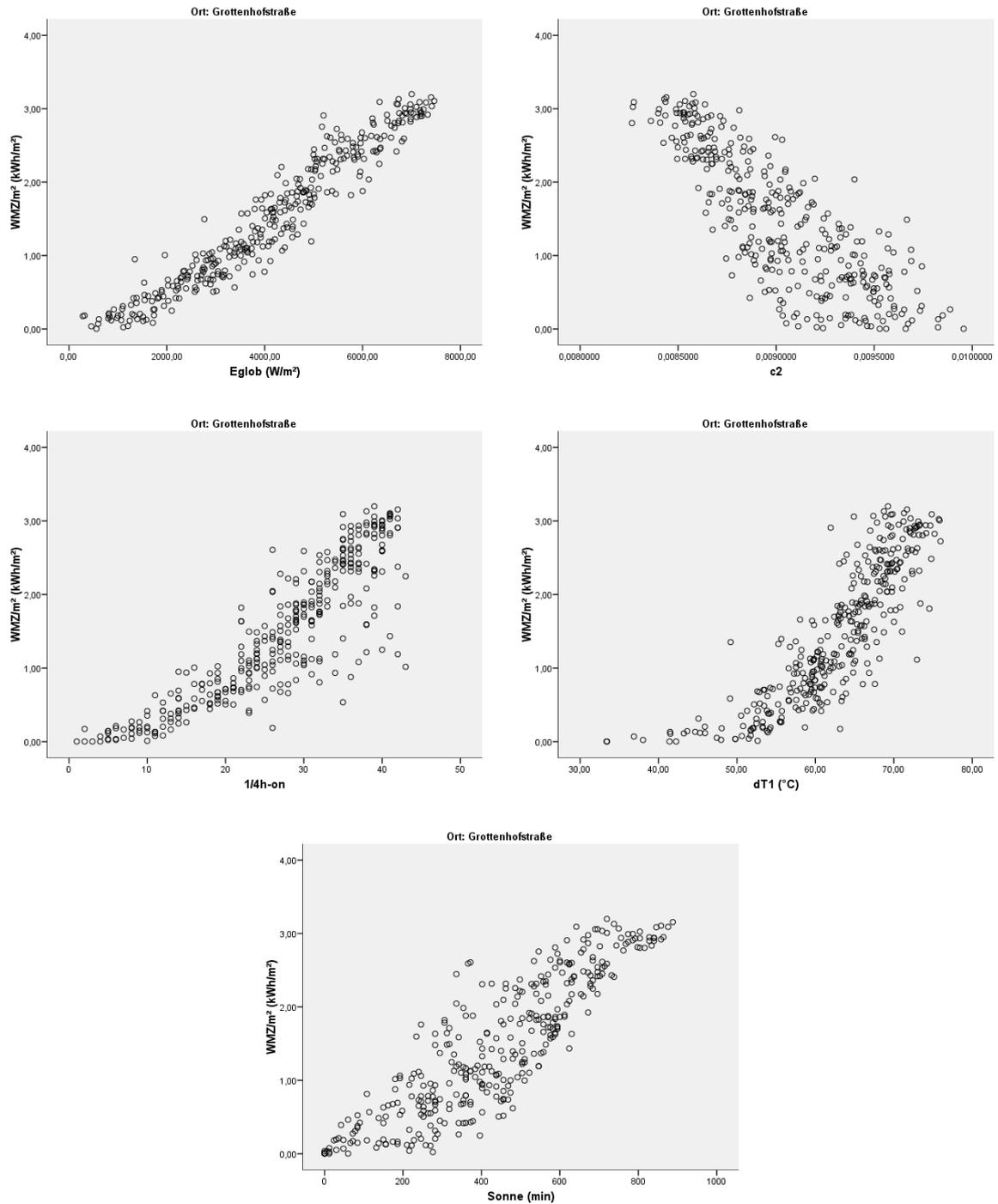


Abbildung 4.2: Streudiagramme der Regressoren mit WMZ/m² (Grottenhof)

kennwert), **1/4h-on**, **dT1** (= Mittelwert zwischen primärseitigen Vor- und Rücklauf-temperaturen) und **Sonne**. Eine mittlere Korrelation weist die Variable **Saison** auf, wie

aus Tabelle 4.3 und Abbildung 4.2 hervorgeht.

Zu **95%** wird bei der Solaranlage in der Grottenhofstraße der Tagesertrag/m² vom Modell mit **sechs Regressoren** erklärt (siehe Tabelle 4.4). Den stärksten Einfluss haben die **Globalstrahlung geneigt** und der berechnete Kollektorkennwert **c2** mit den T-Werten 16, 22 und $-8,75$, gefolgt von der Anzahl der Viertelstunden, an denen die Anlage in Betrieb ist, dem Mittelwert zwischen primärseitigen Vor- und Rücklauftemperaturen ($dT1$), der Sonnenscheindauer und der Saison ($8,75 \geq |T| \geq 2,82$).

Tabelle 4.4: Lineare Regression (Grottenhof)

Modellzusammenfassung ^(b,c)				
R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers	Durbin-Watson-Statistik
,975(a)	0,951	0,950	0,201	1,485

a. Einflussvariablen: (Konstante), Saison, Sonne [min], $dT1$ [°C], c2, 1/4h-on, Eglob [W/m²]

b. Abhängige Variable: WMZ/m² [kWh/m²]; n = 341

c. Ort = Grottenhofstraße

Koeffizienten ^(a)							
	Nicht standardisierte Koeffizienten	Standardisierte Koeff.	95%-Konfidenzintervall für B				
	B	Standardfehler	Beta	T	Signifikanz	Untergrenze	Obergrenze
(Konstante)	3,81203	0,600		6,35	$< 10^{-3}$		
Eglob [W/m ²]	0,00037	0,000	0,734	16,22	$< 10^{-3}$	0,00032	0,00041
c2	-490,94581	56,136	-0,199	-8,75	$< 10^{-3}$	-601,37096	-380,52067
1/4h-on	-0,01142	0,003	-0,124	-3,98	$< 10^{-3}$	-0,01707	-0,00577
$dT1$ [°C]	0,01078	0,003	0,087	3,79	$< 10^{-3}$	0,00519	0,01638
Sonne [min]	0,00045	0,000	0,102	3,39	0,001	0,00019	0,00071
Saison	0,09228	0,033	0,050	2,82	0,005	0,02783	0,15673

a. Abhängige Variable: WMZ/m² [kWh/m²]

b. Ort = Grottenhofstraße

EGGENBERGER ALLEE

Die Variablen **1/4h-on** und **Edir** (= Direktstrahlung geneigt) zeigen laut Tabelle 4.5 und Abbildung 4.3 bei der Solaranlage in der Eggenberger Allee **starke Zusammen-**

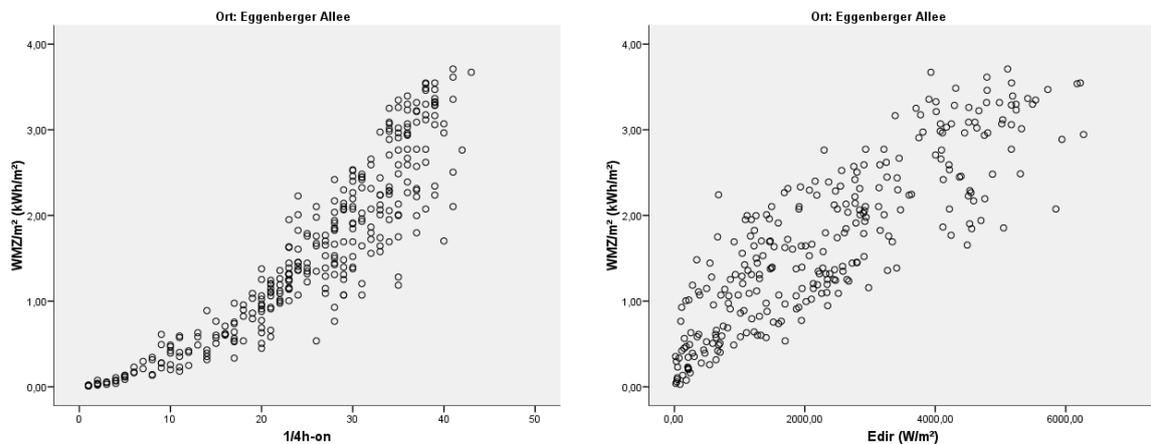
hänge zum Tagesertrag/m². Mittlere Korrelationen weisen die Variablen Sonne und Ediff auf, *schwache Zusammenhänge* gibt es mit den Variablen *eta0_dir*, *Wige* (= Windgeschwindigkeit) und *dT2* (= Temperaturdifferenz zwischen der Außentemperatur zum Zeitpunkt des Sonnenaufgangs und der Variable „dT1“ zum Zeitpunkt des letzten Ausschaltens der Solaranlage).

Tabelle 4.5: Korrelationen der Anlage Eggenberger Allee

	WMZ/m ² [kWh/m ²]						
	1/4h-on	Edir [W/m ²]	eta0_dir	Sonne [min]	Wige [m/s]	dT2 [°C]	Ediff [W/m ²]
Korrelation	,907**	,854**	,236**	,788**	,199**	-,236**	,535**
Signifikanz	0,000	0,000	0,000	0,000	0,000	0,000	0,000
N	317	287	317	317	317	288	317

** Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

Das Modell für die Solaranlage in der Eggenberger Allee erklärt mit **sieben Regressoren** fast **93%** des Tagesertrages/m² (siehe Tabelle 4.6). Den stärksten Einfluss haben die **Anzahl der Viertelstunden**, an denen die Anlage in Betrieb ist, die **Direktstrahlung geneigt** und der berechnete Kollektorkennwert **eta0_dir** mit den T-Werten 11,82, 9,8 und 8,03. Ebenfalls signifikant ins Modell gehen die Sonnenscheindauer, die Windgeschwindigkeit und die Temperaturdifferenz zwischen der Außentemperatur zum Zeitpunkt des Sonnenaufgangs und der Variable „dT1“ zum Zeitpunkt des letzten Ausschaltens der Solaranlage (dT2) ein (3,71 ≥ |T| ≥ 2,75). Die Diffusstrahlung geneigt hat mit T = -2,06 in diesem Modell den geringsten Einfluss.



KAPITEL 4. EINFLUSS DER REGRESSOREN AUF DIE ZIELVARIABLE WMZ/M²

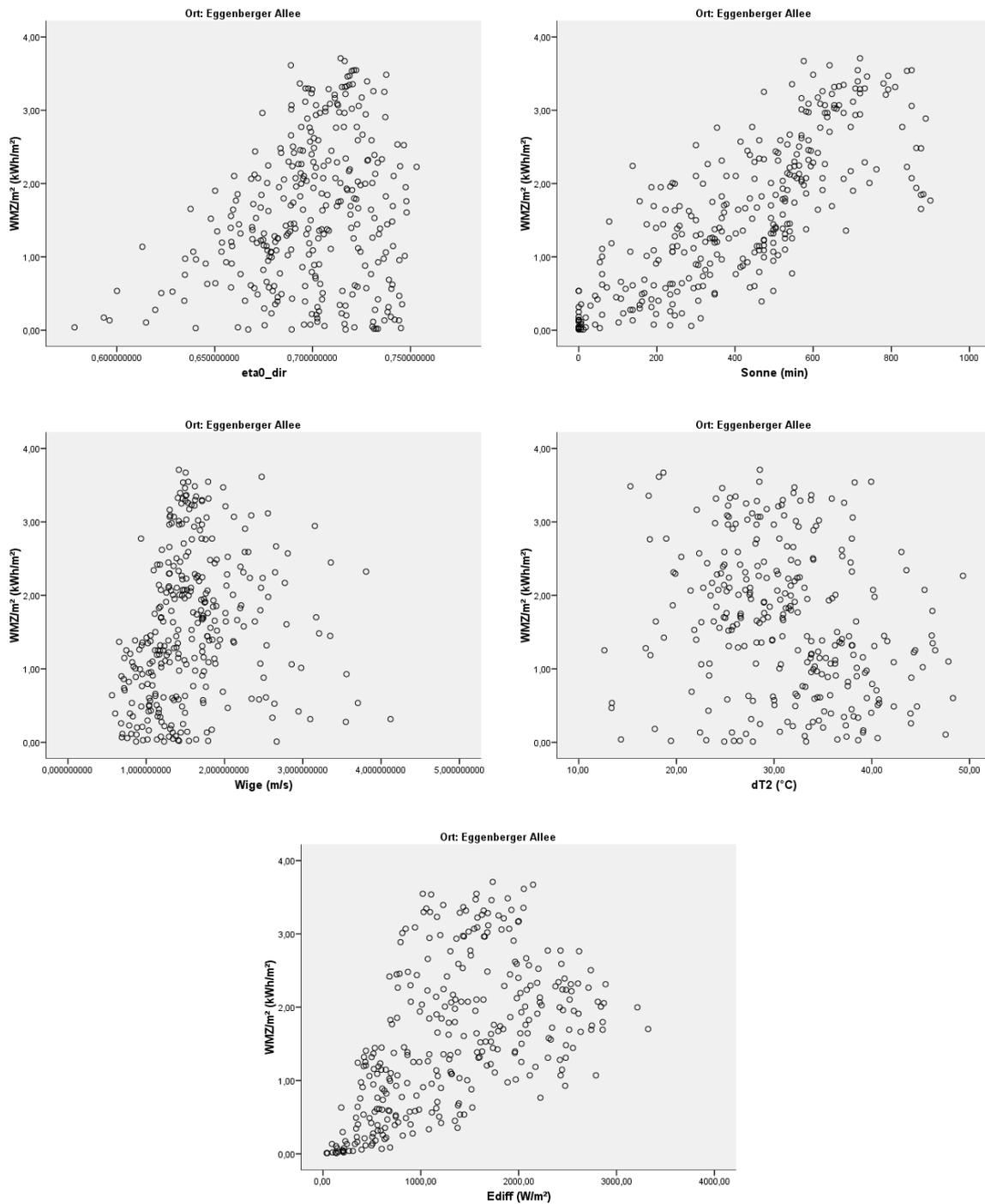


Abbildung 4.3: Streudiagramme der Regressoren mit WMZ/m² (Eggenberg)

Tabelle 4.6: Lineare Regression (Eggenberg)

Modellzusammenfassung^(b,c)

R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers	Durbin-Watson-Statistik
,964(a)	0,930	0,928	0,255	1,677

a. Einflussvariablen: (Konstante), Ediff [W/m²], Edir [W/m²], dT2 [°C], eta0_dir, Wige [m/s], Sonne [min], 1/4h-on
 b. Abhängige Variable: WMZ/m² [kWh/m²]; n = 266
 c. Ort = Eggenberger Allee

Koeffizienten^(a)

	Nicht standardisierte Koeffizienten		Standardisierte Koeff.		95%-Konfidenzintervall für B	
	B	Standardfehler	Beta	T	Untergrenze	Obergrenze
(Konstante)	-3,65798	0,481		-7,61	< 10 ⁻³	
1/4h-on	0,06300	0,005	0,622	11,82	< 10 ⁻³	0,05250 0,07349
Edir [W/m ²]	0,00032	0,000	0,558	9,80	< 10 ⁻³	0,00026 0,00039
eta0_dir	5,40299	0,673	0,155	8,03	< 10 ⁻³	4,07869 6,72729
Sonne [min]	-0,00071	0,000	-0,159	-3,71	< 10 ⁻³	-0,00109 -0,00033
Wige [m/s]	-0,09674	0,032	-0,058	-3,07	0,002	-0,15889 -0,03460
dT2 [°C]	-0,00730	0,003	-0,054	-2,75	0,006	-0,01253 -0,00207
Ediff [W/m ²]	-0,00011	0,000	-0,085	-2,06	0,041	-0,00021 0,00000

a. Abhängige Variable: WMZ/m² [kWh/m²]
 b. Ort = Eggenberger Allee

SCHWARZER WEG

Tabelle 4.7 und Abbildung 4.4 zeigen, dass bei der Anlage am Schwarzen Weg die Variablen **Edir**, **1/4h-on** und **Eglob** **starke Korrelationen** zum **Tagesertrag/m²** aufweisen, während die Variablen **c1**, **eta0_diff** (= berechneter zeitabhängiger Kollektorkennwert) und **Lute** **mittlere Zusammenhänge** zeigen. Nur eine *schwache Korrelation* besteht zu den Variablen **eta0_dir**, **Lufe** (= Luftfeuchtigkeit), **dT1** und **Wige**. Der Tagesertrag/m² ist mit der Luftfeuchtigkeit negativ korreliert.

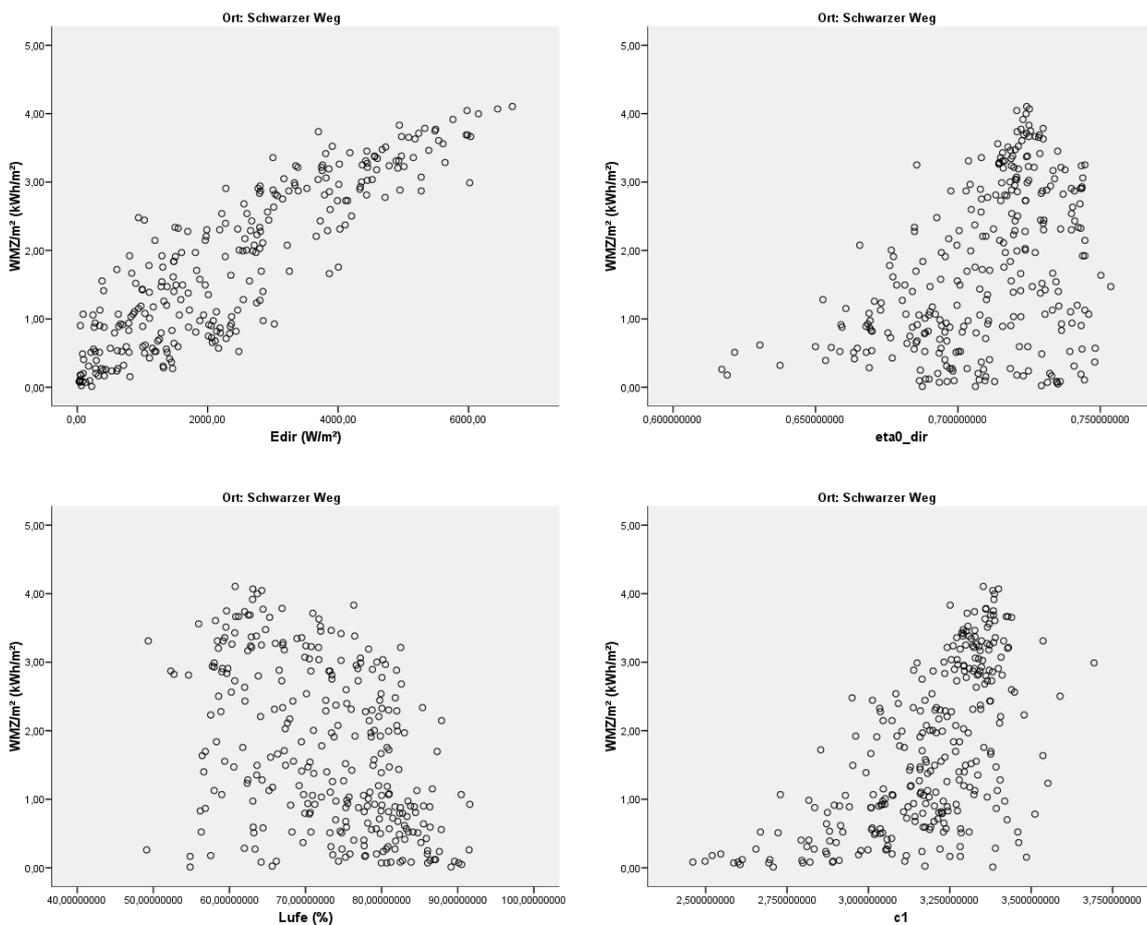
Der Tagesertrag/m² wird bei der Solaranlage am Schwarzen Weg zu fast **95%** vom Modell mit **zehn Regressoren** erklärt (siehe Tabelle 4.8). Den stärksten Einfluss haben die **Direktstrahlung geneigt**, der berechnete Kollektorkennwert **eta0_dir**, die **Luftfeuchtigkeit** sowie der berechnete Kollektorkennwert **c1** ($6,02 \geq |T| \geq 4,09$).

Tabelle 4.7: Korrelationen der Anlage Schwarzer Weg

	WMZ/m ² [kWh/m ²]									
	Edir [W/m ²]	eta0 dir	Lufe [%]	c1	1/4h- on	dT1 [°C]	eta0 diff	Lute [°C]	Wige [m/s]	Eglob [W/m ²]
Korrelation	,879**	,401**	-,434**	,606**	,910**	,334**	,760**	,646**	,226**	,957**
Signifikanz	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
N	276	296	297	296	297	297	296	297	297	276

** Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

Einen ähnlich großen Einfluss haben die Anzahl der Viertelstunden, an denen die Anlage in Betrieb ist, der Mittelwert zwischen primärseitigen Vor- und Rücklauftemperaturen (dT1), der berechnete Kollektorkennwert eta0_diff, die Lufttemperatur und die Windgeschwindigkeit mit T-Werten von ±3,23 bis ±2,73. Den geringsten Einfluss hat hier die Globalstrahlung geneigt mit T = 2,31 und einer Signifikanz von 0,022.



4.1. KORRELATIONEN UND LINEARE REGRESSIONEN DER TAGESWERTE

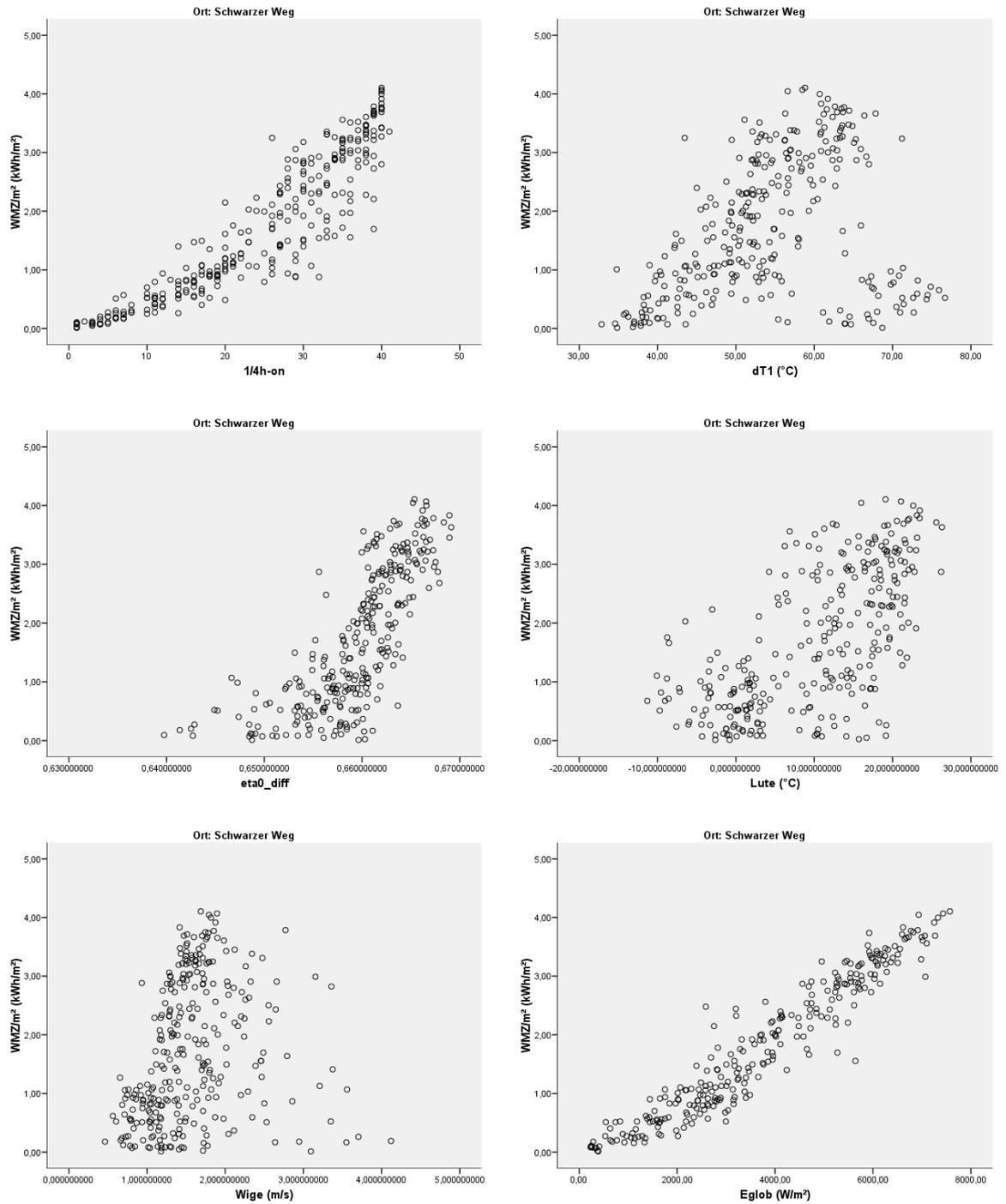


Abbildung 4.4: Streudiagramme der Regressoren mit WMZ/m^2 (Schwarzer Weg)

Tabelle 4.8: Lineare Regression (Schwarzer Weg)

Modellzusammenfassung^(b,c)

R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers	Durbin-Watson-Statistik
,974(a)	0,948	0,946	0,267	1,799

- a. Einflussvariablen: (Konstante), Eglob [W/m²], Wige [m/s], dT1 [°C], eta0_dir, Lute [°C], Lufe [%], c1, 1/4h-on, eta0_diff, Edir [W/m²]
 b. Abhängige Variable: WMZ/m² [kWh/m²]; n = 275
 c. Ort = Schwarzer Weg

Koeffizienten^(a)

	Nicht standardisierte Koeffizienten	Standardisierte Koeff.	Standardfehler	T	Signifikanz	95%-Konfidenzintervall für B	
	B	Beta				Untergrenze	Obergrenze
(Konstante)	-21,18687		6,764	-3,13	0,002		
Edir [W/m ²]	0,00027	0,396	0,000	6,02	< 10 ⁻³	0,00018	0,00036
eta0_dir	4,74088	0,102	0,882	5,38	< 10 ⁻³	3,00440	6,47735
Lufe [%]	-0,01490	-0,119	0,003	-4,43	< 10 ⁻³	-0,02152	-0,00827
c1	-1,00640	-0,158	0,246	-4,09	< 10 ⁻³	-1,49134	-0,52145
1/4h-on	0,02192	0,209	0,007	3,23	0,001	0,00856	0,03528
dT1 [°C]	-0,00587	-0,047	0,002	-2,96	0,003	-0,00978	-0,00196
eta0_diff	34,17001	0,145	11,601	2,95	0,004	11,32694	57,01308
Lute [°C]	0,01141	0,092	0,004	2,75	0,006	0,00323	0,01959
Wige [m/s]	-0,13934	-0,068	0,051	-2,73	0,007	-0,23972	-0,03897
Eglob [W/m ²]	0,00016	0,272	0,000	2,31	0,022	0,00002	0,00030

- a. Abhängige Variable: WMZ/m² [kWh/m²]
 b. Ort = Schwarzer Weg

THEODOR-KÖRNER-STRASSE

Die Solaranlage in der Theodor-Körner-Straße zeigt nur bei der Variablen **Eglob** eine **starke Korrelation** zum **Tagesertrag/m²** und bei den Variablen **Lute** und **Saison** mittlere Zusammenhänge. *Schwache Korrelationen* weisen die Variablen *Ediff*, *dT2*, *Wige* und *eta0_dir* auf, wie aus Tabelle 4.9 und Abbildung 4.5 hervorgeht.

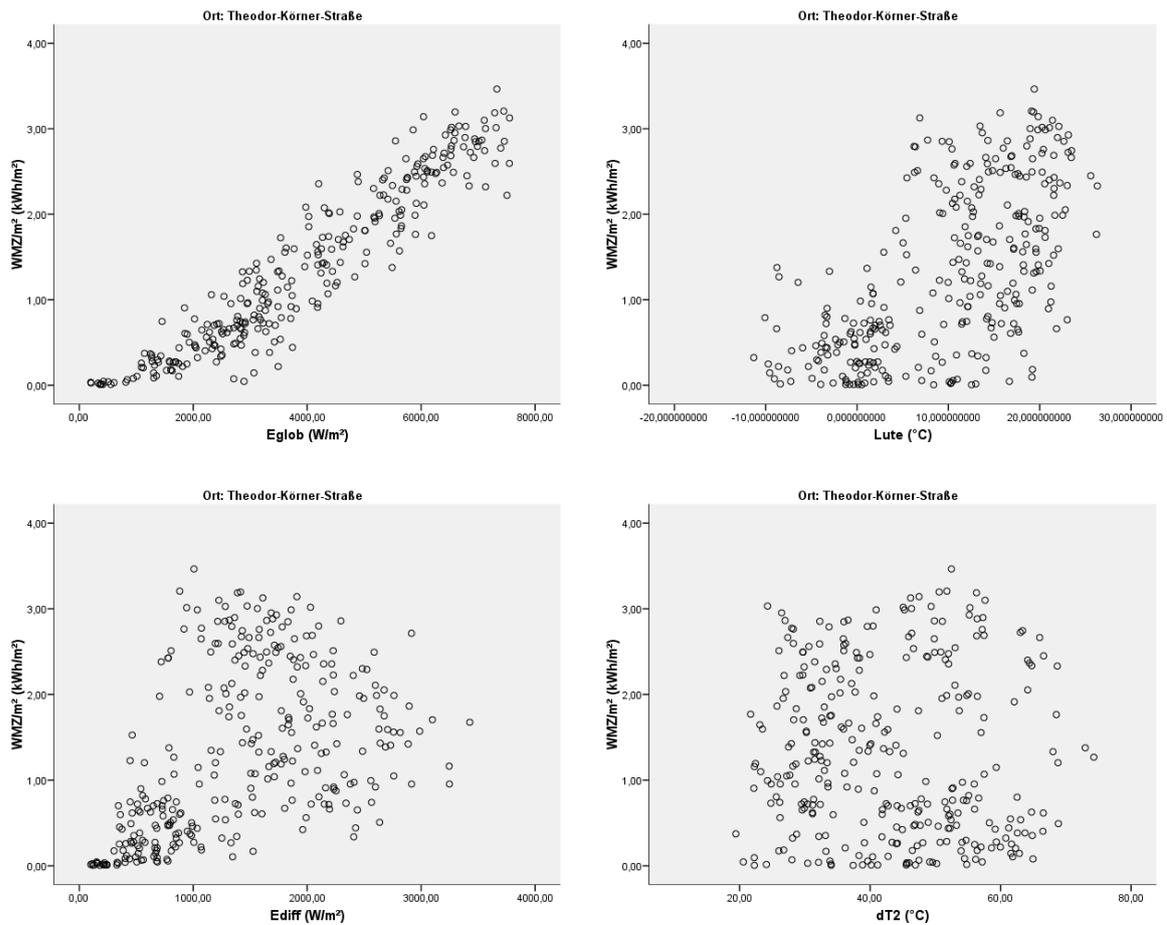
Das Modell mit **sieben Regressoren** erklärt bei der Solaranlage in der Theodor-Körner-Straße zu **93%** den Tagesertrag/m² (siehe Tabelle 4.10). Den stärksten Einfluss hat die **Globalstrahlung geneigt** mit T = 44,41, gefolgt von der Lufttemperatur,

Tabelle 4.9: Korrelationen der Anlage Theodor-Körner-Straße

	WMZ/m ² [kWh/m ²]						
	Eglob [W/m ²]	Lute [°C]	Ediff [W/m ²]	dT2 [°C]	Saison	Wige [m/s]	eta0_ dir
Korrelation	,944**	,618**	,460**	-0,045	,535**	,177**	,237**
Signifikanz	0,000	0,000	0,000	0,433	0,000	0,002	0,000
N	291	313	312	311	313	313	310

** Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.

der Diffusstrahlung geneigt und der Temperaturdifferenz zwischen der Außentemperatur zum Zeitpunkt des Sonnenaufgangs und der Variable „dT1“ zum Zeitpunkt des letzten Ausschaltens der Solaranlage (dT2) ($9,13 \geq |T| \geq 5,01$). Den geringsten Einfluss haben die Saison, die Windgeschwindigkeit und der berechnete Kollektorkennwert eta0_dir mit den T-Werten $-2,79$, $-2,61$ und $2,22$.



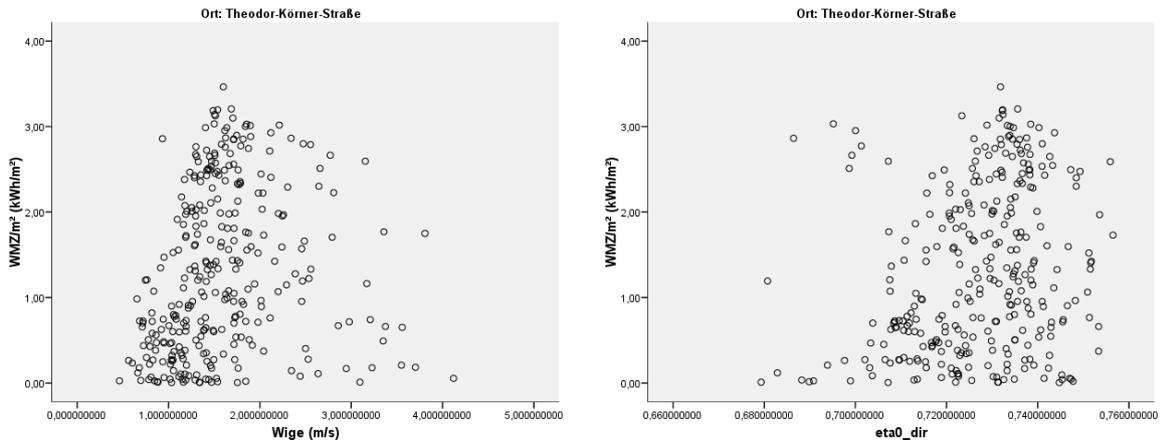


Abbildung 4.5: Streudiagramme der Regressoren mit WMZ/m² (Theodor-Körner-Str.)

Tabelle 4.10: Lineare Regression (Theodor-Körner-Straße)

Modellzusammenfassung^(b,c)

R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers	Durbin-Watson-Statistik
,966(a)	0,933	0,931	0,250	1,354

- a. Einflussvariablen: (Konstante), eta0_dir, dT2 [°C], Saison Eglob [W/m²], Wige [m/s], Ediff [W/m²], Lute [°C]
- b. Abhängige Variable: WMZ/m² [kWh/m²]; n = 289
- c. Ort = Theodor-Körner-Straße

Koeffizienten^(a)

	Nicht standardisierte Koeffizienten		Standardisierte Koeff.		95%-Konfidenzintervall für B		
	B	Standardfehler	Beta	T	Untergrenze	Obergrenze	
(Konstante)	-1,91006	0,885		-2,16	0,032		
Eglob [W/m ²]	0,00045	0,000	0,911	44,41	< 10 ⁻³	0,00043	0,00047
Lute [°C]	0,02755	0,003	0,266	9,13	< 10 ⁻³	0,02161	0,03349
Ediff [W/m ²]	-0,00020	0,000	-0,161	-6,76	< 10 ⁻³	-0,00026	-0,00014
dT2 [°C]	-0,00706	0,001	-0,093	-5,01	< 10 ⁻³	-0,00984	-0,00429
Saison	-0,15639	0,056	-0,082	-2,79	0,006	-0,26663	-0,04615
Wige [m/s]	-0,06989	0,027	-0,044	-2,61	0,010	-0,12266	-0,01712
eta0_dir	2,79869	1,262	0,040	2,22	0,027	0,31355	5,28384

- a. Abhängige Variable: WMZ/m² [kWh/m²]
- b. Ort = Theodor-Körner-Straße

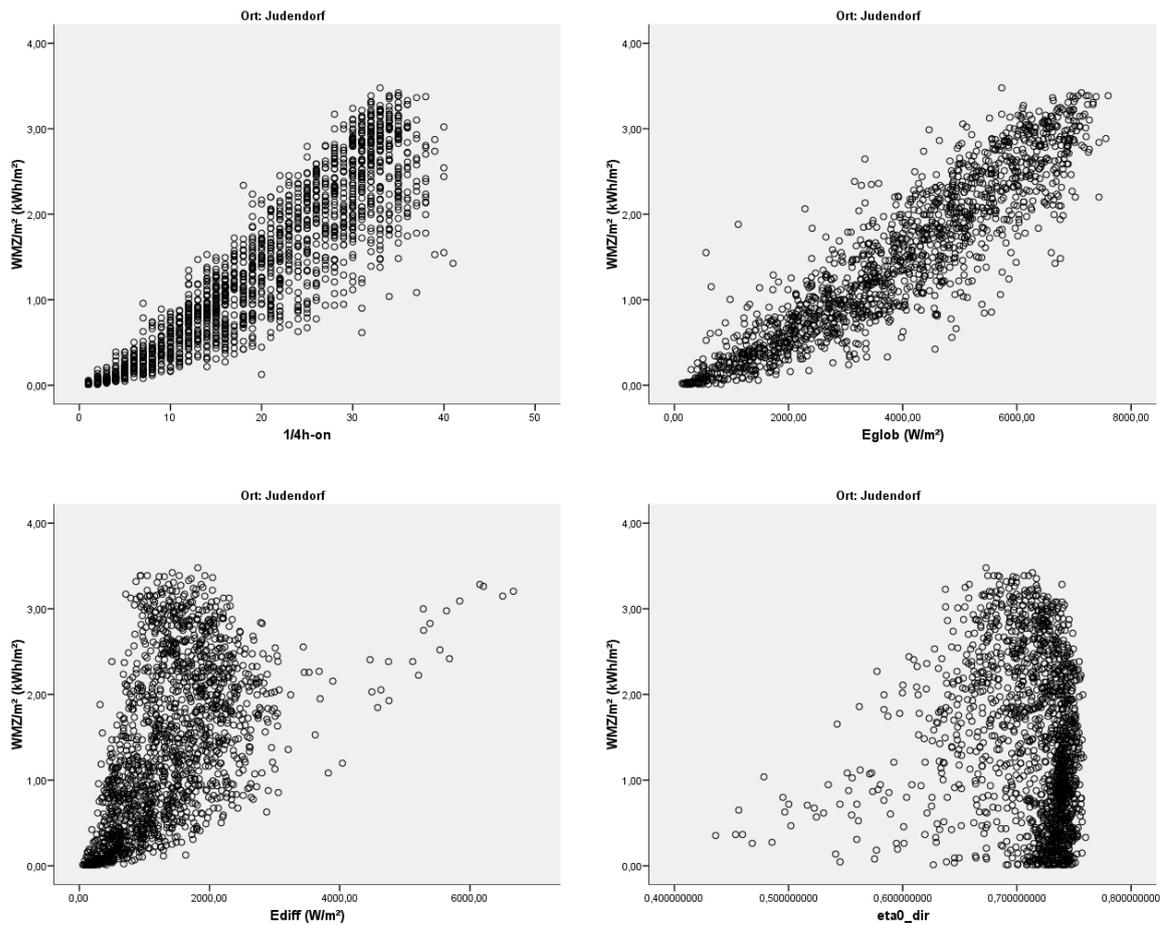
JUDENDORF

Die Variablen **1/4h-on** und **Eglob** zeigen laut Tabelle 4.11 und Abbildung 4.6 bei der Solaranlage in Judendorf **starke Korrelationen** zum **Tagesertrag/m²**. Mittlere Zusammenhänge weisen die Variablen **Ediff**, **Saison**, **Lute**, **eta0_diff** und **dT1** auf, während nur *schwache Zusammenhänge* zwischen den Variablen **eta0_dir**, **Lufe** und **Wige** mit

Tabelle 4.11: Korrelationen der Anlage Judendorf

	WMZ/m ² [kWh/m ²]									
	1/4h-on	Eglob [W/m ²]	Ediff [W/m ²]	eta0_dir	Saison	Lute [°C]	Lufe [%]	eta0_diff	Wige [m/s]	dT1 [°C]
Korrelation	,901**	,909**	,527**	-,111**	,506**	,587**	-,418**	,747**	,247**	,657**
Signifikanz	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000
N	1.658	1.522	1.660	1.646	1.660	1.660	1.660	1.646	1.660	1.660

** Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.



KAPITEL 4. EINFLUSS DER REGRESSOREN AUF DIE ZIELVARIABLE WMZ/M²

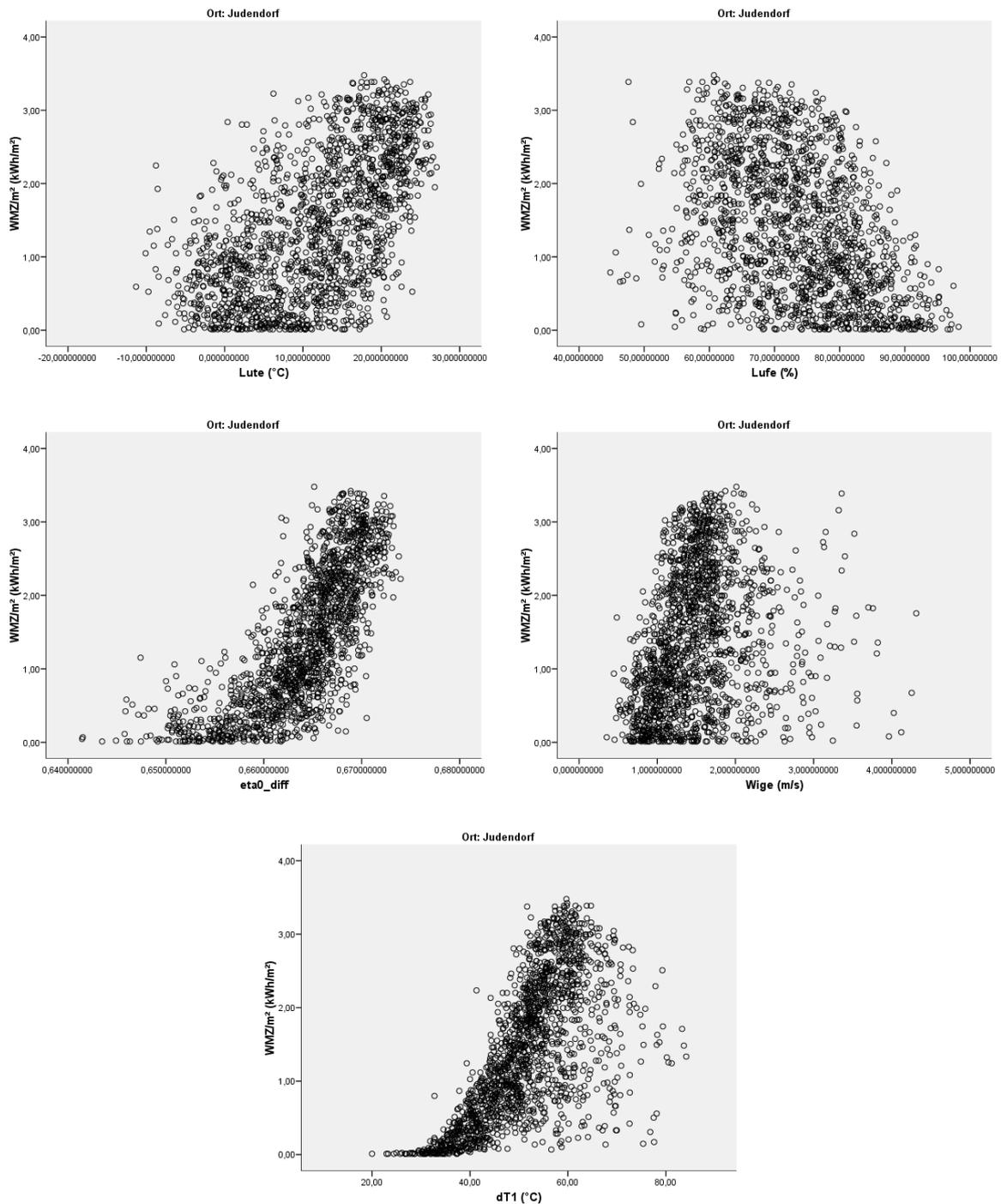


Abbildung 4.6: Streudiagramme der Regressoren mit WMZ/m² (Judendorf)

dem Tagesertrag/m² zu erkennen sind.

Der Tagesertrag/ m^2 bei der Solaranlage in Judendorf wird zu **90%** vom Modell mit **zehn Regressoren** erklärt (siehe Tabelle 4.12). Den stärksten Einfluss haben die **Anzahl der Viertelstunden**, an denen die Anlage in Betrieb ist, die **Global- und Diffusstrahlung geneigt** mit den T-Werten 21, 37, 20, 79 und $-17, 1$, gefolgt von dem berechneten Kollektorkennwert eta0_dir , der Saison, der Lufttemperatur, der Luftfeuchtigkeit und dem berechneten Kollektorkennwert eta0_diff ($6, 84 \geq |T| \geq 3, 74$). Den geringsten Einfluss haben die Windgeschwindigkeit mit $T = 3, 16$ und der Mittelwert zwischen primärseitigen Vor- und Rücklauftemperaturen (dT1) mit $T = 3, 02$.

Tabelle 4.12: Lineare Regression (Judendorf)

Modellzusammenfassung^(b,c)

R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers	Durbin-Watson-Statistik
,950(a)	0,902	0,901	0,286	1,303

- a. Einflussvariablen: (Konstante), dT1 [$^{\circ}\text{C}$], eta0_dir , Saison
Wige [m/s], Ediff [W/m^2], Lufe [%], Eglob [W/m^2],
Lute [$^{\circ}\text{C}$], eta0_diff , $1/4\text{h-on}$
b. Abhängige Variable: WMZ/m^2 [kWh/m^2]; $n = 1.512$
c. Ort = Judendorf

Koeffizienten^(a)

	Nicht standardisierte Koeffizienten		Standardisierte Koeff.		Signifikanz	95%-Konfidenzintervall für B	
	B	Standardfehler	Beta	T		Untergrenze	Obergrenze
(Konstante)	6,48685	2,159		3,00	0,003		
$1/4\text{h-on}$	0,04910	0,002	0,518	21,37	$< 10^{-3}$	0,04459	0,05361
Eglob [W/m^2]	0,00026	0,000	0,534	20,79	$< 10^{-3}$	0,00024	0,00029
Ediff [W/m^2]	-0,00030	0,000	-0,220	-17,10	$< 10^{-3}$	-0,00033	-0,00026
eta0_dir	1,34088	0,196	0,065	6,84	$< 10^{-3}$	0,95633	1,72543
Saison	0,16878	0,027	0,092	6,33	$< 10^{-3}$	0,11646	0,22110
Lute [$^{\circ}\text{C}$]	0,00998	0,002	0,091	5,51	$< 10^{-3}$	0,00643	0,01354
Lufe [%]	0,00547	0,001	0,055	4,40	$< 10^{-3}$	0,00303	0,00792
eta0_diff	-12,78345	3,422	-0,067	-3,74	$< 10^{-3}$	-19,49538	-6,07153
Wige [m/s]	0,05911	0,019	0,036	3,16	0,002	0,02247	0,09576
dT1 [$^{\circ}\text{C}$]	0,00392	0,001	0,040	3,02	0,003	0,00137	0,00648

- a. Abhängige Variable: WMZ/m^2 [kWh/m^2]
b. Ort = Judendorf

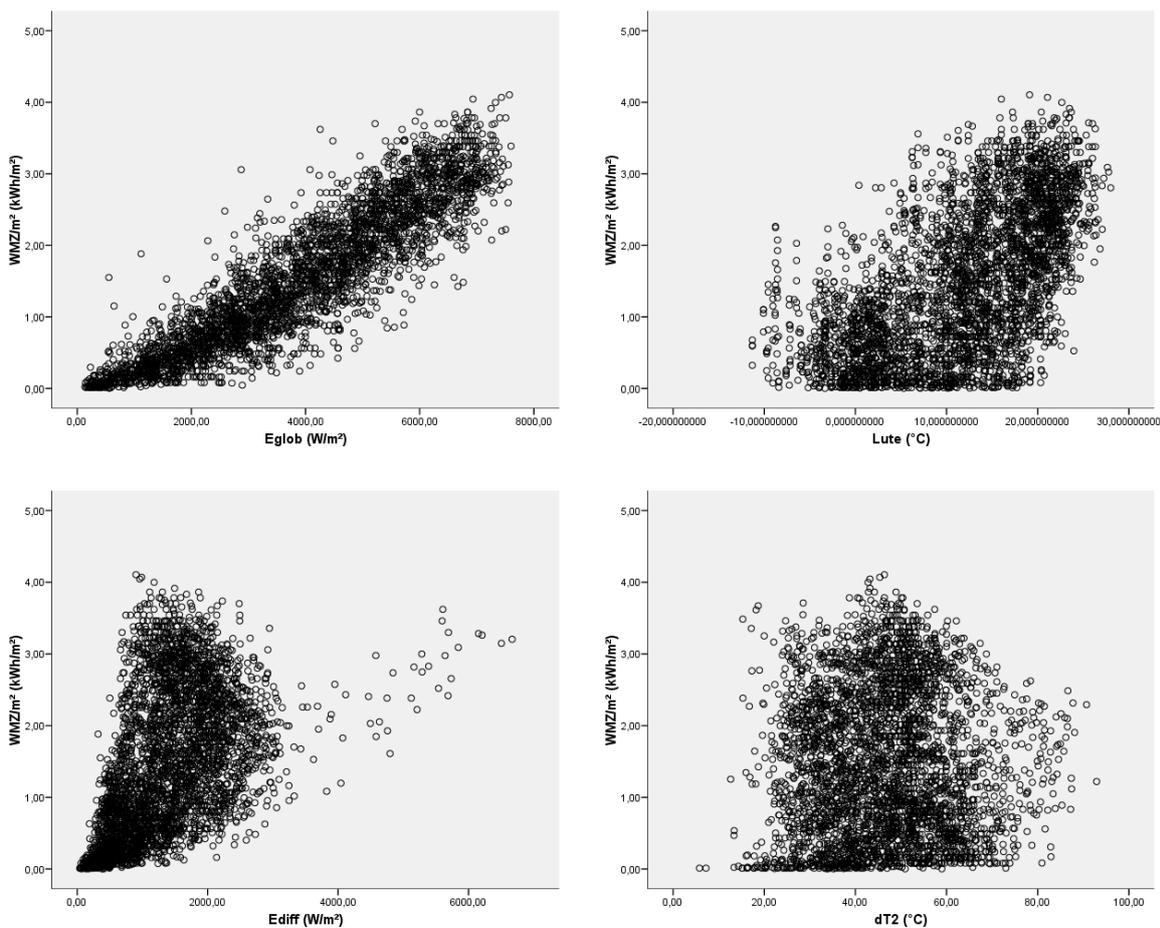
ALLE SOLARANLAGEN

Tabelle 4.13 und Abbildung 4.7 zeigen, dass bei der Analyse der Zusammenhänge aller Solaranlagen zwischen dem **Tagesertrag/m²** und den Regressoren die Variablen **Eglob, Sonne** und **1/4h-on** **starke Korrelationen** aufweisen, die Variablen **Lute** und **Saison** einen mittleren Zusammenhang. *Schwache Korrelationen* besitzen die Variablen *Ediff, Glykolgehalt, Neigungswinkel, dT2, c1* und *eta0_dir*.

Tabelle 4.13: Korrelationen aller Solaranlagen

	WMZ/m ² [kWh/m ²]										
	Eglob	Lute	Ediff	Glykolgehalt	Neigungswinkel	dT2	Saison	c1	eta0_dir	Sonne	1/4h-on
	[W/m ²]	[°C]	[W/m ²]	[%]	[°]	[°C]				[min]	
Korrelation	,903**	,591**	,490**	-0,011	-,065**	0,017	,525**	,339**	,058**	,816**	,823**
Signifikanz	0,000	0,000	0,000	0,489	0,000	0,313	0,000	0,000	0,000	0,000	0,000
N	3.471	3.712	3.711	3.712	3.712	3.662	3.712	3.694	3.694	3.712	3.710

** Die Korrelation ist auf dem Niveau von 0,01 (2-seitig) signifikant.



4.1. KORRELATIONEN UND LINEARE REGRESSIONEN DER TAGESWERTE

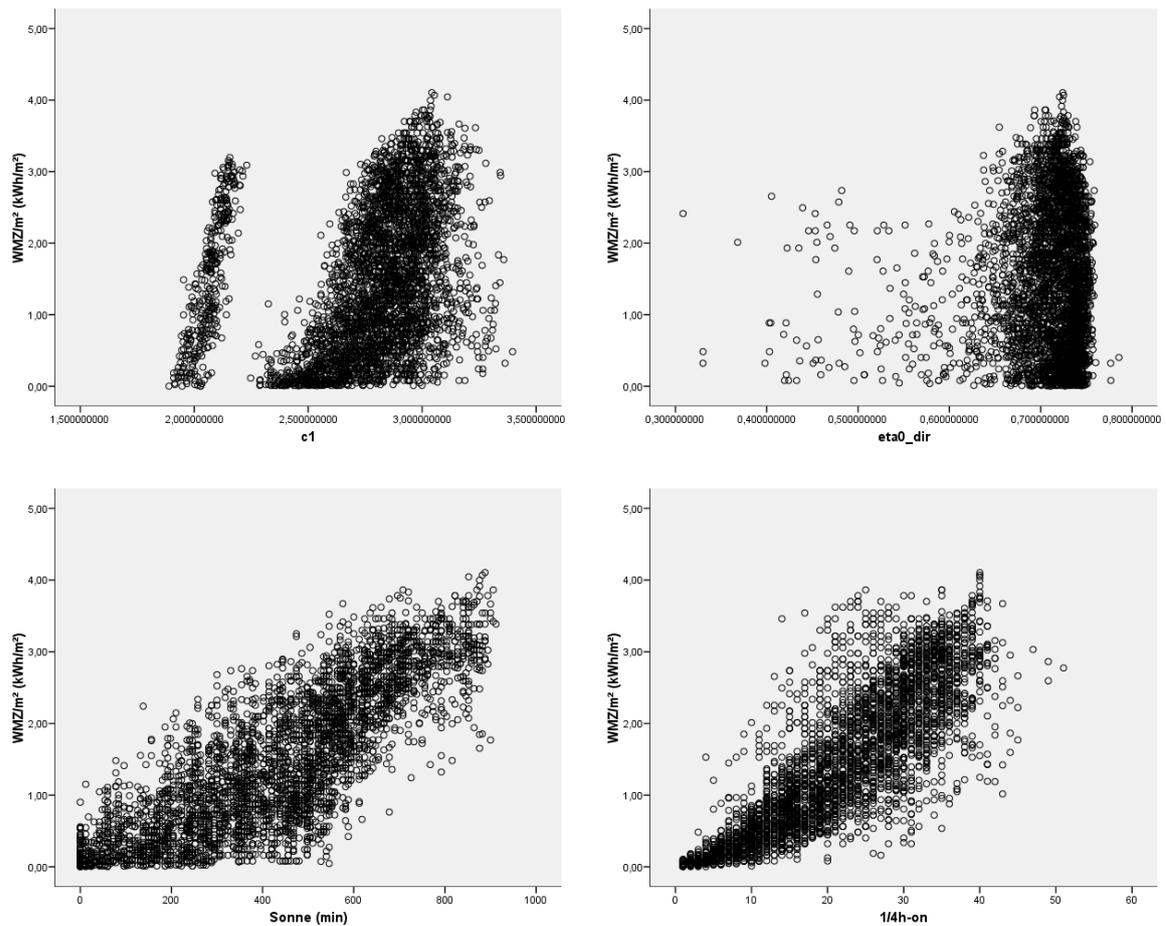


Abbildung 4.7: Streudiagramme der Regressoren mit WMZ/m² (Alle Solaranlagen)

Die Fehlerbalken in Abbildung 4.8 auf der nächsten Seite zeigen, dass im Sommer (April bis September) mit 1,94 kWh/m² mehr als der doppelte mittlere Tagesertrag/m² zu erwarten ist als im Winter (Oktober bis März) mit 0,87 kWh/m². Die Boxplots zeigen, dass der Median im Sommer mit 2 kWh/m² fast dreimal so hoch ist wie jener im Winter mit 0,74 kWh/m².

Der Tagesertrag/m² wird vom Modell für alle Solaranlagen mit **elf Regressoren** zu fast **88%** erklärt (siehe Tabelle 4.14). Den stärksten Einfluss haben die **Globalstrahlung geneigt**, die **Lufttemperatur** und die **Diffusstrahlung geneigt** mit den T-Werten 34, 72, 16, 61 und -13, 3. Der Glykolgehalt, der Kollektorneigungswinkel, die Temperaturdifferenz zwischen der Außentemperatur zum Zeitpunkt des Sonnenaufgangs und der Variable „dT1“ zum Zeitpunkt des letzten Ausschaltens der Solaranlage

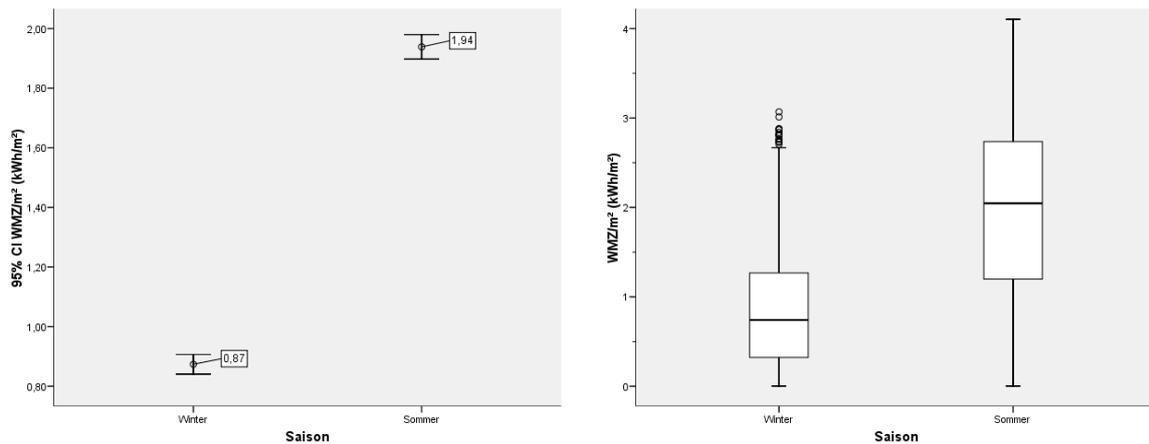


Abbildung 4.8: Fehlerbalken & Boxplots der Tageserträge/m² bzgl. der Saison

(dT2), die Saison, die berechneten Kollektorkennwerte $c1$ und η_{0_dir} , die Sonnenscheindauer und die Anzahl an Viertelstunden, an denen die Anlage in Betrieb ist, gehen ebenfalls hoch signifikant ins Modell ein ($11,79 \geq |T| \geq 4,16$).

Tabelle 4.15 zeigt eine Zusammenfassung der eben vorgestellten linearen Regressionsmodelle, welche die wesentlichen Einflussgrößen beinhalten, sowie die korrigierten R^2 und Standardfehler der Schätzer. Die Zahlen in der Tabelle bestimmen die Reihenfolge der Einflüsse auf die Modelle, d.h. der Regressor mit der Zahl 1 hat den stärksten Einfluss auf das Modell und derjenige mit der höchsten Zahl den geringsten. Die Größe dieser Einflüsse wurden aus der Spalte der Prüfgröße „T“, welche die Quotienten aus den berechneten (nicht standardisierten) Koeffizienten „B“ und deren Standardfehler bilden, der Regressionstabellen entnommen. Darüber hinaus sieht man in der Spalte „Anzahl“, wie oft eine Variable in die Modelle der einzelnen Solaranlagen eingeht.

Tabelle 4.14: Lineare Regression (Alle Solaranlagen)

Modellzusammenfassung ^(b)				
R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers	Durbin-Watson-Statistik
,936(a)	0,876	0,876	0,344	1,158

a. Einflussvariablen: (Konstante), 1/4h-on, η_{0_dir} , $c1$, dT2 [°C], Neigungswinkel [°], Saison, Ediff [W/m²], Lute [°C], Sonne [min], Glykolgehalt [%], Eglob [W/m²]
 b. Abhängige Variable: WMZ/m² [kWh/m²]; n = 3.418

4.1. KORRELATIONEN UND LINEARE REGRESSIONEN DER TAGESWERTE

	Koeffizienten ^(a)						
	Nicht standardisierte Koeffizienten		Standardisierte Koeff.		95%-Konfidenzintervall für B		
	B	Standardfehler	Beta	T	Signifikanz	Untergrenze	Obergrenze
(Konstante)	5,38302	0,533		10,11	$< 10^{-3}$		
Eglob [W/m ²]	0,00041	0,000	0,767	34,72	$< 10^{-3}$	0,00039	0,00043
Lute [°C]	0,02025	0,001	0,174	16,61	$< 10^{-3}$	0,01786	0,02264
Ediff [W/m ²]	-0,00018	0,000	-0,128	-13,30	$< 10^{-3}$	-0,00021	-0,00015
Glykolgehalt [%]	-0,14156	0,012	-0,184	-11,79	$< 10^{-3}$	-0,16510	-0,11803
Neigungswinkel [°]	-0,01500	0,001	-0,083	-11,64	$< 10^{-3}$	-0,01753	-0,01247
dT2 [°C]	-0,00484	0,001	-0,065	-8,03	$< 10^{-3}$	-0,00602	-0,00366
Saison	0,14691	0,021	0,074	7,04	$< 10^{-3}$	0,10598	0,18785
c1	-0,31101	0,045	-0,111	-6,91	$< 10^{-3}$	-0,39928	-0,22275
eta0_dir	0,82735	0,124	0,041	6,68	$< 10^{-3}$	0,58465	1,07005
Sonne [min]	0,00032	0,000	0,068	4,63	$< 10^{-3}$	0,00018	0,00045
1/4h-on	0,00680	0,002	0,069	4,16	$< 10^{-3}$	0,00360	0,01000

a. Abhängige Variable: WMZ/m² [kWh/m²]

Tabelle 4.15: Zusammenfassung der linearen Regressionen aller Solaranlagen

	S O L A R A N L A G E ^(a)							Alle Anlagen
	L	G	E	S	T	J	Anzahl	
Saison	3	6			5	5	4	7
dT1 [°C]		4		6		10	3	
1/4h-on	7	3	1	5		1	5	11
dT2 [°C]			6		4		2	6
eta0_dir	8		3	2	7	4	5	9
eta0_diff				7		8	2	
c1	6			4			2	8
c2		2					1	
Edir [W/m ²]			2	1			2	
Ediff [W/m ²]	4		7		3	3	4	3
Eglob [W/m ²]	1	1		10	1	2	5	1
Sonne [min]	5	5	4				3	10
Lute [°C]	2			8	2	6	4	2
Wige [m/s]			5	9	6	9	4	
Lufe [%]				3		7	2	
Glykolgehalt [%]								4
Neigungswinkel [°]								5
korr. R²	0,911	0,950	0,928	0,946	0,931	0,901		0,876
Standardfehler	0,318	0,201	0,255	0,267	0,250	0,286		0,344

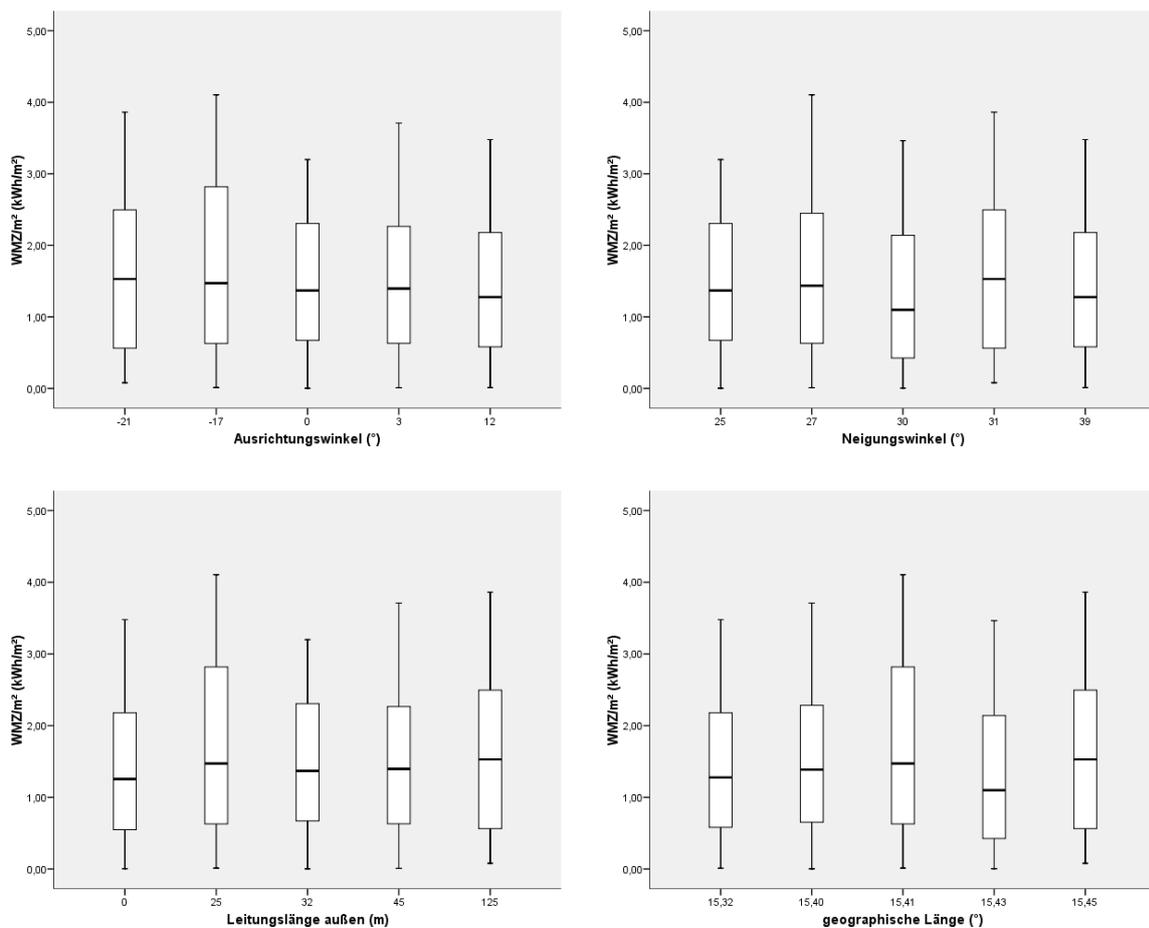
a. L ... Stadion Liebenau, G ... Grottenhofstraße, E ... Eggenberger Allee, S ... Schwarzer Weg, T ... Theodor-Körner-Straße, J ... Judendorf

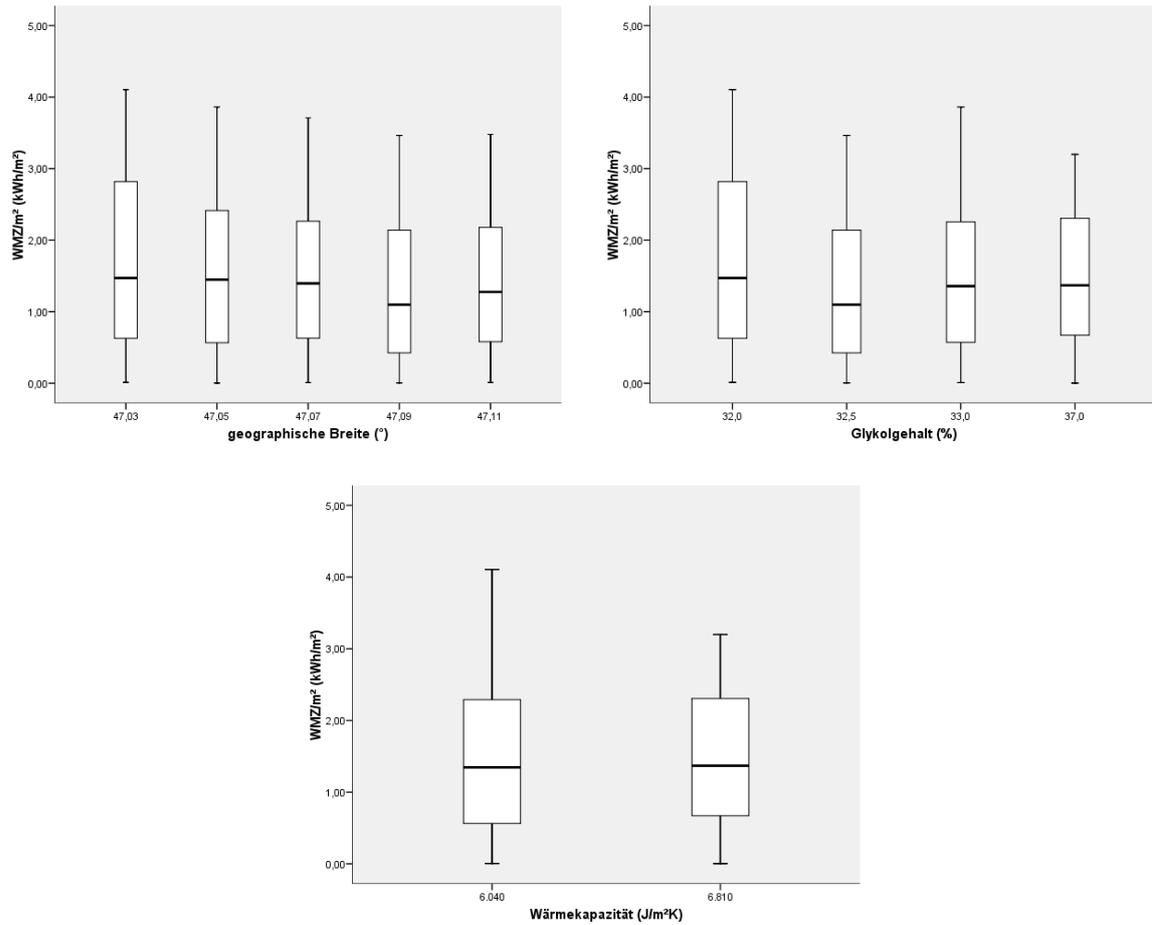
4.2 Analyse der Anlagedaten

Die Anlagedaten Ausrichtungswinkel [°], Neigungswinkel Kollektor [°], Leitungslänge außen [m], geographische Länge [°], geographische Breite [°], Glykolgehalt [%] und Wärmekapazität [J/m²K] sind bei den sechs Solaranlagen teilweise sehr unterschiedlich und können so ebenfalls eine große Rolle bei der linearen Regression für den Tagesertrag/m² spielen.

Bei der ersten linearen Regression für alle Solaranlagen gehen die Variablen Glykolgehalt und Neigungswinkel als Regressoren ein, wie aus Tabelle 4.14 auf der vorherigen Seite hervorgeht.

Die Boxplots der Tageserträge/m² bezüglich dem Ausrichtungswinkel (-21° bis 12°) weist auf einen leichten Abfall im Median hin, bezüglich dem Neigungswinkel (25° bis



Abbildung 4.9: Boxplots der Tageserträge/m² bzgl. der Anlagedaten

39°) zeigen sie keine eindeutige Tendenz. Die Boxplots bezüglich der Leitungslänge außen (0m bis 125m) und geographischen Länge (15,32° bis 15,45°) zeigen eine leichte Steigung im Median mit je einer Ausnahme in der Grottenhofstraße (25m) und der Theodor-Körner-Straße (15,43°), die Boxplots der geographischen Breite (47,03° bis 47,11°) und Glykolgehalt (32% bis 37%) eher eine abfallende Tendenz im Median mit ebenfalls je einer Ausnahme in der Theodor-Körner-Straße (47,09° und 32,5%). Die Boxplots der Wärmekapazität (6.040 J/m²K und 6.810 J/m²K) deuten auf einen leichten Anstieg im Median hin.

5 Prognosen für den Tagesertrag pro m²

5.1 Lineare Regressionsmodelle mit berechneten & vorhersagbaren Variablen

5.1.1 Modelle mit geneigter Strahlung auf Kollektorebene (berechnet)

Tabelle 5.1: Lineare Regressionsansätze für die Prognosen aller Solaranlagen

	S O L A R A N L A G E ^(a)						Anzahl	Alle Anlagen
	L	G	E	S	T	J		
Saison	x				x	x	3	x
Edir [W/m ²]	x	x		x			3	
Ediff [W/m ²]					x	x	2	x
Eglob [W/m ²]	x	x	x	x	x	x	6	x
Lufttemperatur [°C]	x	x		x	x	x	5	x
Windgeschwindigkeit [m/s]	x	x		x	x		4	x
Luftfeuchtigkeit [%]			x	x		x	3	x
geogr. Breite [°]								x
geogr. Länge [°]								x
Glykolgehalt [%]								x
Neigungswinkel [°]								x
korr. R²	0,908	0,943	0,871	0,933	0,925	0,866		0,875
Standardfehler	0,324	0,214	0,339	0,297	0,260	0,335		0,345

a. **L** ... Stadion Liebenau, **G** ... Grottenhofstraße, **E** ... Eggenberger Allee, **S** ... Schwarzer Weg, **T** ... Theodor-Körner-Straße, **J** ... Judendorf

Tabelle 5.1 zeigt die Regressoren, die korrigierten R^2 und Standardfehler der Schätzer der linearen Regressionsanalysen unter der Methode „**Schrittweise**“ (vgl. Seite 105). Das bedeutet, dass als erstes die Variable in die Gleichung aufgenommen wird, die mit der Zielvariablen am höchsten (positiv oder negativ) korreliert ist. Danach werden die partiellen Korrelationen der restlichen Variablen mit der Zielvariable berechnet, also „bereinigt“ von schon aufgenommenen Regressoren, und die Variable mit der höchsten (positiven oder negativen) partiellen Korrelation geht ins Modell ein.⁵

Tabelle 5.2: Prognosemodell mit berechneten Regressoren für alle Solaranlagen

Modellzusammenfassung ^(b)						
R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers	Durbin-Watson-Statistik	F	Signifikanz
,935(a)	0,875	0,874	0,347	1,193	3.015	$< 10^{-3}$

a. Einflussvariablen: (Konstante), Saison, Ediff [W/m²], Eglob [W/m²], Neigungswinkel [°], geogr. Breite [°], geogr. Länge [°], Glykolgehalt [%], Lufttemperatur [°C]
 b. Abhängige Variable: WMZ/m² [kWh/m²]; n = 3.471

Koeffizienten ^(a)							
	Nicht standardisierte Koeffizienten		Standardisierte Koeff.		Signifikanz	95%-Konfidenzintervall für B	
	B	Standardfehler	Beta	T		Untergrenze	Obergrenze
(Konstante)	266,25498	24,212		11,00	$< 10^{-3}$		
Saison	0,16116	0,021	0,081	7,76	$< 10^{-3}$	0,12042	0,20189
Ediff [W/m ²]	-0,00016	0,000	-0,111	-13,62	$< 10^{-3}$	-0,00018	-0,00013
Eglob [W/m ²]	0,00043	0,000	0,818	113,56	$< 10^{-3}$	0,00043	0,00044
Neigungswinkel [°]	-0,02748	0,003	-0,153	-10,21	$< 10^{-3}$	-0,03276	-0,02220
geogr. Breite [°]	-4,38858	0,476	-0,134	-9,23	$< 10^{-3}$	-5,32125	-3,45590
geogr. Länge [°]	-3,60488	0,235	-0,202	-15,36	$< 10^{-3}$	-4,06496	-3,14481
Glykolgehalt [%]	-0,10800	0,006	-0,140	-18,81	$< 10^{-3}$	-0,11925	-0,09674
Lufttemperatur [°C]	0,02442	0,001	0,210	20,26	$< 10^{-3}$	0,02206	0,02679

a. Abhängige Variable: WMZ/m² [kWh/m²]

Durch die Regression für **alle Solaranlagen mit zehn Regressoren** (siehe Tabelle 5.1 rechts) kann **87,5%** der Variabilität von WMZ/m² erklärt werden. Nimmt man die Windgeschwindigkeit und Luftfeuchtigkeit aus diesem Modell, da nur Solaranlagen in und um Graz zur Analyse zur Verfügung stehen und hier durch die „Kessellage“ spezielle Wetterbedingungen herrschen, und verwendet für die lineare Regression das Verfahren

⁵<http://www.lrz.de/services/schulung/unterlagen/spss/spss-regression/>

„Einschluss“ (vgl. Seite 104), verschlechtert sich das korrigierte R^2 nur geringfügig auf **87,4%**, wie aus Tabelle 5.2 hervorgeht.

DIAGNOSEPLOTS FÜR DIE GÜTE DES MODELLS

Die Residuen (beobachtete Werte minus Modellwerte) im Q-Q-Plot in Abbildung 5.1 links sind mit Ausnahme der Tails annähernd normalverteilt. Auf der rechten Seite dieser Abbildung zeigt der Residuenplot, welcher die unstandardisierten Vorhersagewerte auf der x-Achse gegen die standardisierten Residuen auf der y-Achse aufträgt, in der positiven Ebene der Residuen eine zufällige Struktur und damit keine Verletzung der Modellannahmen und in der negativen Ebene ab ca. 1 kWh/m² ebenfalls. Die Struktur, die sich bei den negativen Residuen bis zum Wert von ca. 1 kWh/m² zeigt, rührt daher, dass es bei der Gewinnung von Solarenergie keine negativen Erträge geben kann.

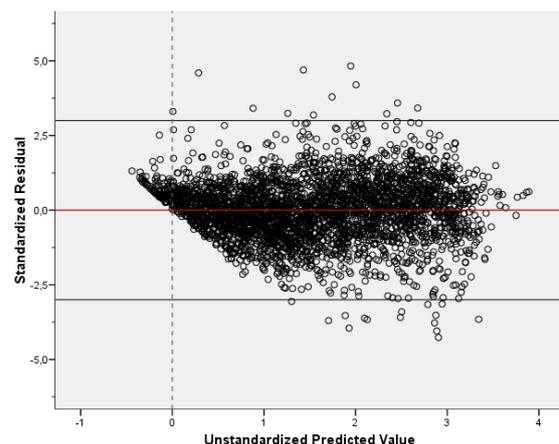
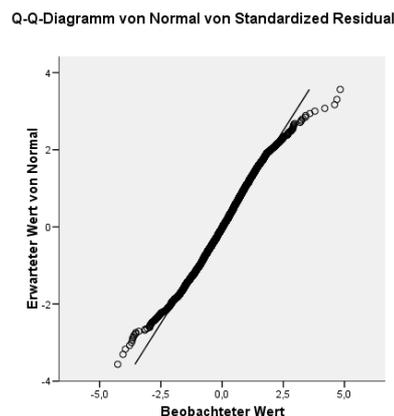


Abbildung 5.1: Q-Q-Diagramm der standardisierten Residuen und Residuenplot

Abbildung 5.2 zeigt den Vergleich der gemessenen Werte (blau) mit den Vorhersagewerten (grün). Die linke Graphik, die den Verlauf der beiden Werte bezüglich der Monate über alle Jahre wiedergibt, zeigt für die Prognose der Tageserträge in kWh/m² durch die Regression mit den berechneten Strahlungswerten E_{diff} [W/m²] und E_{glob} [W/m²], den Anlagedaten Neigungswinkel [°], geographische Breite [°], geographische Länge [°] und Glykolgehalt [%], sowie der Lufttemperatur [°C] eine sehr gute Anpassung. Lediglich in den Monaten Februar, September und November sind kleine Abweichungen zu erkennen. Die rechte Graphik zeigt die Liniendiagramme der Jahre 2002 bis 2008 getrennt und verdeutlicht die gute Anpassung der Prognose an die gemessenen Werte.

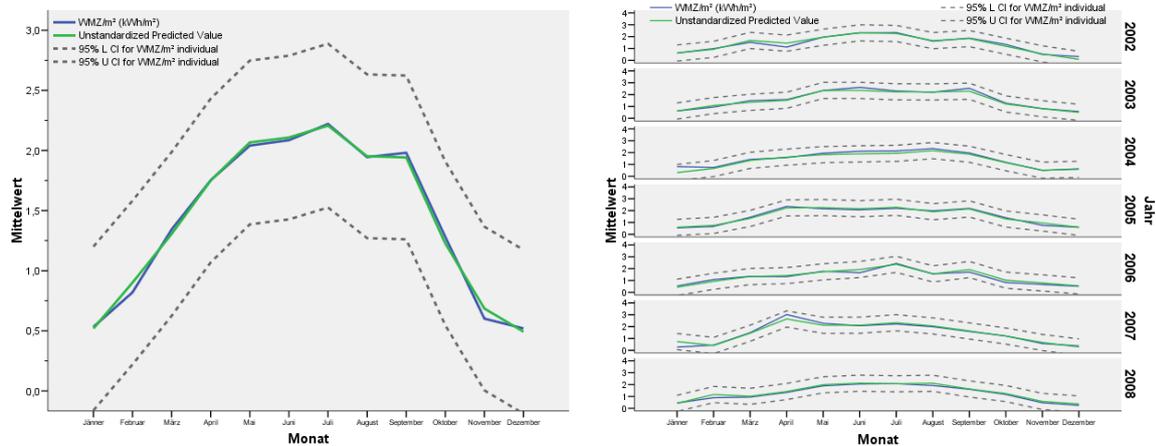


Abbildung 5.2: Liniendiagramme WMZ/m² | geschätzt-WMZ/m²

5.1.2 Modelle mit horizontaler Strahlung (gemessen)

Durch die Regression mit den gemessenen **horizontalen Strahlungswerten** Globalstrahlung [W/m²] und Diffusanteil [%] anstatt der berechneten Eglob [W/m²] und Ediff [W/m²] verringert sich das korrigierte R^2 von **87,4%** auf **82,8%**. Mit den weiteren Regressoren Neigungswinkel [°], geographische Breite [°], geographische Länge [°], Glykolgehalt [%] und Lufttemperatur [°C], welche identisch mit jenen in Tabelle 5.2 auf Seite 54 sind, beträgt die Verschlechterung der Erklärungskraft der Variabilität von WMZ/m² also ca. 5%. Allerdings hat das Modell mit den vorhersagbaren Regressoren den Vorteil, dass diese Strahlungswerte **beobachtbar** und dadurch weniger fehleranfällig sind. Aus diesem Grund scheinen diese Modelle auch für die Berechnung von Prognosen geeigneter zu sein.

Tabelle 5.3: Prognosemodell mit vorhersagbaren Regressoren für alle Solaranlagen

Modellzusammenfassung ^(b)						
R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers	Durbin-Watson-Statistik	F	Signifikanz
,910(a)	0,829	0,828	0,415	1,006	2.238	< 10 ⁻³

a. Einflussvariablen: (Konstante), Saison, Globalstrahlung [W/m²], Diffusanteil [%], Neigungswinkel [°], geogr. Breite [°], geogr. Länge [°], Glykolgehalt [%], Lufttemperatur [°C]
 b. Abhängige Variable: WMZ/m² [kWh/m²]; n = 3.712

5.1. LINEARE REGRESSIONSMODELLE MIT BERECHNETEN & VORHERSAGBAREN VARIABLEN

	Koeffizienten ^(a)						
	Nicht standardisierte Koeffizienten		Standardisierte Koeff.		Signifikanz	95%-Konfidenzintervall für B	
	B	Standardfehler	Beta	T		Untergrenze	Obergrenze
(Konstante)	135,81094	27,875		4,87	$< 10^{-3}$		
Saison	-0,12647	0,025	-0,062	-4,99	$< 10^{-3}$	-0,17620	-0,07675
Globalstrahlung [W/m ²]	0,00035	0,000	0,691	50,83	$< 10^{-3}$	0,00034	0,00036
Diffusanteil [%]	-0,01114	0,000	-0,273	-29,45	$< 10^{-3}$	-0,01188	-0,01039
Neigungswinkel [°]	-0,04166	0,003	-0,226	-13,45	$< 10^{-3}$	-0,04774	-0,03559
geogr. Breite [°]	-1,50777	0,548	-0,045	-2,75	0,006	-2,58139	-0,43415
geogr. Länge [°]	-3,87687	0,273	-0,211	-14,21	$< 10^{-3}$	-4,41170	-3,34204
Glykolgehalt [%]	-0,09867	0,007	-0,122	-14,73	$< 10^{-3}$	-0,11180	-0,08554
Lufttemperatur [°C]	0,01637	0,001	0,137	11,36	$< 10^{-3}$	0,01354	0,01919

a. Abhängige Variable: WMZ/m² [kWh/m²]

DIAGNOSEPLOTS FÜR DIE GÜTE DES MODELLS

Die Residuen im Q-Q-Plot in Abbildung 5.3 links sind mit Ausnahme der kleinsten Werte annähernd normalverteilt. Die Daten im Residuenplot rechts weisen im positiven Teil der standardisierten Residuen eine annähernd zufällige Struktur auf und im negativen Teil ab ca. 1 kWh/m² für die unstandardisierten Vorhersagewerte ebenfalls.

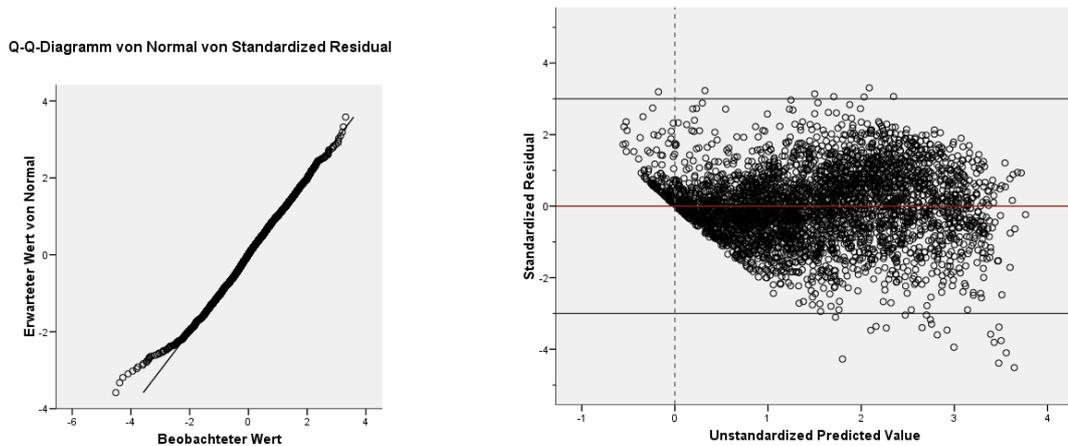


Abbildung 5.3: Q-Q-Plot der standardisierten Residuen und Residuenplot

Die Liniendiagramme in Abbildung 5.4 zeigen die zuvor beschriebene geringfügige Verschlechterung der Anpassung der Prognose an die gemessenen Werte, speziell in den Monaten Juni, Juli und September. Allerdings liegen die Vorhersagewerte immer noch deutlich im Toleranzintervall.

KAPITEL 5. PROGNOSEN FÜR DEN TAGESERTRAG PRO M²

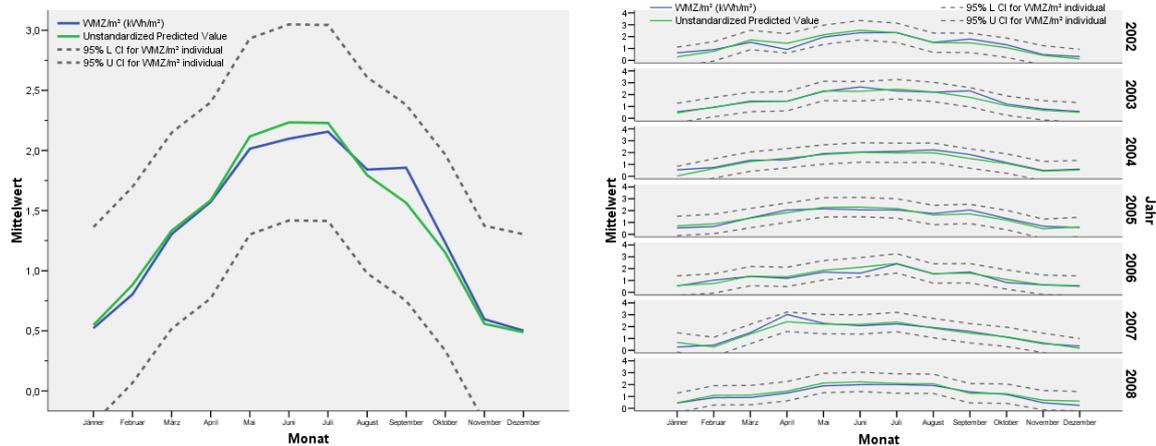


Abbildung 5.4: Liniendiagramme WMZ/m² | geschätzt-WMZ/m²

Da nur Solaranlagen mit ähnlichen geographischen Koordinaten zur Verfügung stehen ist es für das Prognosemodell sinnvoller, diese Daten nicht zu berücksichtigen, da diese

Tabelle 5.4: Prognosemodell ohne geographische Koordinaten der Solaranlagen

Modellzusammenfassung^(b)

R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers	Durbin-Watson-Statistik	F	Signifikanz
,907(a)	0,823	0,823	0,422	0,976	2.466	< 10 ⁻³

a. Einflussvariablen: (Konstante), Saison, Globalstrahlung [W/m²], Diffusanteil [%], Neigungswinkel [°], Leitungslänge außen [m], Wärmekapazität [J/m²K], Lufttemperatur [°C]

b. Abhängige Variable: WMZ/m² [kWh/m²]; n = 3.712

Koeffizienten^(a)

	Nicht standardisierte Koeffizienten		Standardisierte Koeff.	Beta	T	Signifikanz	95%-Konfidenzintervall für B	
	B	Standardfehler					Untergrenze	Obergrenze
(Konstante)	3,91859	0,270			14,49	< 10 ⁻³		
Saison	-0,12598	0,026	-0,062		-4,89	< 10 ⁻³	-0,17647	-0,07550
Globalstrahlung [W/m ²]	0,00035	0,000	0,691		50,09	< 10 ⁻³	0,00034	0,00036
Diffusanteil [%]	-0,01110	0,000	-0,272		-28,92	< 10 ⁻³	-0,01185	-0,01034
Neigungswinkel [°]	-0,02489	0,002	-0,135		-14,32	< 10 ⁻³	-0,02830	-0,02148
Leitungslänge außen [m]	-0,00189	0,000	-0,091		-11,03	< 10 ⁻³	-0,00222	-0,00155
Wärmekapazität [J/m ² K]	-0,00041	0,000	-0,091		-10,86	< 10 ⁻³	-0,00048	-0,00033
Lufttemperatur [°C]	0,01619	0,001	0,135		11,07	< 10 ⁻³	0,01332	0,01906

a. Abhängige Variable: WMZ/m² [kWh/m²]

durch die kleinen Unterschiede zu stark ins Modell eingehen, wie in Tabelle 5.3 auf Seite 56 ersichtlich. Im Prognosemodell in Tabelle 5.4 **ohne geographische Koordinaten** ist vor allem eine starke Veränderung der Konstante bemerkbar. Hingegen verschlechtert sich das korrigierte R^2 nur leicht von **82,8%** auf **82,3%**.

Die Diagnoseplots zur Güte des Modells zeigen kaum Veränderungen zu den Abbildungen 5.3 und 5.4 und sind daher hier nicht abgebildet.

5.2 Lineare Regressionsmodelle ohne Kenntnis der Anlagedaten

Im Folgenden finden die für die sechs Solaranlagen in und um Graz bekannten Anlagedaten bei der Berechnung der Prognosemodelle keine Berücksichtigung, um einen Eindruck davon bekommen zu können, wie gut der Tagesertrag in kWh/m² durch die **Meteorologie** alleine erklärt werden kann.

5.2.1 Modelle mit geneigter Strahlung auf Kollektorebene (berechnet)

Das korrigierte R^2 fällt im Vergleich zum Modell in Tabelle 5.2 auf Seite 54 beim Prognosemodell mit den Regressoren Saison, Ediff [W/m²], Eglob [W/m²] und Lufttemperatur [°C] um nur 1,7% von **87,4%** auf **85,7%**. Auch der Standardfehler des Schätzers wird von 0,346 kWh/m² auf 0,373 kWh/m² nur geringfügig größer.

Tabelle 5.5: Prognosemodell mit den Regressoren Saison, Ediff, Eglob und Lute

Modellzusammenfassung ^(b)						
R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers	Durbin-Watson-Statistik	F	Signifikanz
,925(a)	0,855	0,855	0,373	1,033	5.100	< 10 ⁻³

a. Einflussvariablen: (Konstante), Saison, Ediff [W/m²], Eglob [W/m²], Lufttemperatur [°C]
 b. Abhängige Variable: WMZ/m² [kWh/m²]; n = 3.471

	Koeffizienten ^(a)						
	Nicht standardisierte Koeffizienten		Standardisierte Koeff.	T	Signifikanz	95%-Konfidenzintervall für B	
	B	Standardfehler	Beta			Untergrenze	Obergrenze
(Konstante)	-0,28959	0,017		-17,08	$< 10^{-3}$		
Saison	0,18522	0,022	0,094	8,31	$< 10^{-3}$	0,14153	0,22891
Ediff [W/m ²]	-0,00015	0,000	-0,110	-12,52	$< 10^{-3}$	-0,00018	-0,00013
Eglob [W/m ²]	0,00043	0,000	0,818	105,70	$< 10^{-3}$	0,00043	0,00044
Lufttemperatur [°C]	0,02180	0,001	0,187	17,06	$< 10^{-3}$	0,01930	0,02431

a. Abhängige Variable: WMZ/m² [kWh/m²]

DIAGNOSEPLOTS FÜR DIE GÜTE DES MODELLS

Das Histogramm und der Q-Q-Plot in Abbildung 5.5 zeigen, dass die Residuen annähernd normalverteilt sind. Die Normalverteilungskurve im Histogramm links überdeckt die aufgetragenen standardisierten Residuen sehr gut und auch im Q-Q-Plot rechts ist mit Ausnahme der Tails eine sehr gute Anpassung zu erkennen.

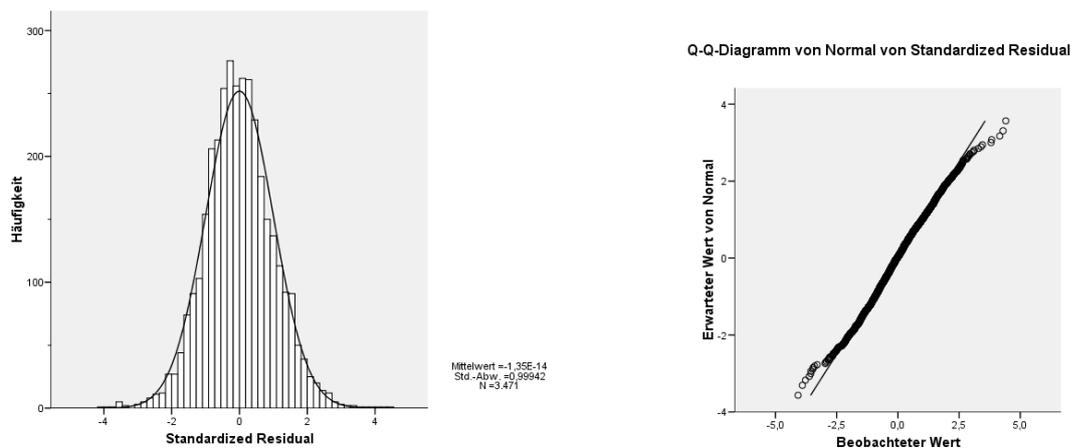


Abbildung 5.5: Histogramm und Q-Q-Plot der standardisierten Residuen

Der Residuenplot in Abbildung 5.6 links zeigt in der positiven Ebene eine zufällige Struktur und in der negativen Ebene ab ca. 1 kWh/m² ebenfalls, mit nur wenigen Ausreißern nach oben oder unten. Der Vergleich zwischen den gemessenen Tageserträgen/m² und den Vorhersagewerten im Liniendiagramm rechts zeigt eine sehr gute Anpassung, vor allem in den für den solaren Ertrag wichtigen Sommermonaten. Das Prognosemodell mit den Regressoren Saison, Ediff [W/m²], Eglob [W/m²] und Lufttemperatur [°C] scheint den wahren Wert im Februar und September etwas zu überschätzen.

5.2. LINEARE REGRESSIONSMODELLE OHNE KENNTNIS DER ANLAGEDATEN

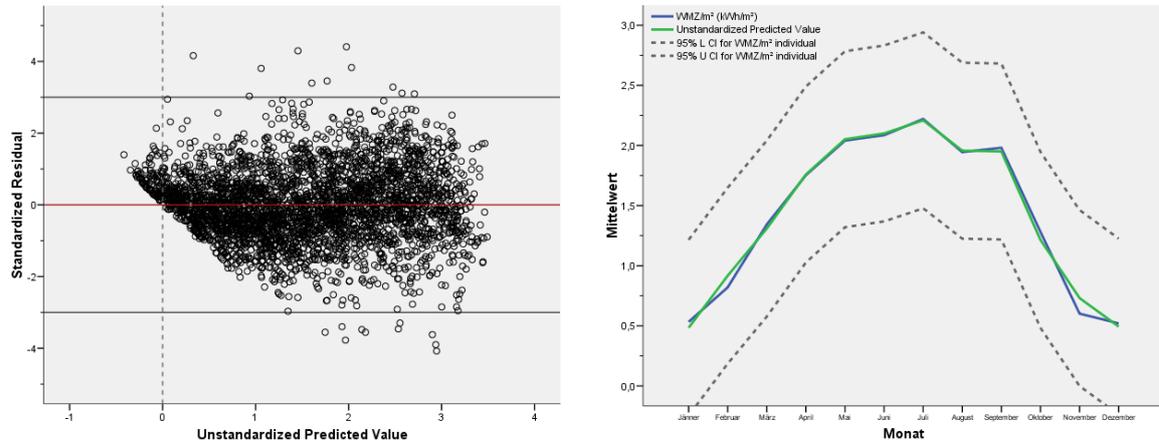


Abbildung 5.6: Residuenplot und Liniendiagramm WMZ/m² | geschätzt-WMZ/m²

5.2.2 Modelle mit horizontaler Strahlung (gemessen)

Werden aus dem Modell mit den gemessenen **horizontalen Strahlungswerten** in Tabelle 5.4 auf Seite 58 die Anlagedaten Neigungswinkel [°], Leitungslänge außen [m] und

Tabelle 5.6: Prognosemodell mit den Regressoren Saison, Globalstrahlung, Diffusanteil und Lufttemperatur

Modellzusammenfassung^(b)

R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers	Durbin-Watson-Statistik	F	Signifikanz
,901(a)	0,812	0,812	0,434	0,922	4.011	< 10 ⁻³

a. Einflussvariablen: (Konstante), Saison, Globalstrahlung [W/m²], Diffusanteil [%], Lufttemperatur [°C]

b. Abhängige Variable: WMZ/m² [kWh/m²]; n = 3.712

Koeffizienten^(a)

	Nicht standardisierte Koeffizienten	Standardfehler	Standardisierte Koeff.	Beta	T	Signifikanz	95%-Konfidenzintervall für B	
							Untergrenze	Obergrenze
(Konstante)	0,55959	0,036			15,68	< 10 ⁻³		
Saison	-0,10353	0,026		-0,051	-3,91	< 10 ⁻³	-0,15546	-0,05160
Globalstrahlung [W/m ²]	0,00035	0,000	0,689	0,689	48,60	< 10 ⁻³	0,00033	0,00036
Diffusanteil [%]	-0,01111	0,000	-0,272	-0,272	-28,22	< 10 ⁻³	-0,01188	-0,01034
Lufttemperatur [°C]	0,01363	0,001	0,114	0,114	9,14	< 10 ⁻³	0,01071	0,01656

a. Abhängige Variable: WMZ/m² [kWh/m²]

Wärmekapazität [J/m²K] genommen, so verringert sich das korrigierte R^2 um nur 1,1% von **82,3%** auf **81,2%**. Auch der Standardfehler des Schätzers vergrößert sich beim Prognosemodell mit den Regressoren Saison, Globalstrahlung [W/m²], Diffusanteil [%] und Lufttemperatur [°C] von 0,422 kWh/m² auf 0,434 kWh/m² nur geringfügig.

Dieses Modell wird in Kombination mit dem Modell aus Tabelle 5.4 auf Seite 58 für den Excel Tagesertrags-Prognosefile „**PROGNOSE_Tag**“ verwendet (siehe Anhang A.1).

DIAGNOSEPLOTS FÜR DIE GÜTE DES MODELLS

Die standardisierten Residuen aus dem Modell mit horizontaler Strahlung sind annähernd normalverteilt, wie in Abbildung 5.7 sowohl im Histogramm rechts als auch im Q-Q-Plot links ersichtlich. Die Anpassung an die Normalverteilung scheint zumindest an den Tails sogar eine bessere zu sein als beim Modell mit geneigter Strahlung (siehe Abbildung 5.5 auf Seite 60).

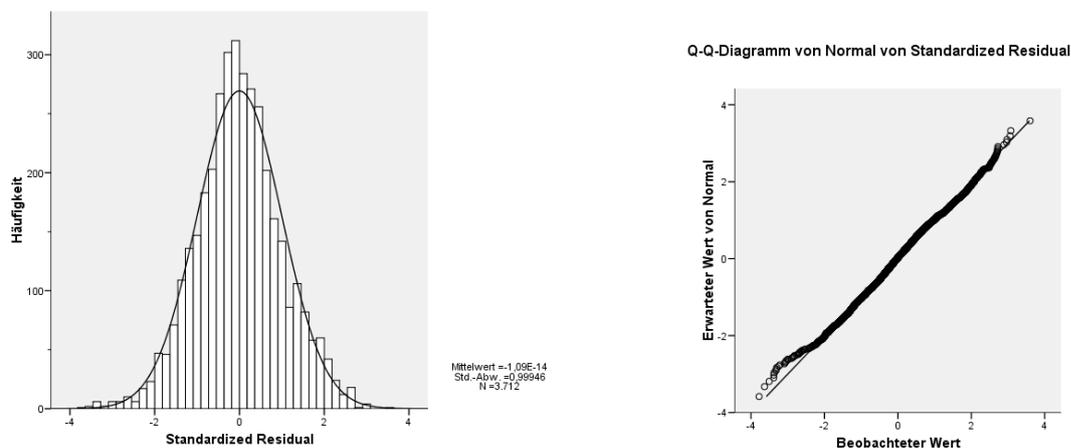


Abbildung 5.7: Histogramm und Q-Q-Plot der standardisierten Residuen

Die positiven standardisierten Residuen im Residuenplot in Abbildung 5.8 links weisen mit nur zwei Ausreißern nach oben ein zufälliges Muster auf. Auch die negativen standardisierten Residuen scheinen mit wenigen Ausreißern nach unten zufällig verteilt zu sein. Der Vergleich zwischen den gemessenen Tageserträgen/m² pro Monat und den

5.2. LINEARE REGRESSIONSMODELLE OHNE KENNTNIS DER ANLAGEDATEN

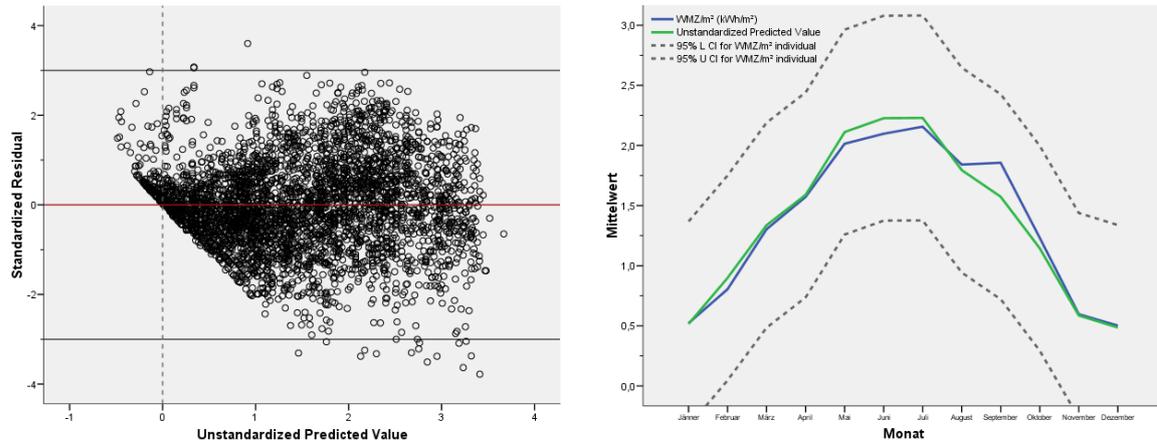
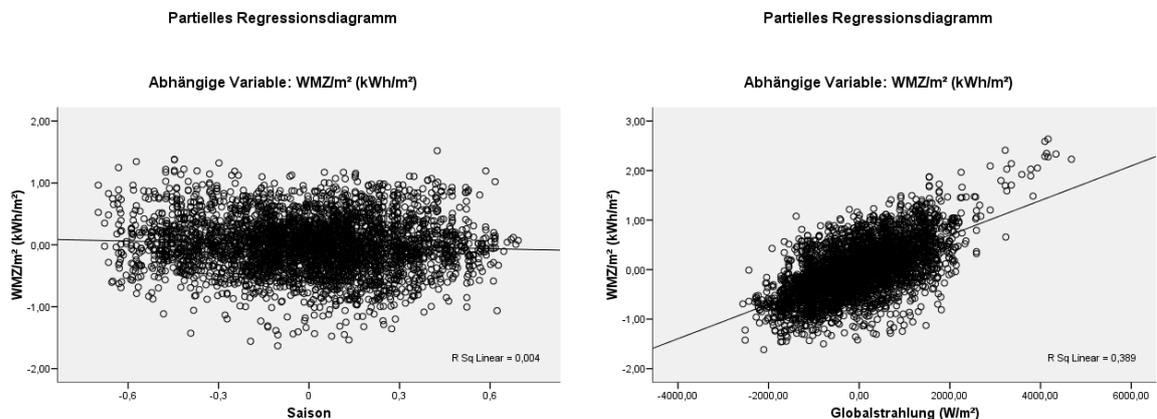


Abbildung 5.8: Residuenplot und Liniendiagramm WMZ/m^2 | geschätzt- WMZ/m^2

Vorhersagewerten im Liniendiagramm rechts zeigt trotz des Weglassens der Anlagendaten immer noch eine gute Anpassung und die Vorhersagewerte liegen deutlich im Toleranzintervall.

Die partiellen Regressionsdiagramme (vgl. Seite 102) in Abbildung 5.9 zeigen einen starken positiven Zusammenhang des Tagesertrages [kWh/m^2] zur Globalstrahlung [W/m^2], einen schwachen positiven zur Lufttemperatur [$^{\circ}C$] und einen negativen zum Diffusanteil [%]. Die Saison weist nur mehr einen schwachen negativen Zusammenhang zum Tagesertrag auf.



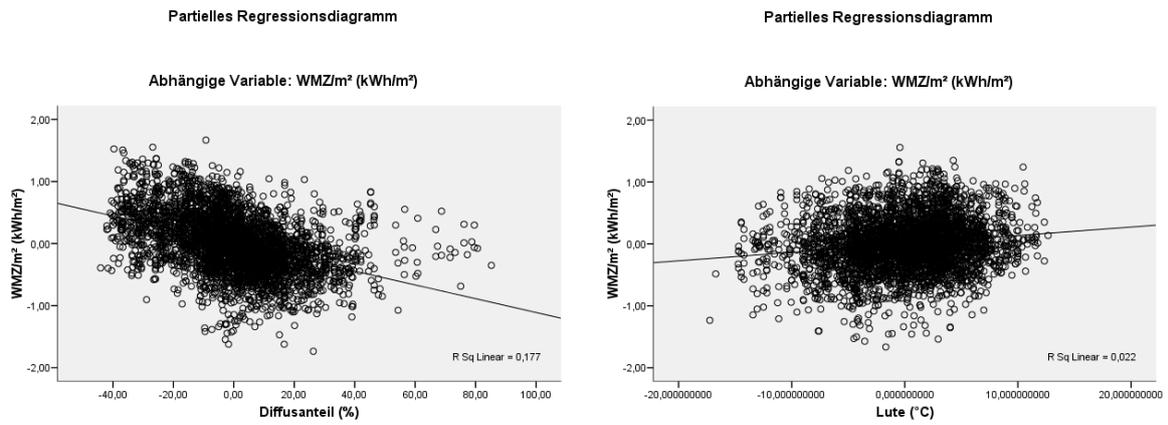


Abbildung 5.9: Partielle Regressionsdiagramme der Regressoren und WMZ/m^2

6 Prognosen für den Monatsertrag pro m²

6.1 Lineare Regressionsmodelle mit berechneten & vorhersagbaren Variablen

Tabelle 6.1: Zusammenfassung der linearen Regressionen aller Solaranlagen

	S O L A R A N L A G E ^(a)						Anzahl	Alle Anlagen
	L	G	E	S	T	J		
Saison	3				1		2	5
dT1 [°C]							0	
1/4h-on			4			2	2	
dT2 [°C]	6			1			2	3
eta0_dir	4						1	
eta0_diff							0	
c1	5		2				2	
c2	7				2		2	
Edir [W/m ²]			3				1	4
Ediff [W/m ²]		3					1	
Eglob [W/m ²]				2	3		2	
Sonnenscheindauer [min]	2	2	1			1	4	1
Lufttemperatur [°C]	1	1					2	2
Windgeschwindigkeit [m/s]							0	
Luftfeuchtigkeit [%]							0	
Ausrichtungswinkel [°]								6
korr. R²	0,971	0,985	0,997	0,992	0,986	0,943		0,934
Standardfehler	3,595	2,596	1,117	2,246	2,475	4,894		5,422

a. **L** ... Stadion Liebenau, **G** ... Grottenhofstraße, **E** ... Eggenberger Allee, **S** ... Schwarzer Weg, **T** ... Theodor-Körner-Straße, **J** ... Judendorf

Tabelle 6.1 zeigt für die Berechnung der **Monatserträge in kWh/m²** die Regressoren und die Reihenfolge, in der diese mittels der **schrittweisen linearen Regression** pro Solaranlage getrennt und alle Anlagen gemeinsam in die Modelle eingehen, sowie die korrigierten R^2 und Standardfehler der Schätzer.

6.1.1 Modelle mit geneigter Strahlung auf Kollektorebene (berechnet)

Tabelle 6.2 zeigt in den Spalten 2 bis 7 die Regressoren, die korrigierten R^2 und die Standardfehler der Schätzer der **schrittweisen linearen Regressionen** je Solaranlage. Die letzte Spalte enthält diese Merkmale für alle Solaranlagen gemeinsam.

Tabelle 6.2: Lineare Regressionsansätze für die Prognosen aller Solaranlagen

	S O L A R A N L A G E ^(a)						Anzahl	Alle Anlagen
	L	G	E	S	T	J		
Saison	x						1	x
Edir [W/m ²]	x		x				2	
Ediff [W/m ²]							0	x
Eglob [W/m ²]		x		x	x	x	4	x
Lufttemperatur [°C]	x	x		x	x	x	5	x
Windgeschwindigkeit [m/s]							0	
Luftfeuchtigkeit [%]	x		x			x	3	
Glykolgehalt [%]								x
Neigungswinkel [°]								x
korr. R^2	0,934	0,981	0,978	0,99	0,969	0,951		0,915
Standardfehler	5,400	2,972	3,200	2,531	3,656	4,556		6,172

a. **L** ... Stadion Liebenau, **G** ... Grottenhofstraße, **E** ... Eggenberger Allee, **S** ... Schwarzer Weg, **T** ... Theodor-Körner-Straße, **J** ... Judendorf

Das in Tabelle 6.3 dargestellte lineare Regressionsmodell für alle Solaranlagen wurde mit der Methode „**Einschluss**“ berechnet und beinhaltet die Variable Saison, die berechneten Strahlungswerte Ediff [W/m²] und Eglob [W/m²], die Anlagedaten Neigungswinkel [°] und Glykolgehalt [%], sowie die Lufttemperatur [°C]. Mit diesem Prognosemodell können **90,9%** der Variabilität des Monatsertrags in kWh/m² erklärt werden, mit einem Standardfehler des Schätzers von 6,34 kWh/m².

6.1. LINEARE REGRESSIONSMODELLE MIT BERECHNETEN & VORHERSAGBAREN VARIABLEN

Tabelle 6.3: Prognosemodell mit berechneten Regressoren für alle Solaranlagen

Modellzusammenfassung^(b)

R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers	Durbin-Watson-Statistik	F	Signifikanz
,955(a)	0,913	0,909	6,340	1,179	279	$< 10^{-3}$

a. Einflussvariablen: (Konstante), Saison, Ediff [W/m²], Eglob [W/m²], Neigungswinkel [°], Glykolgehalt [%], Lufttemperatur [°C]
 b. Abhängige Variable: WMZ/m² [kWh/m²]; n = 167

Koeffizienten^(a)

	Nicht standardisierte Koeffizienten		Standardisierte Koeff.		Signifikanz	95%-Konfidenzintervall für B	
	B	Standardfehler	Beta	T		Untergrenze	Obergrenze
(Konstante)	83,09659	15,653		5,31	$< 10^{-3}$		
Saison	-6,88824	1,983	-0,163	-3,47	0,001	-10,80383	-2,97266
Ediff [W/m ²]	-0,00018	0,000	-0,169	-2,41	0,017	-0,00033	-0,00003
Eglob [W/m ²]	0,00031	0,000	0,640	9,45	$< 10^{-3}$	0,00024	0,00037
Neigungswinkel [°]	-0,39286	0,102	-0,099	-3,86	$< 10^{-3}$	-0,59400	-0,19171
Glykolgehalt [%]	-1,66332	0,420	-0,101	-3,96	$< 10^{-3}$	-2,49331	-0,83333
Lufttemperatur [°C]	1,11129	0,136	0,399	8,18	$< 10^{-3}$	0,84305	1,37954

a. Abhängige Variable: WMZ/m² [kWh/m²]

DIAGNOSEPLOTS FÜR DIE GÜTE DES MODELLS

Das in Abbildung 6.1 links dargestellte Histogramm der standardisierten Residuen zeigt eine hinreichend gute Anpassung an die eingezeichnete Normalverteilungskurve. Die

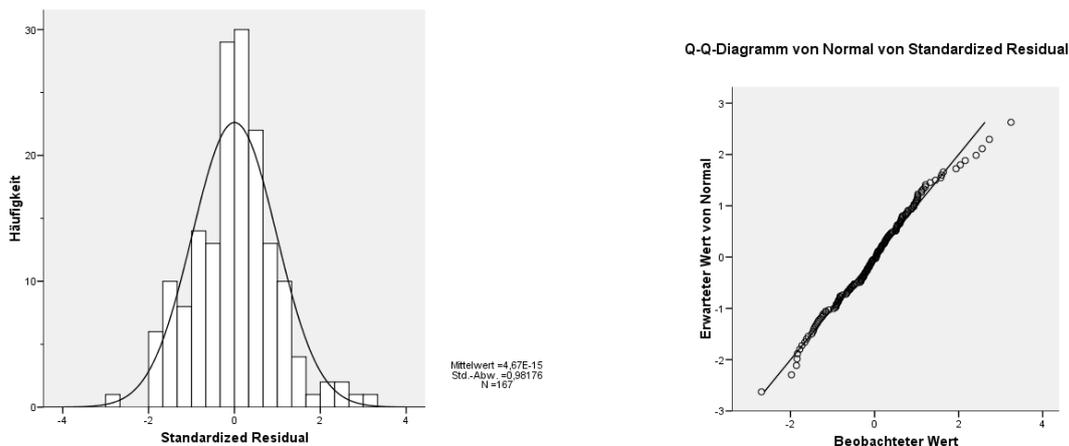


Abbildung 6.1: Histogramm und Q-Q-Plot der standardisierten Residuen

im Q-Q-Plot rechts aufgetragenen standardisierten Residuen zeigen diese annähernde Normalverteilung, mit Ausnahme der Tails.

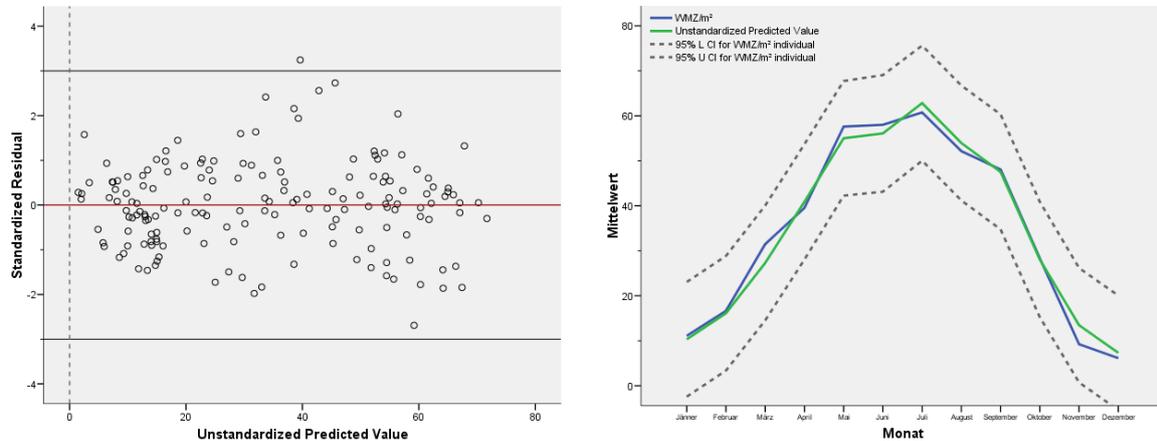


Abbildung 6.2: Residuenplot und Liniendiagramm WMZ/m^2 | geschätzt- WMZ/m^2

Die Daten im Residuenplot für die Prognose der Monaterträge in kWh/m^2 in Abbildung 6.2 links weisen eine annähernd zufällige Struktur auf, mit nur einem standardisierten Residuum größer als $3 kWh/m^2$ im Mai 2007 am Dach des Liebenauer Stadions. Dies bedeutet, dass keine grobe Verletzung der Modellvoraussetzungen vorliegt. Das Liniendiagramm rechts in dieser Abbildung zeigt den Vergleich der gemessenen Monaterträge in kWh/m^2 in blau mit den Vorhersagewerten in grün samt des 95%-Prognoseintervalls bezüglich der einzelnen Monate über alle Jahre. Die Prognosewerte aus dem Modell in Tabelle 6.3 liegen deutlich im Toleranzintervall.

6.1.2 Modelle mit horizontaler Strahlung (gemessen)

Beim Prognosemodell in Tabelle 6.4 ersetzen die gemessenen **horizontalen Strahlungswerte** Globalstrahlung [W/m^2] und Diffusanteil [%] die berechneten Werte E_{glob} [W/m^2] und E_{diff} [W/m^2] aus Tabelle 6.3 auf der vorherigen Seite. Dadurch gehen statt der Saison und des Neigungswinkels [°] die Regressoren Leitungslänge außen [m], geographische Breite [°] und Wärmekapazität [J/m^2K] in das Modell ein. Zusammen mit den identischen Regressoren Glykolgehalt [%] und Lufttemperatur [°C] verbessert sich das korrigierte R^2 von **90,9%** auf **92,0%**. Auch der Standardfehler des Schätzers sinkt von $6,34 kWh/m^2$ auf $5,945 kWh/m^2$.

6.1. LINEARE REGRESSIONSMODELLE MIT BERECHNETEN & VORHERSAGBAREN VARIABLEN

Tabelle 6.4: Prognosemodell mit vorhersagbaren Regressoren für alle Solaranlagen

Modellzusammenfassung^(b)

R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers	Durbin-Watson-Statistik	F	Signifikanz
,961(a)	0,924	0,920	5,945	1,267	275	$< 10^{-3}$

a. Einflussvariablen: (Konstante), Globalstrahlung [W/m²], Diffusanteil [%], Leitungslänge außen [m], geogr. Breite [°], Glykolgehalt [%], Wärmekapazität [J/m²K], Lufttemperatur [°C]
 b. Abhängige Variable: WMZ/m² [kWh/m²]; n = 167

Koeffizienten^(a)

	Nicht standardisierte Koeffizienten		Standardisierte Koeff.		Signifikanz	95%-Konfidenzintervall für B	
	B	Standardfehler	Beta	T		Untergrenze	Obergrenze
(Konstante)	17.602,99490	3.680,332		4,78	$< 10^{-3}$		
Globalstrahlung [W/m ²]	0,00025	0,000	0,629	12,95	$< 10^{-3}$	0,00022	0,00029
Diffusanteil [%]	-0,23509	0,055	-0,111	-4,27	$< 10^{-3}$	-0,34390	-0,12628
Leitungslänge außen [m]	-0,20987	0,039	-0,522	-5,35	$< 10^{-3}$	-0,28737	-0,13237
geogr. Breite [°]	-371,20986	77,721	-0,529	-4,78	$< 10^{-3}$	-524,70811	-217,71160
Glykolgehalt [%]	16,95302	5,041	1,034	3,36	0,001	6,99611	26,90992
Wärmekapazität [J/m ² K]	-0,10956	0,031	-1,214	-3,59	$< 10^{-3}$	-0,16990	-0,04922
Lufttemperatur [°C]	0,79545	0,131	0,285	6,09	$< 10^{-3}$	0,53761	1,05329

a. Abhängige Variable: WMZ/m² [kWh/m²]

DIAGNOSEPLOTS FÜR DIE GÜTE DES MODELLS

Die standardisierten Residuen aus dem Modell mit horizontaler Strahlung scheinen annähernd normalverteilt zu sein, wie aus Abbildung 6.3 sowohl im Histogramm rechts als auch im Q-Q-Plot links hervorgeht. Die Anpassung an die Normalverteilung scheint vor allem an den Tails eine bessere zu sein als beim Modell mit geneigter Strahlung (siehe Abbildung 6.1 auf Seite 67).

Die Daten im Residuenplot in Abbildung 6.4 links weisen eine annähernd zufällige Struktur ohne Ausreißer sowohl nach oben als auch nach unten auf. Der Vergleich zwischen den gemessenen Monatserträgen/m² und den Vorhersagewerten im Liniendiagramm rechts zeigt eine gute Anpassung und die Vorhersagewerte liegen deutlich im Toleranzintervall. Auch hier zeigt sich, dass das Modell mit den gemessenen horizontalen Strahlungswerten besser zu sein scheint als jenes mit den berechneten geneigten Werten, wie in Abbildung 6.2 auf der vorherigen Seite rechts dargestellt. Die Prognose

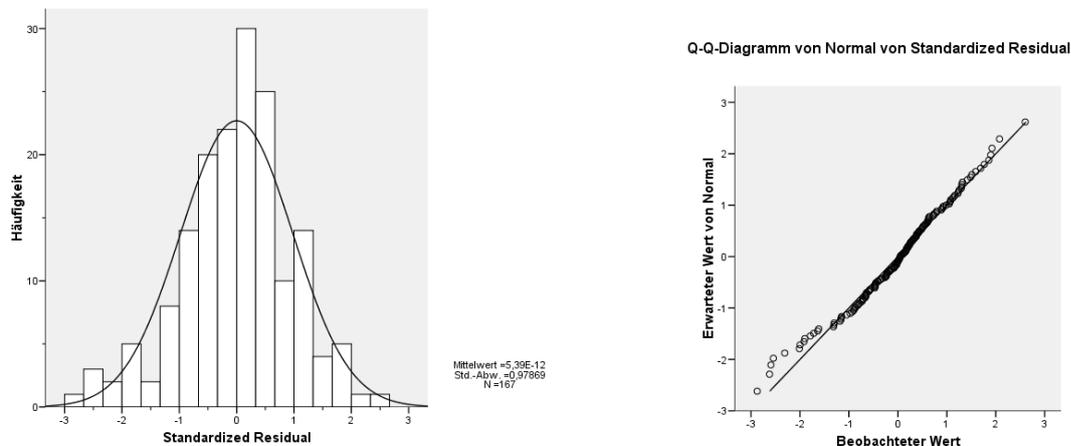


Abbildung 6.3: Histogramm und Q-Q-Plot der standardisierten Residuen

se verbessert sich sogar ein wenig für die Monate März, Mai, August und November, wogegen sie sich nur im September verschlechtert.

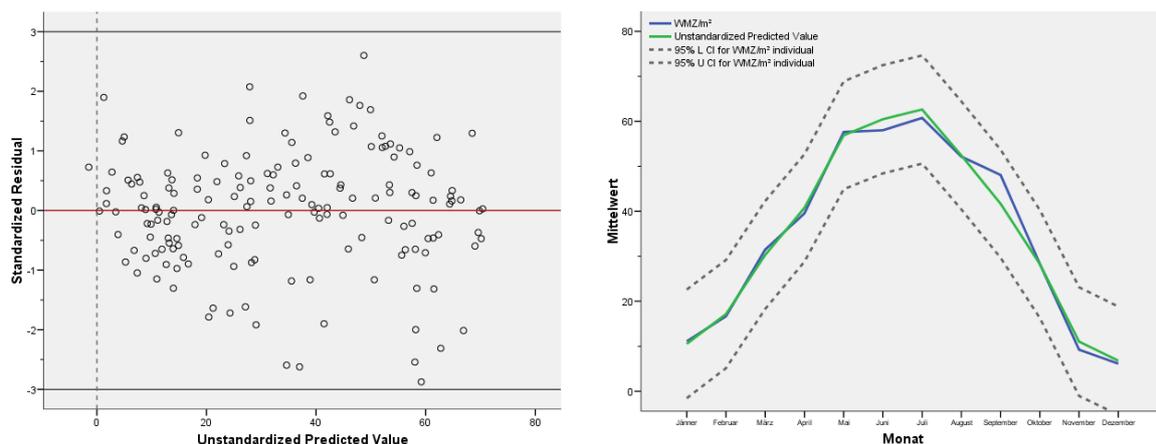


Abbildung 6.4: Residuenplot und Liniendiagramm WMZ/m^2 | geschätzt- WMZ/m^2

Im Prognosemodell in Tabelle 6.4 auf der vorherigen Seite ist der zu starke Einfluss der geographischen Breite - in Kapitel 5.1.2 erklärt - deutlich ersichtlich, weshalb für die Prognose der Monatserträge in kWh/m^2 ein Modell **ohne geographische Koordinaten** verwendet wird, siehe Tabelle 6.5. Nach dieser Herausnahme fällt ebenfalls der Glykolgehalt [%] als nicht signifikant aus dem Modell, wogegen der Neigungswinkel [°] als neuer Regressor ins Modell eingeht. Auch hier ist wie bei den Tagesprognosen vor allem die starke Veränderung der Konstante bemerkbar. Das korrigierte R^2 verschlechtert sich nur geringfügig von **92%** auf **91,6%** und der Standardfehler des Schätzers

vergrößert sich von 5,945 kWh/m² auch nur auf 6,118 kWh/m².

Tabelle 6.5: Prognosemodell ohne geographische Koordinaten der Solaranlagen

Modellzusammenfassung^(b)

R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers	Durbin-Watson-Statistik	F	Signifikanz
,959(a)	0,919	0,916	6,118	1,217	302	< 10 ⁻³

a. Einflussvariablen: (Konstante), Globalstrahlung [W/m²], Diffusanteil [%], Neigungswinkel [°], Leitungslänge außen [m], Wärmekapazität [J/m²K], Lufttemperatur [°C]
 b. Abhängige Variable: WMZ/m² [kWh/m²]; n = 167

Koeffizienten^(a)

	Nicht standardisierte Koeffizienten	Standardisierte Koeff.	T	Signifikanz	95%-Konfidenzintervall für B	
	B	Standardfehler	Beta		Untergrenze	Obergrenze
(Konstante)	88,85521	19,065		4,66	< 10 ⁻³	
Globalstrahlung [W/m ²]	0,00025	0,000	0,631	12,63	< 10 ⁻³	0,00021 0,00029
Diffusanteil [%]	-0,23182	0,057	-0,109	-4,09	< 10 ⁻³	-0,34384 -0,11980
Neigungswinkel [°]	-0,60027	0,124	-0,152	-4,85	< 10 ⁻³	-0,84493 -0,35561
Leitungslänge außen [m]	-0,06307	0,011	-0,157	-5,78	< 10 ⁻³	-0,08464 -0,04151
Wärmekapazität [J/m ² K]	-0,00882	0,003	-0,098	-3,44	0,001	-0,01388 -0,00376
Lufttemperatur [°C]	0,79239	0,134	0,284	5,89	< 10 ⁻³	0,52681 1,05796

a. Abhängige Variable: WMZ/m² [kWh/m²]

Die Diagnoseplots zur Güte des Modells zeigen ebenfalls kaum Veränderungen zu den Abbildungen 6.3 und 6.4 und werden daher hier weggelassen.

6.2 Lineare Regressionsmodelle ohne Kenntnis der Anlagedaten

Wie in Kapitel 5.2 auf Seite 59 finden im Folgenden die Anlagedaten der sechs Solaranlagen in und um Graz bei der Berechnung der Prognosemodelle für den Monatsertrag in kWh/m² keine Berücksichtigung.

6.2.1 Modelle mit geneigter Strahlung auf Kollektorebene (berechnet)

Das korrigierte R^2 fällt im Vergleich zum Modell in Tabelle 6.3 auf Seite 67 beim Prognosemodell mit den Regressoren Saison, Ediff [W/m²], Eglob [W/m²] und Lufttemperatur [°C] um nur 1,1% von **90,9%** auf **89,8%**. Auch der Standardfehler des Schätzers wird von 6,34 kWh/m² auf 6,717 kWh/m² nur geringfügig größer.

Tabelle 6.6: Prognosemodell mit den Regressoren Saison, Ediff, Eglob und Lute

Modellzusammenfassung ^(b)							
R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers	Durbin-Watson-Statistik	F	Signifikanz	
,949(a)	0,901	0,898	6,717	1,078	368	< 10 ⁻³	
a. Einflussvariablen: (Konstante), Saison, Ediff [W/m ²], Eglob [W/m ²], Lufttemperatur [°C]							
b. Abhängige Variable: WMZ/m ² [kWh/m ²]; n = 167							
Koeffizienten ^(a)							
	Nicht standardisierte Koeffizienten	Standardisierte Koeff.			95%-Konfidenzintervall für B		
	B	Standardfehler	Beta	T	Signifikanz	Untergrenze	Obergrenze
(Konstante)	17,67199	4,610		3,83	< 10 ⁻³		
Saison	-7,99423	2,085	-0,190	-3,83	< 10 ⁻³	-12,11127	-3,87720
Ediff [W/m ²]	-0,00019	0,000	-0,178	-2,40	0,018	-0,00035	-0,00003
Eglob [W/m ²]	0,00030	0,000	0,623	8,71	< 10 ⁻³	0,00023	0,00037
Lufttemperatur [°C]	1,08223	0,143	0,388	7,55	< 10 ⁻³	0,79909	1,36537
a. Abhängige Variable: WMZ/m ² [kWh/m ²]							

DIAGNOSEPLOTS FÜR DIE GÜTE DES MODELLS

Das Histogramm in Abbildung 6.5 links weist eine recht gute Anpassung der standardisierten Residuen an die eingezeichnete Normalverteilungskurve auf. Die standardisierten Residuen im Q-Q-Plot scheinen annähernd normalverteilt zu sein, mit Ausnahme der Tails. Der Vergleich mit Abbildung 6.1 auf Seite 67 zeigt sogar einen bessere Anpassung an die Modellvoraussetzungen.

6.2. LINEARE REGRESSIONSMODELLE OHNE KENNTNIS DER ANLAGEDATEN

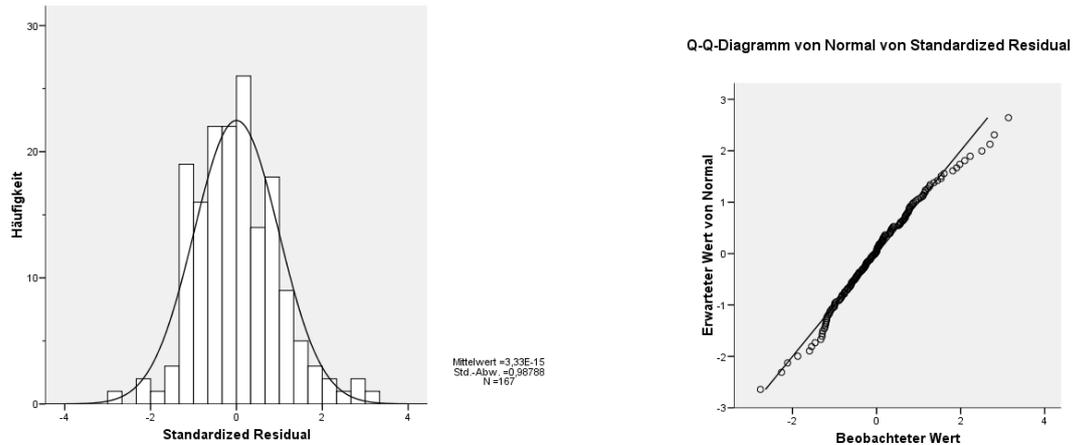


Abbildung 6.5: Histogramm und Q-Q-Plot der standardisierten Residuen

Die Daten im Residuenplot für die Prognose der Monatserträge in kWh/m² in Abbildung 6.6 links weisen eine annähernd zufällige Struktur auf, mit wieder nur einem standardisierten Residuum größer als 3 kWh/m² im Mai 2007 am Dach des Liebenauer Stadions. Der Vergleich zwischen den gemessenen Monatserträgen/m² und den Vorhersagewerten im Liniendiagramm rechts zeigt eine gute Anpassung und die Vorhersagewerte liegen deutlich im Toleranzintervall.

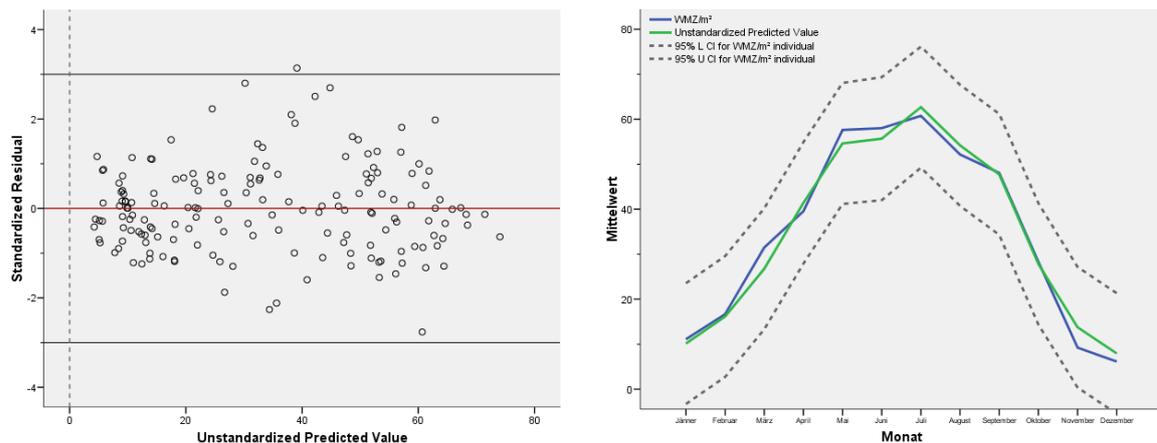


Abbildung 6.6: Residuenplot und Liniendiagramm WMZ/m² | geschätzt-WMZ/m²

6.2.2 Modelle mit horizontaler Strahlung (gemessen)

Werden aus dem Modell mit den gemessenen **horizontalen Strahlungswerten** in Tabelle 6.5 auf Seite 71 die Anlagedaten Neigungswinkel [°], Leitungslänge außen [m] und Wärmekapazität [J/m²K] genommen, so verringert sich das korrigierte R^2 um nur 1,8% von **91,6%** auf **89,8%**. Auch der Standardfehler des Schätzers vergrößert sich beim Prognosemodell mit den drei verbliebenen Regressoren Globalstrahlung [W/m²], Diffusanteil [%] und Lufttemperatur [°C] von 6,118 kWh/m² auf 6,735 kWh/m² nur geringfügig.

Der Vergleich der beiden Modelle zeigt einzig eine Veränderung der Konstante von 89 kWh/m² auf 12 kWh/m².

Tabelle 6.7: Prognosemodell mit den Regressoren Globalstrahlung, Diffusanteil und Lufttemperatur

Modellzusammenfassung ^(b)						
R	R-Quadrat	Korrigiertes R-Quadrat	Standardfehler des Schätzers	Durbin-Watson-Statistik	F	Signifikanz
,949(a)	0,900	0,898	6,735	0,968	487	< 10 ⁻³
a. Einflussvariablen: (Konstante), Globalstrahlung [W/m ²], Diffusanteil [%], Lufttemperatur [°C]						
b. Abhängige Variable: WMZ/m ² [kWh/m ²]; n = 167						

Koeffizienten ^(a)								
	Nicht standardisierte Koeffizienten	Standardfehler	Standardisierte Koeff.	Beta	T	Signifikanz	95%-Konfidenzintervall für B	
							Untergrenze	Obergrenze
(Konstante)	12,22485	4,030			3,03	0,003		
Globalstrahlung [W/m ²]	0,00027	0,000	0,665	12,28	< 10 ⁻³	0,00023	0,00031	
Diffusanteil [%]	-0,23154	0,062	-0,109	-3,73	< 10 ⁻³	-0,35429	-0,10880	
Lufttemperatur [°C]	0,67648	0,146	0,243	4,65	< 10 ⁻³	0,38899	0,96397	
a. Abhängige Variable: WMZ/m ² [kWh/m ²]								

Dieses Modell wird in Kombination mit dem Modell aus Tabelle 6.5 auf Seite 71 für den Excel Monatsertrags-Prognosefile „**PROGNOSE_Monat**“ verwendet (siehe Anhang A.2).

DIAGNOSEPLOTS FÜR DIE GÜTE DES MODELLS

Die standardisierten Residuen aus dem Modell mit horizontaler Strahlung für die Monatserträge in kWh/m² scheinen mit Ausnahme der Tails annähernd normalverteilt, wie in Abbildung 6.7 sowohl im Histogramm links als auch im Q-Q-Plot rechts ersichtlich.

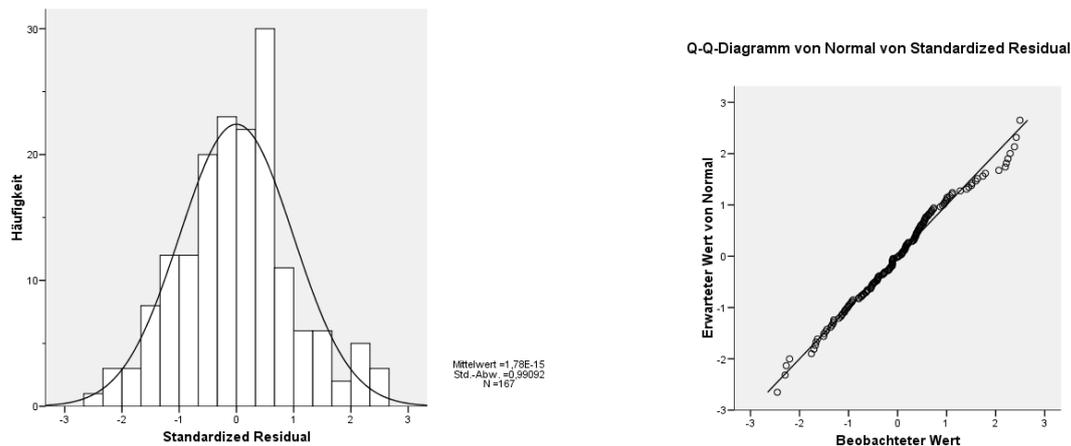


Abbildung 6.7: Histogramm und Q-Q-Plot der standardisierten Residuen

Die Daten im Residuenplot in Abbildung 6.8 links weisen eine annähernd zufällige Struktur auf. Der Vergleich zwischen den gemessenen Monatserträgen/m² und den Vorhersagewerten zeigt eine gute Anpassung und die Vorhersagewerte liegen deutlich im Toleranzintervall.

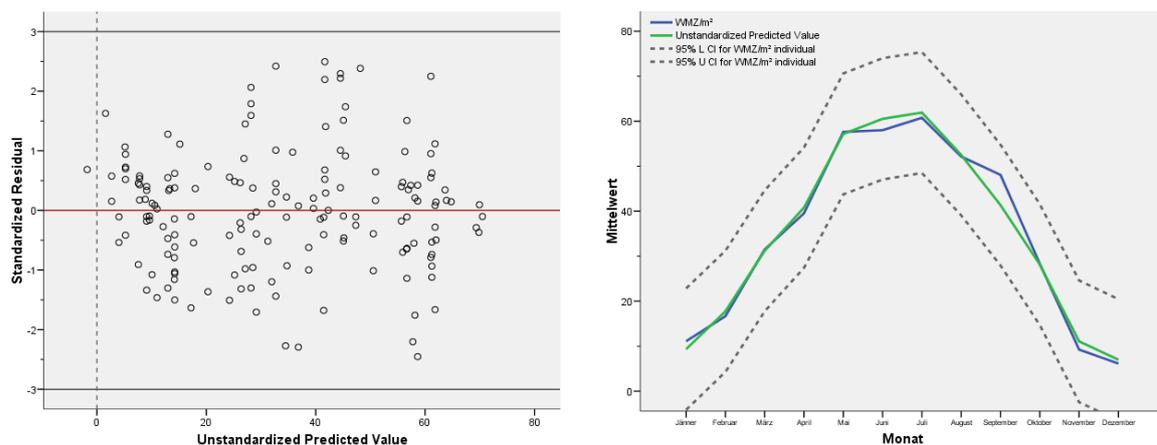


Abbildung 6.8: Residuenplot und Liniendiagramm WMZ/m² | geschätzt-WMZ/m²

Die partiellen Regressionsdiagramme (vgl. Seite 102) in Abbildung 6.9 zeigen positive Zusammenhänge der Globalstrahlung $[W/m^2]$ und Lufttemperatur $[^\circ C]$ zum Tagesertrag $[kWh/m^2]$. Der Diffusanteil $[\%]$ weist hingegen einen negativen Zusammenhang zum Tagesertrag auf.

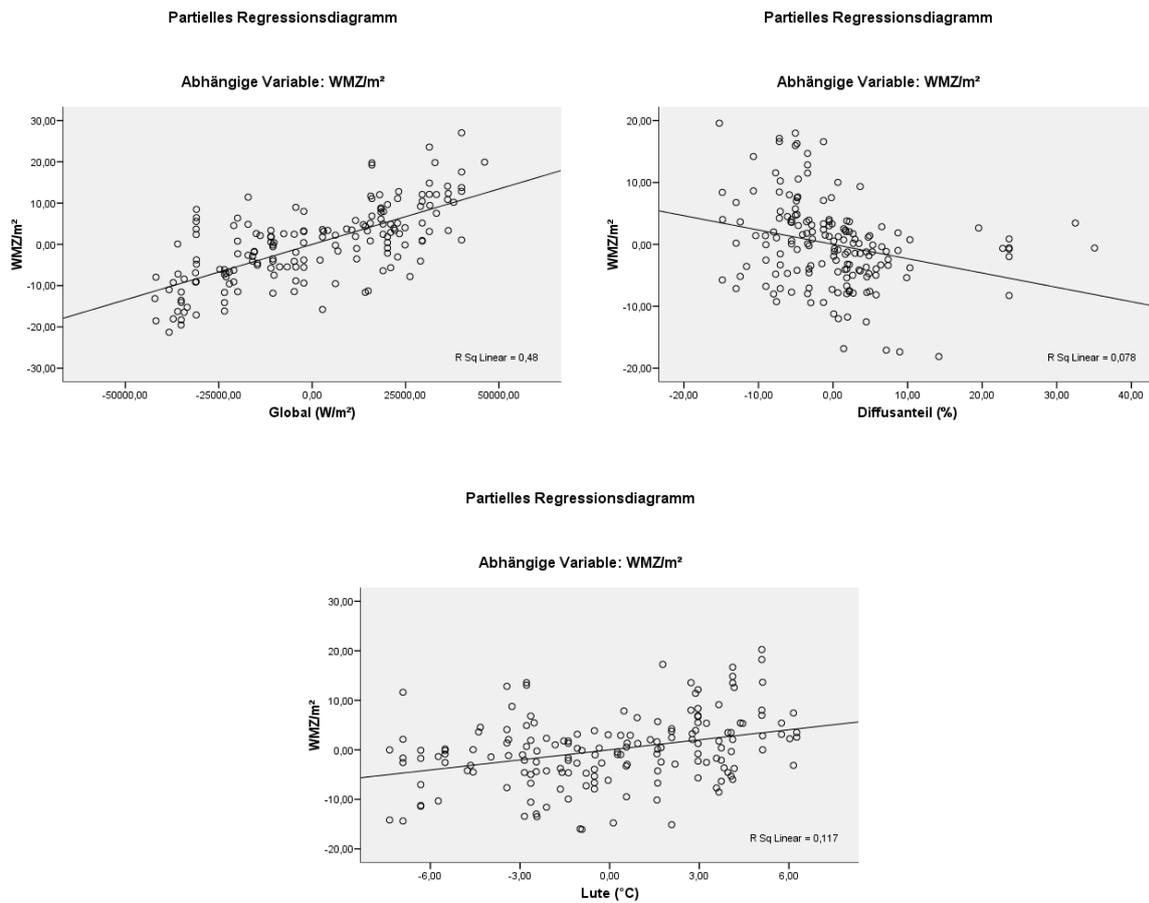


Abbildung 6.9: Partielle Regressionsdiagramme der Regressoren und WMZ/m^2

7 Mathematische und Statistische Grundlagen^{6,7,8,9,10,11,12,13,14,15,16,17}

Dieses Kapitel soll einen Einblick in die für diese Arbeit verwendeten statistischen Methoden geben. Dem mathematisch versierten Leser mögen diese theoretischen Grundlagen vertraut sein. In Anbetracht der Tatsache, dass es sich hierbei jedoch um eine „Industrie-Diplomarbeit“ in Kooperation mit der Firma S.O.L.I.D. GmbH handelt, werden im Folgenden die mathematischen und statistischen Grundlagen derart erklärt, dass auch einem mit dem Thema nicht vertrauten Leser ein Zugang ermöglicht wird.

7.1 Daten, Kenngrößen und Graphiken

7.1.1 Datenaspekt

Vor jeder statistischen Analyse von konkreten Daten sollte deren Messniveau und die daraus resultierenden Gegebenheiten genau betrachtet werden. Der Ausgangspunkt sind Daten in Form von Messreihen, die als *Stichproben* aus einer größeren Einheit (*Grundgesamtheit*) aufgefasst werden können.

⁶[1] Bosius, 1998

⁷[2] Bühl, Zöfel, 1996

⁸[3] Fahrmeir, Künstler, Pigeot, Tutz, 1999

⁹[4] Friedl, 1993

¹⁰[5] Hussain, 1998

¹¹[6] Müller, 2010

¹²[8] Sponer, 2009

¹³[9] Stadlober, 2001-2011

¹⁴[10] Stadlober, 2011

¹⁵[11] Stadlober, 2011

¹⁶[12] Stadlober & Schauer, 2007-2011

¹⁷<http://www.uni-potsdam.de/u/soziologie/methoden/mitarbeiter/shk/Michael/2/index.htm>

Diskrete und stetige Merkmale

Kann ein Merkmal nur endlich oder abzählbar unendlich viele Ausprägungen annehmen, so heißt das Merkmal **diskret**. Kann ein Merkmal hingegen alle Werte eines Intervalls annehmen, heißt es **stetig**.

Metrisches Merkmal

Als **metrische** Merkmale werden Merkmalsausprägungen bezeichnet, die aus konkreten Zahlen bestehen, eine Dimension und einen Nullpunkt besitzen und deren Eigenschaften anhand der Darstellung von quantitativen Daten erfolgt. Metrische Merkmale können auch als **kardinale** oder **quantitative** Merkmale bezeichnet werden und beinhalten die Daten aus einer *Intervall-* oder *Verhältnisskala*.

Nur mit metrischen Merkmalen lassen sich alle Rechenoperationen wirklich aussagekräftig durchführen und etwa Mittelwerte, Korrelationen oder Regressionen berechnen. Sie sind daher auch die Grundlage für die Anwendung der Normalverteilung.¹⁸

Repräsentative Stichprobe

Das Merkmal X sei eine *Zufallsvariable* mit der Verteilung F auf der Grundgesamtheit. Ein Stichprobenvektor (X_1, \dots, X_n) heißt **repräsentativ**, wenn $X_i \stackrel{iid}{\sim} F$, d.h. wenn alle X_i dieselbe Verteilung F besitzen (Verteilung der Grundgesamtheit) und stochastisch unabhängig sind (Zufallsauswahl).¹⁹

Die konkrete Auswahl einer Stichprobe liefert eine bestimmte *Realisierung* (x_1, \dots, x_n) des Stichprobenvektors (X_1, \dots, X_n) .

Schätzfunktion

Eine **Schätzfunktion** dient dazu, aufgrund von vorhandenen Daten einer Stichprobe einen *Schätzwert* zu ermitteln und dadurch Informationen über unbekannte Parameter einer Grundgesamtheit zu erhalten (vgl. Seite 86). Schätzfunktionen sind die Basis zur Berechnung von Konfidenzintervallen und Teststatistiken in Hypothesentests. Sie

¹⁸<http://www.bwl-bote.de/20080627.htm>

¹⁹*iid = independent and identically distributed*

sind spezielle *Stichprobenfunktionen*, also Zufallsvariablen, die Funktionen des Stichprobenvektors (X_1, \dots, X_n) sind, und können durch Schätzverfahren, z.B. die Maximum-Likelihood-Methode, die Methode der kleinsten Quadrate oder die Momentenmethode, bestimmt werden.²⁰

7.1.2 Statistische Kenngrößen

Es seien x_1, \dots, x_n Realisationen der Stichprobe X_1, \dots, X_n und $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ sei Realisation der geordneten Stichprobe $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$.

Quantile

Sei $(x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)})$ eine geordnete Stichprobe, so heißt

$$q_p = (1 - g) x_{(\lfloor (n+1)p \rfloor)} + g x_{(\lfloor (n+1)p \rfloor + 1)} \quad \text{mit} \quad g = (n + 1)p - \lfloor (n + 1)p \rfloor$$

mit den Spezialfällen

- $(n + 1)p$ ganzzahlig $\Rightarrow g = 0: q_p = x_{(\lfloor (n+1)p \rfloor)}$
- np ganzzahlig $\Rightarrow g = p: q_p = (1 - p) x_{(\lfloor (n+1)p \rfloor)} + p x_{(\lfloor (n+1)p \rfloor + 1)}$

p -tes Quantil ($0 < p < 1$) dieser Stichprobe (Berechnungsformel bei SPSS).

Das Quantil q_p hat die Eigenschaft, dass *mindestens* ein relativer Anteil von p Daten *kleiner gleich* q_p und *höchstens* ein relativer Anteil von $(1 - p)$ *größer als* q_p ist.

Arithmetisches Mittel (Mittelwert)

Sei (x_1, \dots, x_n) eine Stichprobe aus *metrischen* Daten, so heißt

$$\bar{x} = \frac{1}{n}(x_1 + \dots + x_n) = \frac{1}{n} \sum_{i=1}^n x_i$$

arithmetisches Mittel dieser Stichprobe und ist ein Maß für die Lage (Lokation).

²⁰<http://de.wikipedia.org/wiki/Schätzfunktion>

Median (Zentralwert)

Sei $(x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)})$ eine geordnete Stichprobe, so heißt

$$x_{med} = \tilde{x} = q_{0,5} = \begin{cases} x_{(\frac{n+1}{2})}, & \text{falls } n \text{ ungerade} \\ \frac{1}{2} \left(x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)} \right), & \text{falls } n \text{ gerade} \end{cases}$$

Median dieser Stichprobe und ist ein Maß für die Lage (Lokation).

Er gehört zur Gruppe der Quantile, kann auch als 0,5-Quantil bzw. 2. Quartil bezeichnet werden, halbiert eine geordnete Stichprobe und hat, verglichen mit dem arithmetischen Mittel, den Vorteil **robuster** gegenüber Ausreißern zu sein.

Varianz, Standardabweichung und Standardfehler

Sei (x_1, \dots, x_n) eine Stichprobe aus *metrischen* Daten, so heißen

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

Stichprobenvarianz und

$$s = +\sqrt{s^2}$$

Standardabweichung dieser Stichprobe. Beide sind Maße für die Streuung (Skalierung).

Weiters gibt

$$\frac{s}{\sqrt{n}}$$

den **Standardfehler** des arithmetischen Mittels an.

Korrelationskoeffizient von Pearson

Sei $(x_1, y_1), \dots, (x_n, y_n)$ eine paarige Stichprobe von *metrischen* Merkmalen X und Y , so heißt

$$r = r_{XY} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} = \frac{s_{XY}}{s_X s_Y} = r_{YX}$$

Korrelationskoeffizient von Pearson, wobei

$$s_X = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \quad \text{und} \quad s_Y = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2}$$

für die *empirischen Standardabweichungen* der Merkmale X und Y stehen und

$$s_{XY} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

die *empirische Kovarianz* von (X, Y) bezeichnet.

Partielle Korrelation

Sei $(x_1, y_1, z_1), \dots, (x_n, y_n, z_n)$ eine Stichprobe von *metrischen* Merkmalen X , Y und Z , so heißt

$$r_{XY.Z} = \frac{r_{XY} - r_{XZ} \times r_{YZ}}{\sqrt{1 - r_{XZ}^2} \sqrt{1 - r_{YZ}^2}}$$

partieller Korrelationskoeffizient von X und Y unter Z .

Die Partialkorrelation ist ein Verfahren, mit dem ein linearer Zusammenhang zwischen zwei Variablen X und Y um den linearen Einfluss einer Drittvariablen Z bereinigt wird.

Eigenschaften:

- Die Gesamtkorrelation geht nur auf den Einfluss der Drittvariablen Z zurück, wenn $r_{XY.Z} = 0$ und $r_{XY} \neq 0$.
- Die Drittvariable Z hat keinen Einfluss, wenn $r_{XY.Z} = r_{XY}$.

7.1.3 Explorative Graphiken

Boxplot

Der **Boxplot** ist eine Summationsdarstellung der Verteilung der Daten durch 5 Kenngrößen, die alle auf *Quantilen* basieren. Die Box beinhaltet die zentralen 50% der

Daten, die definitionsgemäß im interquantilen Bereich $iqr = q_{0,75} - q_{0,25}$ liegen. Der Median x_{med} ist als horizontaler Strich dargestellt. Die Länge der Linien unter- und oberhalb der Box sind höchstens $1,5 \times iqr$. Liegen *alle* Datenpunkte *innerhalb* dieses Bereichs, dann erstrecken sich die Linien nur bis zum *minimalen Wert* $x_{(1)}$ bzw. *maximalen Wert* $x_{(n)}$. Ansonsten werden die Linien durch folgende Datenpunkte begrenzt:

- Unterhalb der Box: $\min_{\ell} x_{(\ell)}$ mit $x_{(\ell)} \geq q_{0,25} - 1,5 \times iqr$
- Oberhalb der Box: $\max_k x_{(k)}$ mit $x_{(k)} \leq q_{0,75} + 1,5 \times iqr$.

Datenpunkte *außerhalb* dieses Bereichs werden als *Ausreißer* bezeichnet und durch „*“ oder „o“ markiert.

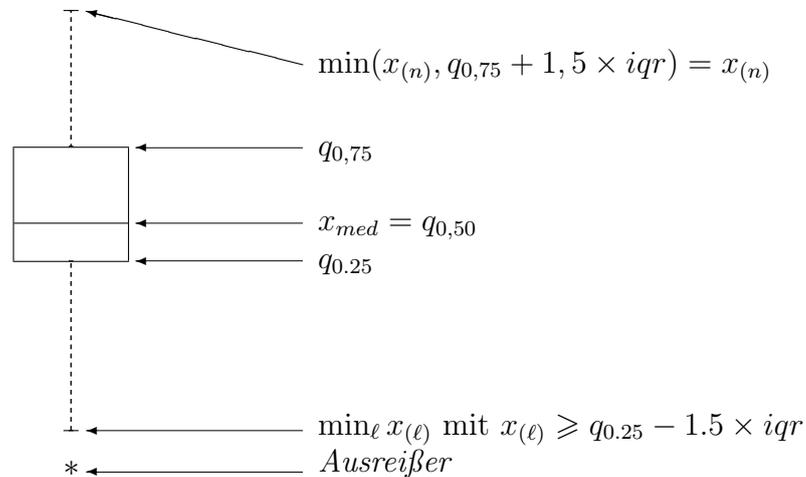


Abbildung 7.1: Definition des Boxplots²¹

Das Statistikprogramm SPSS kennzeichnet Werte, die um mehr als drei iqr -Längen außerhalb der Box liegen, mit einem Stern und solche, die um mehr als anderthalb iqr -Längen außerhalb liegen, mit einem Kreis.

Fehlerbalken

Im Fehlerbalken werden die Daten durch Standardgrößen (Mittelwert \bar{x} für Lokation, Standardabweichung s oder 95%-Konfidenzintervalle des Mittelwertes als Streubereiche) charakterisiert.

²¹[12] Stadlober & Schauer, 2011, S. 35

Streudiagramme (Scatterplots) und Scatterplotmatrizen

Streudiagramme oder Scatterplots geben paarige Stichproben (x_i, y_i) , $i = 1, \dots, n$ als Punkte, Kreuze, Kreise oder sonstige Symbole in einem $(x - y)$ -Koordinatensystem wieder. Mehrere Paare von Stichproben können in einer Scatterplotmatrix zusammengefasst werden.

Quantil-Quantil-Plot (Q-Q-Plot)

Mit Hilfe des Q-Q-Plots kann die Anpassung einer theoretischen Verteilung an die gegebene Stichprobe beurteilt werden, wobei das Idealmuster (Referenz) eine Gerade darstellt. Bestimmte Abweichungen von der Geraden können verteilungsmäßig charakterisiert werden.

Es werden die geordneten Stichprobenwerte

$$x_{(i)} \quad \text{gegen die theoretischen Quantile} \quad F^{-1}\left(\frac{i}{n+1}\right), \quad i = 1, \dots, n$$

aufgetragen. Wenn F eine Verteilungsfunktion mit $F(x) = G\left(\frac{x-\mu}{\sigma}\right)$ ist, dann kann man zeigen, dass

$$X_{(i)} \approx F^{-1}\left(\frac{i}{n+1}\right) = \sigma G^{-1}\left(\frac{i}{n+1}\right) + \mu$$

gilt.

In der Praxis werden die Parameter μ und σ durch die Schätzungen \bar{x} und s ersetzt. Die Statistiksoftware **SPSS** trägt das Paar $(x_{(i)}, F^{-1}\left(\frac{i}{n+1}\right))$ und das Statistikprogramm **R** das Paar $(x_{(i)}, G^{-1}\left(\frac{i}{n+1}\right))$ auf. Im Falle der Normalverteilung hat man $G = \Phi \equiv N(0, 1)$, bzw. $F \equiv N(\bar{x}, s)$.

Histogramm

Der gesamte Bereich $[x_{(1)}, x_{(n)}]$ der geordneten Stichprobe $(x_{(1)} \leq \dots \leq x_{(n)})$ wird in k äquidistante Klassen mit Breite b aufgeteilt: $[c_0 \leq x_{(1)}, c_1)$, $[c_1, c_2)$, \dots , $[c_{k-1}, c_k \geq x_{(n)})$, wobei $c_j = c_0 + j \times b$, $j = 0, \dots, k$.

Das **Histogramm** ergibt sich als Zusammensetzung von k Rechtecken mit Breite b und Länge h_j , die über den entsprechenden Klassen $[c_{j-1}, c_j)$ aufgetragen werden:

$$h(x) = \begin{cases} h_j & c_{j-1} \leq x < c_j, j = 1, \dots, k; c_j - c_{j-1} = b \\ 0 & \text{sonst} \end{cases} .$$

Die Wahl von h_j richtet sich nach der gewünschten Skalierung. Sei n_j mit $\sum_{j=1}^k n_j = n$ die Anzahl der Datenpunkte in der j -ten Klasse und $H = \int_{c_0}^{c_k} h(x) dx$ die Fläche des Histogramms. Folgende Definitionen von h_j sind dann gebräuchlich:

$$h_j = \begin{cases} n_j & \text{absolute Häufigkeiten} & H = nb \\ f_j = \frac{n_j}{n} & \text{relative Häufigkeiten} & H = b \\ \frac{n_j}{nb} & \text{Dichteschätzer} & H = 1 \end{cases} .$$

7.2 Stetige Verteilungen

7.2.1 Dichte und Verteilungsfunktion

Die Verteilung einer *stetigen* Zufallsvariablen X wird durch eine stückweise stetige Funktion $f(x) \geq 0$, der (Wahrscheinlichkeits-) **Dichte** oder **Dichte** (-funktion), charakterisiert. Die Wahrscheinlichkeit, dass X einen Wert aus dem Intervall $[a, b]$ annimmt, lässt sich durch

$$P(a \leq X \leq b) = \int_a^b f(x) dx$$

berechnen. Die Gesamtfläche zwischen x-Achse und der Dichte $f(x)$ ist gleich 1, also

$$\int_{-\infty}^{\infty} f(x) dx = 1.$$

Die Funktion $F : \mathbb{R} \rightarrow [0, 1]$ mit

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(t) dt \quad \text{für alle } x \in \mathbb{R}$$

heißt **Verteilungsfunktion** von X und ist eine monoton steigende, stetige Funktion. Daher existiert auch die inverse Funktion (Umkehrfunktion) $x = F^{-1}(u)$ und

$x_\alpha = F^{-1}(\alpha)$ ist das (theoretische) α -Quantil. Links von x_α liegt genau die Wahrscheinlichkeitsmasse α , d.h.

$$F(x_\alpha) = P(X \leq x_\alpha) = \int_{-\infty}^{x_\alpha} f(x) dx = \alpha.$$

7.2.2 Normalverteilung $N(\mu, \sigma)$

Normalverteilungen bilden eine Klasse von besonders wichtigen Dichtefunktionen. Sie sind symmetrisch, unimodal und glockenförmig. Sie heißen nach dem deutschen Mathematiker Gauß, der ihre Bedeutung erkannte, auch **Gauß-Verteilungen**.

Eine stetige Zufallsvariable X mit *Lokationsparameter* $\mu \in \mathbb{R}$ und *Skalierungsparameter* $\sigma > 0$ mit der Dichte

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, \quad x \in \mathbb{R},$$

heißt **normalverteilt**, kurz $X \sim N(\mu, \sigma)$.

Die *standardisierte* Zufallsvariable $Z = \frac{X-\mu}{\sigma}$ besitzt die Dichtefunktion

$$f_Z(z) = \phi(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$$

einer Normalverteilung mit $\mu = 0$ und $\sigma = 1$. Diese Normalverteilung heißt **Standardnormalverteilung** und die Variable Z entsprechend **standardnormalverteilt**, d.h. $Z \sim N(0, 1)$.

Lineare Transformation

Für $X \sim N(\mu, \sigma)$ ist die linear transformierte Variable $Y = aX + b$ wieder normalverteilt mit

$$Y \sim N(a\mu + b, a\sigma).$$

Zentraler Grenzwertsatz

Seien X_1, \dots, X_n identisch verteilte unabhängige Zufallsvariable mit $\mu = E(X_i)$, $\sigma^2 = \text{Var}(X_i) > 0$ und sei $S_n = \sum_{i=1}^n X_i$. Dann strebt die Verteilung der standardisierten

Summe

$$Z_n = \frac{S_n - E(S_n)}{\sqrt{\text{Var}(S_n)}} \quad \text{gegen die } N(0, 1)\text{-Verteilung,}$$

wobei $E(S_n) = n\mu$ und $\sqrt{\text{Var}(S_n)} = \sigma\sqrt{n}$.

Für das arithmetische Mittel \bar{X} gilt, dass $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ asymptotisch $N(\mu, \frac{\sigma}{\sqrt{n}})$ -verteilt ist und

$$\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \xrightarrow{n \rightarrow \infty} N(0, 1).$$

Aus diesem Grund kann das arithmetische Mittel von unabhängigen Zufallsvariablen X_1, \dots, X_n als approximativ normalverteilt betrachtet werden.

Freiheitsgrade

In der Statistik gibt der Begriff **Freiheitsgrad** darüber Auskunft, aus wie vielen unabhängigen $N(0, 1)$ -verteilten Zufallsvariablen die betrachtete Zufallsgröße zusammengesetzt ist.

7.3 Parameterschätzung

7.3.1 Punktschätzung

Schätzfunktion

Schätzt eine Funktion $T = T(X_1, \dots, X_n)$ der Stichprobenvariablen X_1, \dots, X_n aus einer Verteilung F einen unbekanntem Parameter θ , so heißt T **Schätzfunktion** (*Schätzer*) und der aus den Realisationen x_1, \dots, x_n resultierende numerische Wert $t = t(x_1, \dots, x_n)$ ist der zugehörige *Schätzwert*.

Erwartungstreue und Bias

Sei X_1, \dots, X_n eine Stichprobe aus der Verteilung F mit Parameter θ und $T = T(X_1, \dots, X_n)$ ein Schätzer für den Parameter θ . Dann heißt der Schätzer T **erwartungstreu** (asym-

ptotisch erwartungstreu), wenn

$$E(T) = \theta \quad \left(\lim_{n \rightarrow \infty} E(T) = \theta \right)$$

gilt. Sonst heißt der Schätzer *verzerrt* (biased) und der **Bias** (*systematische Fehler*) von T ist definiert durch

$$b(T, \theta) := E(T) - \theta.$$

Konsistenz

Eine Folge von Schätzern, $(T_n)_{n \in \mathbb{N}}$ heißt (schwach) **konsistent**, falls gilt:

$$\lim_{n \rightarrow \infty} P(|T_n - \theta| > \varepsilon) = 0 \quad \text{für alle } \varepsilon > 0.$$

Effizient

Der Schätzer T heißt **effizient** (*wirksam*), wenn für alle anderen Schätzer T' von θ

$$E((T - \theta)^2) \leq E((T' - \theta)^2)$$

gilt, d.h. T hat den kleinsten quadratischen Fehler $E((T - \theta)^2)$. Für erwartungstreue Schätzer stimmt der quadratische Fehler mit der Varianz überein.

Momentenschätzer

Sei $X_i \stackrel{iid}{\sim} F$, dann ist

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

erwartungstreu, *konsistenter* und *wirksamster* linearer Schätzer für den **Erwartungswert** $E(X_i) = \mu$ (= erstes Moment um Null), falls $Var(X_i) < \infty$ und

$$S^2 = \frac{1}{n-1} \sum (X_i - \bar{X})^2$$

erwartungstreuer und *konsistenter* Schätzer für die **Varianz** $\text{Var}(X_i) = \sigma^2$ (= zweites zentrales Moment), falls $E(X_i^4) < \infty$.

Die **Kovarianz** σ_{XY} ist das zentrale Moment der Ordnung 2 und hat als Schätzer

$$S_{XY} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}).$$

Als Schätzer für den **Korrelationskoeffizienten** $\rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$ bekommt man

$$R = R_{XY} = \frac{S_{XY}}{S_X S_Y} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}}.$$

7.3.2 Intervallschätzung

Konfidenzintervalle

Sei X_1, \dots, X_n eine Stichprobe aus der Verteilung F mit unbekanntem Parameter θ . Dann heißt das zufällige Intervall $[K_1 \leq \theta \leq K_2]$ zweiseitiges $(1-\alpha)$ -Konfidenzintervall (KONF, KIV) für den Parameter θ , falls für alle θ

$$P(K_1 \leq \theta \leq K_2) \geq 1 - \alpha \quad \text{mit} \quad K_i = K_i(X_1, \dots, X_n)$$

gilt. Es überdeckt den wahren Parameter θ mit einer Wahrscheinlichkeit von mindestens $1 - \alpha$.

Konfidenzintervall für μ , falls σ^2 unbekannt

Die Stichprobenvariablen $X_i, i = 1, \dots, n$, seien aus der Normalverteilung $N(\mu, \sigma)$. Der Parameter σ^2 sei unbekannt und muss durch den Schätzer S^2 ersetzt werden. Da

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$

gilt, ergibt sich mit dem t_{n-1} -Quantil $t_{n-1;\alpha}$ das zweiseitige $(1-\alpha)$ -Konfidenzintervall

$$\left[\bar{x} - \frac{s}{\sqrt{n}} t_{n-1;1-\frac{\alpha}{2}}; \bar{x} + \frac{s}{\sqrt{n}} t_{n-1;1-\frac{\alpha}{2}} \right].$$

Für $\alpha = 0,05$ und $n \rightarrow \infty$ hat das $t_{n-1;0,975}$ -Quantil den Wert 1,96 und damit erhält man das zweiseitige 95%-KONF

$$\left[\bar{x} - \frac{s}{\sqrt{n}} \times 1,96; \bar{x} + \frac{s}{\sqrt{n}} \times 1,96 \right].$$

Als grobe *Faustregel* kann man also angeben: Trägt man den doppelten Standardfehler zu beiden Seiten des Mittelwertes auf, so wird mit etwa 95%-iger Wahrscheinlichkeit der Mittelwert der Grundgesamtheit von diesem Intervall überdeckt.

7.4 Testverteilungen

Bei statistischen Tests benötigt man so genannte *Prüfgrößen* oder *Teststatistiken*. Setzt man eine normalverteilte Grundgesamtheit voraus, dann besitzen solche Teststatistiken Verteilungen, die näherungsweise oder exakt durch theoretische Verteilungsmodelle beschrieben werden können und im Folgenden definiert werden. Für Schätzer und Testverfahren werden insbesondere die zugehörigen Quantile benötigt.

7.4.1 Chi-Quadrat-Verteilung χ_n^2

Viele Teststatistiken besitzen unter geeigneten Voraussetzungen zumindest approximativ eine Chi-Quadrat-Verteilung.

Seien X_1, \dots, X_n unabhängige und identisch $N(0,1)$ -verteilte Zufallsvariablen, dann heißt die Zufallsvariable

$$Y = \sum_{i=1}^n X_i^2$$

χ_n^2 -verteilt, kurz $Y \sim \chi_n^2$. Die Dichte

$$f_n(y) = \frac{1}{2^{n/2}\Gamma(n/2)} y^{\frac{n}{2}-1} e^{-y/2} \quad \text{für } y > 0$$

heißt Chi-Quadrat-Verteilung mit n Freiheitsgraden, kurz χ_n^2 -Verteilung. Es gilt

$$E(Y) = n \quad \text{und} \quad \text{Var}(Y) = 2n.$$

7.4.2 Student-t-Verteilung t_n

Die Student-t-Verteilung findet besonders bei Parameter-tests und bei Konfidenzintervallen für Parameter Verwendung.

Seien X und Y unabhängige Zufallsvariablen mit $X \sim N(0, 1)$ und $Y \sim \chi_n^2$, dann heißt die Zufallsvariable

$$T = \frac{X}{\sqrt{Y/n}}$$

t_n -verteilt, kurz $T \sim t_n$. Die Dichte

$$f_n(t) = \frac{\Gamma((n+1)/2)}{\Gamma(n/2)\sqrt{n\pi}} (1+t^2/n)^{-(n+1)/2}, \quad -\infty < t < \infty,$$

heißt t-Verteilung mit n Freiheitsgraden, kurz t_n -Verteilung. Es gilt

$$E(T) = 0, \quad \text{falls } n > 1 \quad \text{und} \quad \text{Var}(T) = \frac{n}{n-2}, \quad \text{falls } n > 2.$$

Für $n \rightarrow \infty$ strebt die t_n -Verteilung gegen die Standardnormalverteilung $N(0, 1)$. Im Grenzfall gilt also $\phi = t_\infty$.

7.4.3 Fishers F-Verteilung $F_{m,n}$

Quantile der Fisher F-Verteilung werden vor allem bei Testverfahren der Regressions- und Varianzanalyse benötigt.

Seien X und Y unabhängige χ_m^2 bzw. χ_n^2 verteilte Zufallsvariablen, dann heißt die Zufallsvariable

$$F = \frac{X/m}{Y/n}$$

$F_{m,n}$ -verteilt, kurz $F \sim F_{m,n}$. Die Dichte

$$f_{m,n}(t) = \frac{\Gamma((m+n)/2)}{\Gamma(m/2)\Gamma(n/2)} \left(\frac{m}{n}\right) \left(\frac{m}{n}t\right)^{m/2-1} \left(1 + \frac{m}{n}t\right)^{-(m+n)/2} \quad \text{für } t > 0$$

Fisher- oder F-Verteilung mit m und n Freiheitsgraden, kurz $F_{m,n}$ -Verteilung. Es gilt

$$E(T) = \frac{n}{n-2}, \quad \text{falls } n > 2 \quad \text{und} \quad \text{Var}(T) = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}, \quad \text{falls } n > 4.$$

7.5 Lineare Regression

7.5.1 Einfache lineare Regression

Für die Datenpaare $(x_1, Y_1), \dots, (x_n, Y_n)$ soll ein funktionaler Zusammenhang hergestellt werden, wobei x_1, \dots, x_n gegebene deterministische Werte oder Realisierungen einer *metrischen* Zufallsvariablen X (= Regressor) sind und $Y = (Y_1, \dots, Y_n)$ eine von X abhängige beobachtbare metrische Zufallsvariable (= Zielvariable), deren Realisation mit $y = (y_1, \dots, y_n)$ bezeichnet wird.

Das einfache lineare Regressionsmodell hat die Form

$$Y_i = \alpha + \beta x_i + \varepsilon_i \quad \text{mit} \quad \varepsilon_i \stackrel{iid}{\sim} N(0, \sigma), \quad i = 1, \dots, n,$$

bzw.

$$E(Y_i) = \mu(x_i) = \alpha + \beta x_i, \quad i = 1, \dots, n,$$

wobei ε_i die nicht beobachtbaren Fehler wiedergeben. Für Y_i gilt $Y_i \stackrel{ind}{\sim} N(\mu(x_i), \sigma)$.²² Die Eigenschaft gleicher Varianz der Fehlervariablen ($\text{Var}(\varepsilon_i) = \sigma^2$) wird auch als *Homoskedastizität* (Varianzhomogenität) bezeichnet.

Die Regressionskoeffizienten α (= Regressionskonstante, Schnittpunkt mit der y -Achse) und β (= Steigungskoeffizient), sowie die Varianz σ^2 sind *unbekannte Parameter*, die aus den Daten (x_i, y_i) , $i = 1, \dots, n$, zu schätzen sind.

Schätzung der Parameter

Die Regressionskoeffizienten α und β können durch die *Kleinste-Quadrate-Methode* (*Least Squares Method*) geschätzt werden. Dabei wird die Quadratsumme der Fehler

²²*ind = independent but not identically distributed*

$\varepsilon_i = Y_i - (\alpha + \beta x_i)$, also die Quadratsumme aller vertikalen Abweichungen zwischen den beobachteten Werten Y_i und der Regressionsgeraden $\mu(x_i)$, minimiert.

Durch die Minimierung der Fehler-Quadratsumme (Sum of Squared Errors)

$$SSE(\alpha, \beta) = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (Y_i - \alpha - \beta x_i)^2 \rightarrow \min_{\alpha, \beta}$$

ergeben sich die *Kleinste-Quadrate-Schätzer* $\left\{ \begin{array}{l} \hat{\beta} = \frac{\sum_{i=1}^n (x_i - \bar{x})(Y_i - \bar{Y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{s_{xY}}{s_x^2} \\ \hat{\alpha} = \bar{Y} - \hat{\beta} \bar{x}. \end{array} \right.$

Die Abweichungen $\hat{\varepsilon}_i = Y_i - \hat{Y}_i$, $i = 1, \dots, n$, zwischen den tatsächlichen Werten $Y = (Y_1, \dots, Y_n)$ und den geschätzten (gefitteten) Werten $\hat{Y}_i = \hat{\alpha} + \hat{\beta} x_i$ werden als *Residuen* bezeichnet. Als Schätzer für die Varianz σ^2 wird die gemittelte Residuenquadratsumme

$$\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^n \hat{\varepsilon}_i^2 = \frac{1}{n-2} \sum_{i=1}^n (Y_i - \hat{\alpha} - \hat{\beta} x_i)^2 = \frac{SSE(\hat{\alpha}, \hat{\beta})}{n-2}$$

verwendet. Da

$$E(\hat{\alpha}) = \alpha, \quad E(\hat{\beta}) = \beta, \quad E(\hat{\sigma}^2) = \sigma^2$$

gilt, sind $\hat{\alpha}$, $\hat{\beta}$ und $\hat{\sigma}^2$ *erwartungstreue* Schätzer. Weiters gilt

$$Var(\hat{\alpha}) = \sigma_{\hat{\alpha}}^2 = \sigma^2 \frac{\sum_{i=1}^n x_i^2}{n \sum_{i=1}^n (x_i - \bar{x})^2}, \quad Var(\hat{\beta}) = \sigma_{\hat{\beta}}^2 = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2},$$

und gilt für $n \rightarrow \infty$

$$\sum_{i=1}^n (x_i - \hat{x})^2 \rightarrow \infty,$$

so sind sie auch *konsistent*.

Bemerkungen:

- Die Gerade $\hat{y} = \hat{\alpha} + \hat{\beta}x$ geht durch den Punkt (\bar{x}, \bar{y}) und heißt (gefittete) *Regressionsgerade* bei gegebenen Daten $(x_1, y_1), \dots, (x_n, y_n)$.
- Der Korrelationskoeffizient r_{XY} stimmt mit $\hat{\beta} \frac{s_x}{s_y}$ überein.
- Der geschätzte *Standardfehler der Regression* ist durch $\hat{\sigma} = \sqrt{\frac{SSE(\hat{\alpha}, \hat{\beta})}{n-2}}$ gegeben.

- Für die *standardisierte Residuen* $\hat{\varepsilon}_i^* = \frac{\hat{\varepsilon}_i}{\hat{\sigma}}$, $i = 1, \dots, n$, gilt: $\text{Var}(\hat{\varepsilon}_i^*) \approx 1$.
- Für einen x -Wert x_0 ergibt sich für $Y_0 = \alpha + \beta x_0 + \varepsilon_0$ unter der Voraussetzung

$$\frac{Y_0 - \hat{Y}_0}{\hat{\sigma}_{\hat{Y}_0}} \sim t_{n-2} \quad \text{mit} \quad \hat{\sigma}_{\hat{Y}_0} = \hat{\sigma} \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}$$

das *95%-Prognoseintervall* bei gegebenen Daten

$$[\hat{y}_0 - \hat{\sigma}_{\hat{y}_0} t_{n-2;0,975}; \hat{y}_0 + \hat{\sigma}_{\hat{y}_0} t_{n-2;0,975}].$$

Bei SPSS (englische Version) werden die Werte, die die Grenzen des Prognoseintervalls festlegen, mit „95% L CI for Y individual“ und „95% U CI for Y individual“ bezeichnet. Dabei steht das „L CI“ für „lower confidence interval“ (untere Grenze) und „U CI“ für „upper confidence interval“ (obere Grenze).

7.5.2 Multiple lineare Regression

Hauptunterschied zwischen dem einfachen und dem multiplen Modell ist, dass neben der Zielvariablen Y mehrere *erklärende Variablen* (Regressoren) X_1, \dots, X_k betrachtet werden. Zu diesen Variablen werden jeweils n Werte $y_i, x_{i1}, \dots, x_{ik}$, $i = 1, \dots, n$, gemessen oder beobachtet. Sei $p = k + 1$ die Anzahl der Parameter im Modell mit $p \ll n$ und $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)$ unbekannte, nicht beobachtbare Fehlerterme. Dann beschreibt ε_i die Abweichung der i -ten Beobachtung y_i von der wahren Regressionsebene.

Unter Normalverteilungsannahme kann das **multiple lineare Regressionsmodell** geschrieben werden als

$$Y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} + \varepsilon_i \quad \text{mit} \quad \varepsilon_i \stackrel{iid}{\sim} N(0, \sigma), \quad i = 1, \dots, n,$$

oder in Vektorschreibweise als

$$\mathbf{Y}_{n \times 1} = \mathbf{X}_{n \times p} \boldsymbol{\beta}_{p \times 1} + \boldsymbol{\varepsilon}_{n \times 1} \quad \text{mit} \quad \boldsymbol{\varepsilon} \sim N_n(\mathbf{0}, \sigma \mathbf{I}_n),$$

wobei

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1k} \\ 1 & x_{21} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{nk} \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}, \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

sind und die Einsen in der ersten Spalte der Designmatrix \mathbf{X} dazu dienen, das Modell mit *Intercept* zu modellieren.

Kompakter lässt sich das Modell in Matrixschreibweise darstellen als

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad \text{mit} \quad \boldsymbol{\varepsilon} \sim N_n(\mathbf{0}, \sigma^2 \mathbf{I}_n),$$

bzw. äquivalent dazu als

$$E(\mathbf{Y}) = \boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta} \quad \text{mit} \quad \mathbf{Y} \sim N_n(\boldsymbol{\mu}, \sigma^2 \mathbf{I}_n).$$

Es sind die p Regressionskoeffizienten β_0, \dots, β_k und die Varianz σ^2 zu schätzen.

Schätzung der Parameter

Die Parameter β_0, \dots, β_k werden wie in der einfachen linearen Regressionsanalyse mit der Kleinsten-Quadrate-Methode geschätzt. Der Schätzvektor $\hat{\boldsymbol{\beta}}$ wird durch Minimierung der Fehler-Quadratsumme

$$SSE(\beta_0, \dots, \beta_k) = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (Y_i - (\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}))^2 \rightarrow \min_{\beta_0, \dots, \beta_k}$$

bzw.

$$SSE(\boldsymbol{\beta}) = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})^t(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}) = \mathbf{Y}^t\mathbf{Y} - 2\boldsymbol{\beta}^t\mathbf{X}^t\mathbf{Y} + \boldsymbol{\beta}^t\mathbf{X}^t\mathbf{X}\boldsymbol{\beta}$$

bestimmt. Der Vektor der ersten Ableitungen lässt sich schreiben als

$$\frac{\partial SSE(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \begin{pmatrix} \frac{\partial SSE(\boldsymbol{\beta})}{\partial \beta_0} \\ \vdots \\ \frac{\partial SSE(\boldsymbol{\beta})}{\partial \beta_k} \end{pmatrix} = -2\mathbf{X}^t\mathbf{Y} + 2\mathbf{X}^t\mathbf{X}\boldsymbol{\beta} \stackrel{!}{=} \mathbf{0}.$$

Daraus ergeben sich die Normalgleichungen $\mathbf{X}^t\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}^t\mathbf{Y}$ und unter der Annahme, dass die Matrix \mathbf{X} vollen Rang p hat, bekommt man als Lösung

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^t\mathbf{X})^{-1}\mathbf{X}^t\mathbf{Y} \quad \text{mit} \quad \hat{\boldsymbol{\beta}} \sim N_p\left(\boldsymbol{\beta}, \sigma\sqrt{(\mathbf{X}^t\mathbf{X})^{-1}}\right).$$

Als Schätzer für σ^2 wird wieder die gemittelte Residuenquadratsumme

$$\hat{\sigma}^2 = \frac{1}{n-p} \sum_{i=1}^n \hat{\varepsilon}_i^2 = \frac{1}{n-p} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \frac{SSE(\hat{\beta}_0, \dots, \hat{\beta}_k)}{n-p}$$

verwendet.

Bemerkungen:

- Der geschätzte *Standardfehler der Regression* ist durch $\hat{\sigma} = \sqrt{\frac{SSE(\hat{\beta}_0, \dots, \hat{\beta}_k)}{n-p}}$ gegeben und wird bei SPSS in der *Modellzusammenfassungs*-Tabelle ausgegeben.
- Die geschätzten Regressionskoeffizienten $\hat{\beta}_0, \dots, \hat{\beta}_k$ werden bei SPSS in der *Koeffizienten*-Tabelle unter „Nicht standardisierte Koeffizienten“ in der Spalte „B“ angegeben. Diese geben an, um welchen Beitrag sich die abhängige Variable ändert, wenn sich die unabhängige Variable um eine Einheit ändert.
- Die geschätzten *Standardfehler* $\hat{\sigma}_0, \dots, \hat{\sigma}_k$ der einzelnen Regressoren mit

$$\hat{\sigma}_j = \sqrt{\widehat{Var}(\hat{\beta}_j)} = \hat{\sigma} \sqrt{(X^t X)^{-1}_{jj}}, \quad j = 0, \dots, p-1,$$

sind bei SPSS in der zweiten Spalte unter „Nicht standardisierte Koeffizienten“ eingetragen.

- Die geschätzten Regressionskoeffizienten $\hat{\beta}_0, \dots, \hat{\beta}_k$ sind abhängig von der Einheit in der die Regressoren X_1, \dots, X_k gemessen wurden. Daher können diese in extrem unterschiedlichen Größenordnungen auftreten und ihr Betrag sagt nichts darüber aus, wie hoch die Erklärungskraft des Regressors ist. Um zwei Regressionskoeffizienten vergleichen zu können, müssen die geschätzten *standardisierten Beta* betrachtet werden, die bei SPSS in der *Koeffizienten*-Tabelle unter „Standardisierte Koeffizienten“ ausgegeben werden. Diese geben an, um wie viele Standardabweichungen sich die abhängige Variable verändert, wenn die unabhängige Variable sich um eine Standardabweichung verändert.

- Da unter der Normalverteilungsannahme für die standardisierten Schätzer

$$\frac{\hat{\beta}_j - \beta_j}{\hat{\sigma}_j} \sim t_{n-p}, \quad j = 0, \dots, p-1,$$

gilt, ergeben sich daraus die *95%-Konfidenzintervalle* für β_j

$$\left[\hat{\beta}_j - \hat{\sigma}_j t_{n-p;0,975}; \hat{\beta}_j + \hat{\sigma}_j t_{n-p;0,975} \right], \quad j = 0, \dots, p-1.$$

- Durch den Standardfehler der Residuen $\hat{\sigma}_{\hat{\varepsilon}_i} = \hat{\sigma} \sqrt{1 - h_{ii}}$ mit $\mathbf{H} = \mathbf{X}(\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t$ als Hat-Matrix und $h_{ii} = \text{diag}(\mathbf{H})$ mit $\frac{1}{n} \leq h_{ii} \leq 1$ ergeben sich die *standardisierten Residuen*

$$\hat{\varepsilon}_i^* = \frac{\hat{\varepsilon}_i - \bar{\varepsilon}}{\hat{\sigma}_{\hat{\varepsilon}_i}} \stackrel{\bar{\varepsilon}=0}{=} \frac{\hat{\varepsilon}_i}{\hat{\sigma}_{\hat{\varepsilon}_i}}, \quad i = 1, \dots, n,$$

mit deren Hilfe sich z.B. mittels Residuenplot (siehe Seite 103) Ausreißer identifizieren lassen. Diese können einen starken Einfluss auf das Modell haben und die Normalverteilungsannahme stören, weswegen solche Ausreißer auszuschließen sind und das Modell danach neu berechnet werden sollte.

Quadratsummen und Bestimmtheitsmaße

Eine wesentliche Rolle in der Regressionsanalyse spielt die Zerlegung der **Totalen Quadratsumme** (**S**um of **S**quared **T**otal) in die zwei Komponenten **Regressions-Quadratsumme** (**S**um of **S**quared **R**egression) und **Fehler-Quadratsumme** (**S**um of **S**quared **E**rrors). Diese *Varianz-* bzw. *Streuungszerlegung*

$$SST = SSR + SSE$$

$$\sum_{i=1}^n (Y_i - \bar{Y})^2 = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 + \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

unterteilt die totale Variabilität in \mathbf{Y} in die durch die Regression erklärte (= zwischen den Gruppen) und die durch das Modell nicht erklärte (= innerhalb der Gruppen) Variabilität.

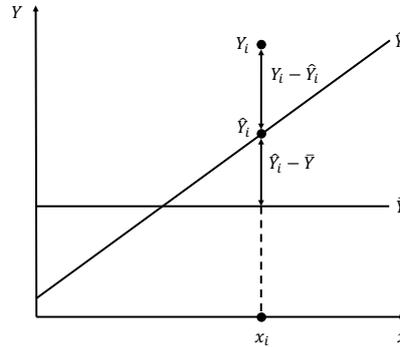


Abbildung 7.2: Quadratsummenzerlegung

Der Quotient aus der erklärten Streuung und der Gesamtstreuung

$$R^2 = \frac{\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

heißt **Bestimmtheitsmaß**, wird zur Beurteilung der *Güte der Anpassung* verwendet und gibt den Anteil der Gesamtvarianz an, der durch das Regressionsmodell von \mathbf{Y} auf \mathbf{X} erklärt wird ($0 \leq R^2 \leq 1$). Das heißt, je näher R^2 bei 1 liegt, desto besser wird die Zielvariable durch die unabhängigen Variablen bestimmt und je näher R^2 bei 0 liegt, desto schlechter können die Regressoren die Variabilität der abhängigen Variablen erklären.

Im Falle des einfachen linearen Modells stimmt R^2 mit dem Quadrat des Korrelationskoeffizienten R_{XY} überein. Bei der multiplen linearen Regression entspricht R^2 dem Quadrat des multiplen Korrelationskoeffizienten. Sind die Regressoren unkorreliert, so ergeben die aufsummierten r^2 -Werte aller Regressoren das Bestimmtheitsmaß, also $R^2 = \sum_{i=1}^k (r_{x_i Y})^2 = r_{x_1 Y}^2 + r_{x_2 Y}^2 + \dots + r_{x_k Y}^2$.

Da das Bestimmtheitsmaß die Eigenschaft hat, dass es mit der Anzahl der unabhängigen Variablen im Modell wächst, unabhängig davon, ob weitere Parameter wirklich das Modell verbessern, muss R^2 skaliert werden. Daraus ergibt sich mit

$$R_{\text{korrr}}^2 = R_{\text{adj}}^2 = 1 - \frac{\frac{1}{n-p} SSE}{\frac{1}{n-1} SST} = 1 - \frac{n-1}{n-p} (1 - R^2)$$

das **korrigierte** (*adjustierte, angepasste* oder *bereinigte*) Bestimmtheitsmaß und es gilt stets $R^2 > R_{\text{korrr}}^2$.

Signifikanztests

Der Signifikanztest testet „gegen“ die Nullhypothese H_0 , d.h. es wird getestet, ob die Nullhypothese verworfen werden kann oder eben nicht verworfen werden kann. Um eine Entscheidung für die Gegenhypothese H_1 oder für die Nullhypothese fällen zu können, wird ein *berechneter empirischer Wert* mit einem *vorgegebenen kritischen Wert* verglichen. Unter einem vorgegebenem α , welches als *Irrtumswahrscheinlichkeit*, *Fehler 1. Art* oder *Signifikanzniveau* des Tests bezeichnet wird und die Wahrscheinlichkeit angibt, dass die Nullhypothese H_0 verworfen wird, obwohl sie in Wirklichkeit zutrifft, führt dieser Vergleich dann entweder zur Ablehnung der Nullhypothese H_0 (= Annahme von H_1) oder Nichtablehnung von H_0 .

Signifikanztest für das Gesamtmodell: Overall-F-Test

Mit dem F-Test wird überprüft, ob die Regressoren überhaupt zur Erklärung der Zielvariablen beitragen. Unter den Hypothesen

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_k = 0 \quad \text{gegen} \quad H_1 : \beta_j \neq 0 \quad \text{für mindestens ein } j$$

wird das Verhältnis von erklärter Streuung zu nicht-erklärter Streuung berechnet und es ergibt sich die Teststatistik

$$F = \frac{\frac{1}{p-1}SSR}{\frac{1}{n-p}SSE} = \frac{n-p}{p-1} \frac{R^2}{1-R^2} \stackrel{H_0}{\sim} F_{p-1;n-p},$$

die in enger Beziehung zum Bestimmtheitsmaß steht. Die Nullhypothese besagt also, dass alle Regressoren X_1, \dots, X_k zusammen keine Varianz in der abhängigen Variablen erklären. Der F-Wert wird mit dem kritischen Wert verglichen, der aus entsprechenden Verteilungstabellen abgelesen werden kann.

Beim Statistikprogramm SPSS steht der F-Wert in der ANOVA-Tabelle und außerdem wird die beobachtete *Signifikanz*, der so genannte *p*-Wert, welcher die Wahrscheinlichkeit angibt, dass die Teststatistik unter H_0 größer als der realisierte Wert der Teststatistik ist, mit ausgegeben. Diese beiden Werte sind in den Kapiteln 5 und 6 dieser Arbeit in der *Modellzusammenfassungs*-Tabelle zu finden.

Je kleiner dieser Signifikanzwert (= *p*-Wert) ist, desto mehr spricht das Ergebnis gegen

Tabelle 7.1: Symbolisierung des Signifikanzniveaus in SPSS

Signifikanz	Bedeutung	Symbol
$> 0,05$	nicht signifikant	
$\leq 0,05$	signifikant	*
$\leq 0,01$	sehr signifikant	**
$\leq 0,001$	höchst signifikant	***

die Nullhypothese. SPSS kennzeichnet diesen Wert bei manchen Verfahren mit Sternchen, wie in Tabelle 7.1 angegeben.

Signifikanz einzelner Regressoren: t-Test

Mittels t-Test kann der Einfluss der einzelnen Regressoren X_1, \dots, X_k auf die Zielvariable getestet werden. Hierfür werden folgende Hypothesen betrachtet:

$$H_0 : \beta_j = 0 \quad \text{gegen} \quad H_1 : \beta_j \neq 0.$$

D.h., wenn H_0 zutrifft, hat der Regressor X_j keinen signifikanten Erklärungswert für \mathbf{Y} und kann aus dem Regressionsansatz entfernt werden. Trifft H_1 zu, muss X_j im Ansatz enthalten sein. Diese Hypothesen werden mit der Teststatistik

$$T_j = \frac{\hat{\beta}_j}{\hat{\sigma}_j} \stackrel{H_0}{\sim} t_{n-p}, \quad j = 0, \dots, p-1,$$

getestet.

Da für $\alpha = 0,05$ und $n \rightarrow \infty$ das $t_{n-p;0,975}$ -Quantil den Wert 1,96 besitzt, bedeutet dies, dass als grobe *Faustregel* angegeben werden kann, dass jene Regressoren signifikant zur Erklärung der Zielvariablen beitragen, für deren T-Werte gilt:

$$-2 \geq T \geq 2.$$

In SPSS sind die empirischen t-Werte und die Signifikanz der Regressoren in der *Koeffizienten*-Tabelle enthalten. Auch hier gibt der Signifikanzwert die Wahrscheinlichkeit an, dass die Nullhypothese nicht abgelehnt werden kann.

Modellverstöße

Lineare Regressionsmodelle basieren auf bestimmten Voraussetzungen. Wird eine Regression durchgeführt, obwohl die Voraussetzungen nicht zutreffen, wird die Modellschätzung fehlerhaft, verzerrt oder ineffizient.

- **Nichtlinearität des Zusammenhangs:** Die lineare Schätzung eines nichtlinearen Zusammenhangs liefert verzerrte, falsche Schätzwerte. Nichtlinearität lässt sich im Streudiagramm diagnostizieren.
- **Multikollinearität der Regressoren:** Die Stärke des Zusammenhangs zwischen den unabhängigen Variablen wird mit Multikollinearität bezeichnet. Hohe Korrelationen zwischen zwei oder mehreren Regressoren führen zu verzerrten und wenig stabilen Schätzwerten und erhöhen den Standardfehler der Regressionskoeffizienten. Daher bietet sich die Durchführung einer schrittweisen Regression (vgl. Seite 105) an.
- **Heteroskedastizität:** Heteroskedastizität bedeutet, dass die Regressionsresiduen verschiedener Schätzwerte sehr stark variieren. Dadurch werden Standardfehler und Konfidenzintervall verfälscht. Diese Eigenschaft kann mittels Residuenplot, der die unstandardisierten Vorhersagewerte gegen die standardisierten Residuen aufträgt (vgl. Seite 103), überprüft werden.
- **Nicht-Normalverteilung der Residuen:** Die Residuen sollen in ihrer Verteilung der Normalverteilung folgen. Ist dies nicht der Fall, so wird die grundlegende Modellvoraussetzung verletzt und sämtliche auf dem Modell basierende Schlussfolgerungen sind inkorrekt. Die Residuen können mittels Histogramm und Q-Q-Plot auf Normalverteilung überprüft werden.
- **Autokorrelation:** Regressionsmodelle basieren auf der Annahme, dass die Residuen eines geschätzten Modells unkorreliert sind. Autokorrelation bedeutet, dass die Residuen voneinander abhängig sind und das gewählte Regressionsmodell den tatsächlichen Zusammenhang nicht vollständig erklären kann. Bei Korrelation der Residuen können die berechneten Standardfehler der Regressionskoeffizienten deutlich kleiner sein als sie in Wirklichkeit sind. Das täuscht dann eine höhere Präzision des berechneten Modells vor. Als Konsequenz daraus sind die Konfi-

denzintervalle und die Signifikanztests der Parameter nicht mehr gültig.²³

SPSS bietet zur Überprüfung einer möglichen Autokorrelation den Durbin-Watson-Test an.

Durbin-Watson-Autokorrelationstest^{24,25}

Durch Aufstellen der Nullhypothese, dass *keine Autokorrelation 1. Ordnung* ($\rho_1 = 0$) vorliegt und deren Gegenhypothese, dass *Autokorrelation 1. Ordnung* ($\rho_1 \neq 0$), d.h. Korrelation zwischen zwei aufeinanderfolgenden Residualgrößen, vorhanden ist, kann mit Hilfe des von Durbin und Watson (1951) vorgeschlagenen Testverfahrens

$$d = \frac{\sum_{i=2}^n (\hat{\varepsilon}_i - \hat{\varepsilon}_{i-1})^2}{\sum_{i=1}^n \hat{\varepsilon}_i^2}, \quad i = 1, \dots, n,$$

die Annahme, dass die Residuen unkorreliert sind, überprüft werden, wobei die $\hat{\varepsilon}_i$ jeweils die Residuen der Regression in der i -ten Periode bezeichnen. Wenn die Differenz zwischen den Residualgrößen sehr klein bzw. sehr groß ist, so liegt positive bzw. negative Autokorrelation vor.

Da als Testentscheidung

- $d = 0 \Leftrightarrow \rho_1 = 1 \quad \dots \quad \text{perfekte positive Autokorrelation}$
- $d = 4 \Leftrightarrow \rho_1 = -1 \quad \dots \quad \text{perfekte negative Autokorrelation}$
- $d = 2 \Leftrightarrow \rho_1 = 0 \quad \dots \quad \text{keine Autokorrelation}$

gilt, hat sich in der Praxis die Faustregel bewährt, dass beim Auftreten eines Teststatistikwertes nahe zwei davon ausgegangen werden kann, dass weder positive noch negative Autokorrelation erster Ordnung vorliegt.

Voraussetzungen für die Anwendung sind, dass das Regressionsmodell eine Konstante (Intercept) beinhaltet, der nicht beobachtbare Fehlerterm normalverteilt ist und in der Modellgleichung die Zielvariable nicht als zeitversetzte erklärende Variable auftaucht.

²³http://www.statistics4u.info/fundstat_germ/ee_durbin_watson.html

²⁴<http://de.wikipedia.org/wiki/Durbin-Watson-Test>

²⁵<http://wirtschaftslexikon.gabler.de/Definition/durbin-watson-autokorrelationstest.html>

7.5.3 Graphische Modelldiagnostik

Für die sorgfältige Beantwortung der Frage nach Angemessenheit eines gewählten Modells ist die Ermittlung des Bestimmtheitsmaßes oder der Wert der F-Statistik allein nicht ausreichend. Mit Hilfe einer Residuenanalyse können schlecht gefittete \hat{Y} -Werte und insbesondere Verletzungen der Voraussetzungen bezüglich der Fehlervariablen ε gefunden werden, z.B. eine Verletzung der Homoskedastizität.

Partielles Regressionsdiagramm

In diesem *Streudiagramm* (vgl. Seite 83) werden die Residuen der abhängigen Variable auf der y -Achse gegen die Residuen einer erklärenden Variablen auf der x -Achse geplottet, die sich ergeben, wenn die beiden Variablen jeweils getrennt von den übrigen unabhängigen Variablen erklärt werden. Diese Auspartialisierung funktioniert wie im Fall der Berechnung der partiellen Korrelationskoeffizienten.

Sei also x_j die j -te Spalte der Designmatrix \mathbf{X} . Dann ist

$$\varepsilon_j^* = \varepsilon + \hat{\beta}_j x_j = Y - \sum_{k \neq j} \hat{\beta}_k x_k$$

der j -te partielle Residuenvektor (Larsen und McCleary, 1972). Somit kann ε_j^* als abhängige Variable, bereinigt um den Effekt aller übrigen Regressoren, mit Ausnahme des j -ten, aufgefasst werden.²⁶

Histogramm

Einen ersten optischer Nachweis, ob die standardisierten Residuen $\hat{\varepsilon}_i^*$ annähernd normalverteilt sind, kann ein **Histogramm** liefern, wie auf Seite 83 erklärt. Weiters kann in SPSS die Normalverteilungsdichte zu Vergleichszwecken eingezeichnet werden.

²⁶[7] Naredi-Rainer, 1986

Q-Q-Plot

Die Verteilungsannahme $\varepsilon_i \stackrel{iid}{\sim} N(0, \sigma)$ wird durch einen Q-Q-Plot, wie auf Seite 83 beschrieben, überprüft. Dafür werden die standardisierten Residuen $\hat{\varepsilon}_i^*$ gegen die theoretischen Quantile $F^{-1}\left(\frac{i}{n+1}\right)$, mit F Verteilungsfunktion von $N(\bar{x}, s)$, aufgetragen. Bei Vorliegen einer Normalverteilung sollten die Punkte ein lineares Muster entlang der 1. Mediane bilden. Ausreißer oder Abweichungen vom linearen Muster deuten auf eine Verletzung der Normalverteilung hin.

- Bilden die Punkte die Gerade $x = y$, so ist die theoretische Verteilung eine gute Approximation der empirischen Verteilung.
- Bilden die Punkte eine Gerade parallel zu $x = y$, so liegt der Unterschied zwischen theoretischer und empirischer Verteilung nur im Lage-Parameter.
- Bilden die Punkte eine Gerade, die $x = y$ schneidet, so beruht der Unterschied im Skalierungs-Parameter.
- Bilden die Punkte kein Geradenmuster, so liegt der empirischen Verteilung eine andere theoretische Verteilung zu Grunde.

Residuenplot

Im Residuenplot werden die unstandardisierten Vorhersagewerte \hat{Y}_i auf der x -Achse gegen die standardisierten Residuen $\hat{\varepsilon}_i^*$ auf der y -Achse aufgetragen und es ergibt sich ein *Scatterplot*, der folgende Aussagen erlaubt:

- Wenn das lineare Modell angemessen ist, also $\hat{\varepsilon}_i = Y_i - \hat{Y}_i$ unabhängig von \hat{Y}_i für $i = 1, \dots, n$, weist der entsprechende Residuenplot der Daten ein zufälliges Muster um die horizontale Gerade $y = 0$ auf.
- Ist die Voraussetzung einer konstanten Varianz σ^2 nicht erfüllt, dann hat der Residuenplot eine erkennbare Struktur, z.B. eine Trichterform, welche auf eine Zunahme der Varianz in Abhängigkeit von den Vorhersagewerten hindeutet.
- Ist das lineare Modell nicht für die vorliegende Datensituation geeignet, dann erkennt man im Residuenplot eine gekrümmte Struktur.

- Als Ausreißer lassen sich nach einer *Faustregel* solche Fälle identifizieren, deren standardisierte Residuen größer als 3,3 sind (das entspricht einem Signifikanzniveau von $\alpha = 0,001$).²⁷

Dies bedeutet, dass die Referenz eine zufällige Struktur ohne erkennbares Muster um die Nulllinie sein soll.

7.5.4 Methoden der linearen Regressionsanalyse im SPSS

Einschluss-Methode

Wenn bekannt ist, welche Regressoren in die Regressionsgleichung aufgenommen werden sollen, so wird die voreingestellte **Einschluss**-Methode verwendet. In diesem Fall werden alle angegebenen Variablen für die lineare Regression benutzt. Diese Methode ist besonders dann angebracht, wenn die unabhängigen Regressoren untereinander unkorreliert sind.²⁸

Vorwärts-Methode

Bei der **Vorwärts**-Methode werden der Gleichung nacheinander die Regressoren mit dem höchsten partiellen Korrelationskoeffizienten mit der abhängigen Zielvariablen hinzugefügt, sofern sie dem F-Signifikanztest zur Aufnahme (Wahrscheinlichkeit zur Aufnahme) genügen. Beim Statistikprogramm SPSS wird für die Aufnahme ein Signifikanzniveau von $\alpha = 0,05$ zugrundegelegt.

Rückwärts-Methode

Bei der **Rückwärts**-Methode wird mit der Lösung angefangen, die alle unabhängigen Variablen enthält und schließt dann jeweils die unabhängigen Variablen mit dem kleinsten partiellen Korrelationskoeffizienten aus, soweit der zugehörige Regressionskoeffizient nicht signifikant ist (Wahrscheinlichkeit zum Ausschluss). SPSS legt für den Ausschluss ein Signifikanzniveau von $\alpha = 0,1$ zugrunde.

²⁷[http://home.arcor.de/sigmatraugott/Skripte/Lineare Regression.html](http://home.arcor.de/sigmatraugott/Skripte/Lineare%20Regression.html)

²⁸<http://www.lrz.de/services/schulung/unterlagen/spss/spss-regression/>

Schrittweise-Methode

Sind die Regressoren untereinander möglicherweise korreliert, so kann die **hierarchische Regressionsanalyse** mittels der **Schrittweise-Methode** angewandt werden. Diese funktioniert ähnlich wie die Vorwärts-Methode, allerdings werden bei der schrittweisen Regressionsanalyse nach jedem Schritt die jeweils aufgenommenen Variablen nach der Rückwärts-Methode untersucht. Das heißt, dass nur die unabhängigen Variablen in das Regressionsmodell aufgenommen werden, die einen signifikanten zusätzlichen Erklärungsbeitrag von \mathbf{Y} leisten und gleichzeitig aufgenommene Regressoren, deren separater Erklärungsbeitrag durch die Einbeziehung weiterer Variablen signifikant gesunken ist, aus dem Modell wieder ausgeschlossen werden.²⁹ Der Prüfprozess wird solange fortgesetzt, bis für keine Variable mehr die Möglichkeit der Aufnahme existiert. Diese Vorgehensweise ermöglicht es ein „sparsames“ Modell zu identifizieren, d.h. mit möglichst wenigen Variablen möglichst viel Varianz in der abhängigen Variable zu erklären.

Variablenauswahl:

1. Aufnahme der unabhängigen Variablen X_1 mit dem größten Korrelationskoeffizienten zur Zielvariablen \mathbf{Y} :

$$r_{X_1\mathbf{Y}} = \frac{\sum_{i=1}^n (x_{i1} - \bar{x}_1)(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (x_{i1} - \bar{x}_1)^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}} = \frac{s_{X_1\mathbf{Y}}}{s_{X_1}s_{\mathbf{Y}}}$$

2. Aufnahme der unabhängigen Variablen X_2 mit dem betragsmäßig größten partiellen Korrelationskoeffizienten zur Zielvariablen \mathbf{Y} :

$$r_{X_2\mathbf{Y}.X_1} = \frac{r_{X_2\mathbf{Y}} - r_{X_2X_1} \times r_{\mathbf{Y}X_1}}{\sqrt{1 - r_{X_2X_1}^2} \sqrt{1 - r_{\mathbf{Y}X_1}^2}}$$

3. Überprüfung der ersten beiden Regressoren X_1 und X_2 auf ihre partiellen F-Werte

$$F_1 = \frac{SSR(\hat{\beta}_1|\hat{\beta}_2)}{\frac{1}{n-3}SSE(\hat{\beta}_1, \hat{\beta}_2)} = \frac{SSR(\hat{\beta}_1, \hat{\beta}_2) - SSR(\hat{\beta}_2)}{\frac{1}{n-3}SSE(\hat{\beta}_1, \hat{\beta}_2)} \quad \dots \text{ F-Test für } X_1$$

²⁹<http://www.ivwl.uni-kassel.de/eckey/lehre/oekonometrie/spss/sitzung8.pdf>

$$F_2 = \frac{SSR(\hat{\beta}_2|\hat{\beta}_1)}{\frac{1}{n-3}SSE(\hat{\beta}_1, \hat{\beta}_2)} = \frac{SSR(\hat{\beta}_1, \hat{\beta}_2) - SSR(\hat{\beta}_1)}{\frac{1}{n-3}SSE(\hat{\beta}_1, \hat{\beta}_2)} \quad \dots \text{ F-Test für } X_2$$

und eventuelle Entfernung einer der Variablen, wenn besagter Wert unter dem Signifikanzniveau ($\alpha = 0, 1$) liegt.

4. Aufnahme der unabhängigen Variablen X_k mit dem betragsmäßig größten partiellen Korrelationskoeffizienten zur Zielvariablen \mathbf{Y} .
5. Überprüfung aller Regressoren des aktuellen Modells auf ihre partiellen F-Werte und Entfernung jener Variablen, deren Wert unter dem vordefinierten Signifikanzniveau liegen. Als Teststatistik für den Regressor X_k , falls X_1, \dots, X_{k-1} schon ins Modell aufgenommen wurden, ergibt sich

$$F_k = \frac{\frac{1}{k-(k-1)}SSR(\hat{\beta}_k|\hat{\beta}_1, \dots, \hat{\beta}_{k-1})}{\frac{1}{n-k-1}SSE(\hat{\beta}_1, \dots, \hat{\beta}_k)} = \frac{SSR(\hat{\beta}_1, \dots, \hat{\beta}_k) - SSR(\hat{\beta}_1, \dots, \hat{\beta}_{k-1})}{\frac{1}{n-k-1}SSE(\hat{\beta}_1, \dots, \hat{\beta}_k)}.$$

6. Wiederholung der Prozedur ab Schritt 4, bis ein vordefiniertes Kriterium zur Beendigung des Verfahrens erreicht ist.

Zusammenfassung

Für die **Analyse von Solaranlagen** standen Daten von **fünf Grazer** Anlagen und **einer in Judendorf** mit unterschiedlichen Aufzeichnungszeiträumen (von 392 bis 2.080 korrekten Aufzeichnungstagen) zwischen 2002 und 2008, sowie **Wetterdaten** für den Zeitraum **01.01.2002 - 31.12.2008** zur Verfügung.

Bei der Aufbereitung der Messparameter der Solaranlagen wurden leider viele und extreme Aufzeichnungsfehler festgestellt. Auf Grund dieser Fehler, vieler Ausreißer und offensichtlich fehlender laufender Qualitätskontrolle der Daten musste viel mehr Arbeit als geplant in die Datenerfassung und Bereinigung der Rohdaten gesteckt werden.

Darüber hinaus wurden mit Hilfe von in VBA programmierten Excel-Files von Herrn DI Philip Ohnewein und Herrn DI Hausner zusätzliche Variablen, wie der *Einfallswinkel* der *Sonnenstrahlen* und die *geneigten Global-* und *Diffusstrahlungen*, berechnet, um daraus dann die Kollektorkennwerte *eta0_dir*, *eta0_diff*, *c1* und *c2* berechnen zu können. Diese Kennwerte stellten sich im Laufe der Analyse als recht aussagekräftig für den Ertrag einer Solaranlage heraus, da sie sich aber teilweise aus Daten von schon erzielten Erträgen zusammensetzen, wurden diese für die Erstellung der Prognosemodelle nicht berücksichtigt.

Die erste Analyse der Zielvariablen WMZ in kWh und kWh/m² zeigte erwartungsgemäß die **höchsten** gesamten und mittleren **Erträge** in den **Sommermonaten** April bis September. Im Mittel erzeugte die Solaranlage am Schwarzen Weg den höchsten Tagesertrag mit 1,7 kWh/m², gefolgt von den Solaranlagen am Dach des Liebenauer Stadions (1,6 kWh/m²) und in der Eggenberger Allee (1,5 kWh/m²). Den niedrigsten mittleren Tagesertrag lieferte die Solaranlage in der Theodor-Körner-Straße mit nur 1,3 kWh/m². Die Anlage in der Grottenhofstraße war im Mittel mit 6,5 Stunden pro Tag am häufigsten in Betrieb, die Anlage am Liebenauer Stadion hingegen nur 4,75 Stunden.

Mittels Korrelationen und linearen Regressionen wurde festgestellt, dass die **Globalstrahlung** (geneigt und horizontal) den **stärksten Einfluss** auf den Ertrag/m² hat, gefolgt von der **Lufttemperatur** und der **Diffusstrahlung** (geneigt oder als Diffusanteil). Auch nicht außer Acht lassen durfte man die **Anlagedaten**, wie *Kollektorneigungswinkel*, die *äußere Leitungslänge*, den *Glykolgehalt* oder die *Wärmekapazität*.

Für die **Prognosemodelle** wurden nur **vorhersagbare** (*horizontale Global-, Diffus-, Direktstrahlung* und *Lufttemperatur*), bzw. aus vorhersagbaren berechnete **Variablen** (*geneigte Global-, Diffus- und Direktstrahlung*), sowie die **Anlagedaten** berücksichtigt. Da nur Solaranlagen mit ähnlichen *geographischen Koordinaten* zur Verfügung standen, waren diese Koordinaten hier nicht anwendbar. Unter der Annahme unbekannter Anlagedaten verschlechterte sich das korrigierte R^2 nur um 1-2%.

Insgesamt lassen sich mittels der Prognosemodelle mit **geneigter Strahlung** über **85%** des **Tagesertrages/m²** und ca. **90%** der **Monatsertrages/m²** erklären. Mit **horizontaler Strahlung** sind es immerhin noch über **81%** des **Tagesertrages/m²** und ebenfalls ca. **90%** des **Monatsertrages/m²**.

Für alle Prognosemodelle wurden prinzipiell immer die **Global- und Diffusstrahlungen** (geneigt, horizontal und als Diffusanteil), sowie die **Lufttemperatur** in Kombinationen mit den Anlagedaten verwendet.

Für die beiden Excel-Prognosefiles „**PROGNOSE_Tag**“ und „**PROGNOSE_Monat**“ wurden die **horizontale Globalstrahlung**, deren **Diffusanteil** und die **Lufttemperatur** verwendet, sowie der **Kollektorneigungswinkel**, die **äußeren Leitungslänge** und die **Wärmekapazität**.

Anhang A

Excel-Prognosefiles

Die in Anhang A.1 bis A.4 dargestellten Screenshots zeigen die Ein- und Ausgabefelder der Excel-Prognosefiles, sowie als graphische Darstellung die Fehlerbalken der jeweiligen Prognosen in kWh/m² und kWh samt ihren Prognoseintervallen.

Anhang	File
A.1	PROGNOSE_Tag
A.2	PROGNOSE_Monat
A.3	PROGNOSE_Tag_Graz
A.4	PROGNOSE_Monat_Graz
A.5	Weitere Tages- und Monats-Prognosefiles

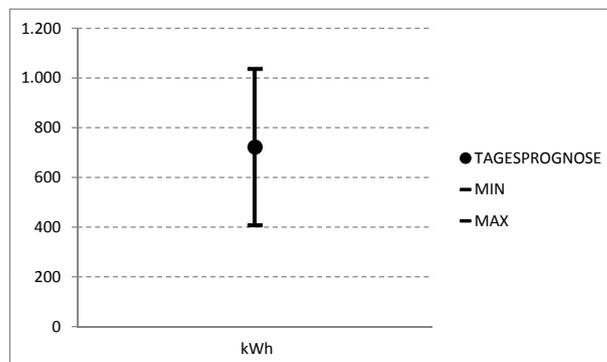
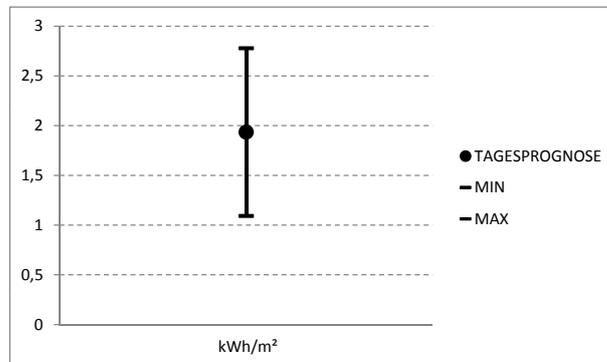
Die Werte in den hellgrünen Zellen der Eingabemasken entsprechen den jeweiligen (Sommer-) Mittelwerten, welche zu Demonstrationszwecken eingegeben wurden.

A.1 PROGNOSE_Tag

Tagesprognose für den Ertrag einer Solaranlage

EINGABE									
SAISON	Sommer Winter April-September								
WETTERPROGNOSE	Globalstrahlung [W/m ²] 5.209								
	Diffusanteil [%] 51								
	Lufttemperatur [°C] 17,3								
ANLAGEDATEN	Neigungswinkel Kollektor [°] 33								
	Leitungslänge außen [m] 35								
	Wärmekapazität [J/m ² K] 6113								
ANLAGE	Kollektorfläche [m ²] 373								
	<table border="1"> <thead> <tr> <th>MIN</th> <th>MAX</th> </tr> </thead> <tbody> <tr> <td>25</td> <td>39</td> </tr> <tr> <td>0</td> <td>125</td> </tr> <tr> <td>6040</td> <td>6810</td> </tr> </tbody> </table>	MIN	MAX	25	39	0	125	6040	6810
MIN	MAX								
25	39								
0	125								
6040	6810								
Eingabe löschen									

AUSGABE		
TAGESPROGNOSE	1,94 kWh/m ²	1,09 2,78
		MIN MAX
TAGESPROGNOSE	722 kWh	408 1.037
		MIN MAX

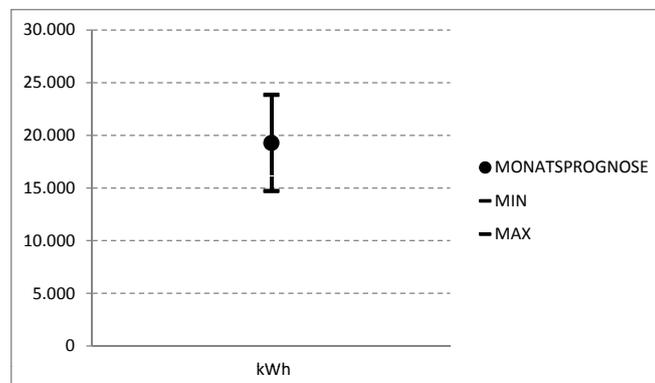
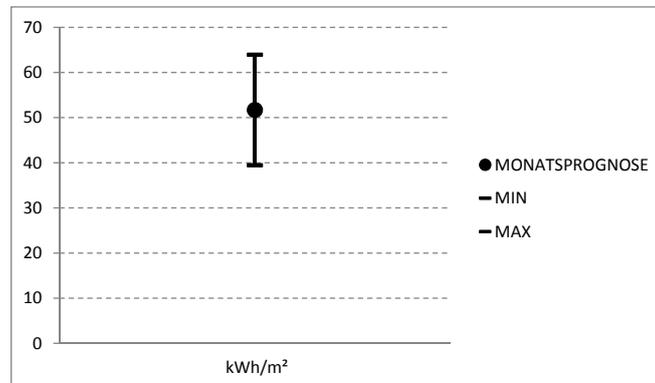


A.2 PROGNOSE_Monat

Monatsprognose für den Ertrag einer Solaranlage

EINGABE				Eingabe löschen	
WETTERPROGNOSE		Globalstrahlung [W/m ²]	148.037		
		Diffusanteil [%]	47		
		Lufttemperatur [°C]	16,6		
ANLAGEDATEN		Neigungswinkel Kollektor [°]	33	MIN	MAX
		Leitungslänge außen [m]	42	25	39
		Wärmekapazität [J/m ² K]	6118	0	125
				6040	6810
ANLAGE		Kollektorfläche [m ²]	373		

AUSGABE					
MONATSPROGNOSE	51,70	kWh/m ²	39,47	63,94	
			MIN	MAX	
MONATSPROGNOSE	19.286	kWh	14.722	23.849	
			MIN	MAX	

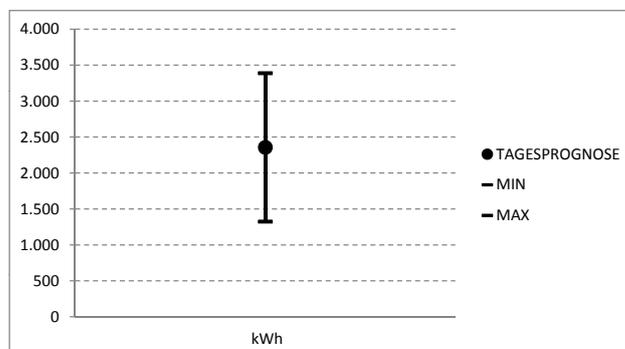
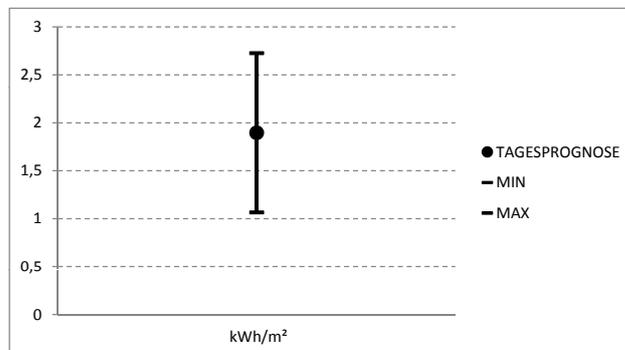


A.3 PROGNOSE _ Tag _ Graz

Tagesprognose für Solaranlagen in Graz / Graz-Umgebung

EINGABE	
ANLAGE	Kollektorfläche [m ²] Stadion Liebenau 1.242,7
SAISON	Sommer Winter April-September
WETTERPROGNOSE	Globalstrahlung [W/m ²] 5.329 MIN 0 MAX 8.858 Diffusanteil [%] 50 11 100 Lufttemperatur [°C] 17,7 -11,3 28
ANLAGEDATEN	geographische Breite [°] 47,05 geographische Länge [°] 15,45 Neigungswinkel Kollektor [°] 31 Glykolgehalt [%] 33

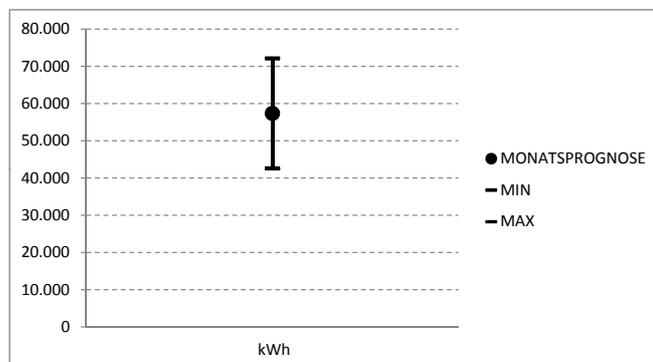
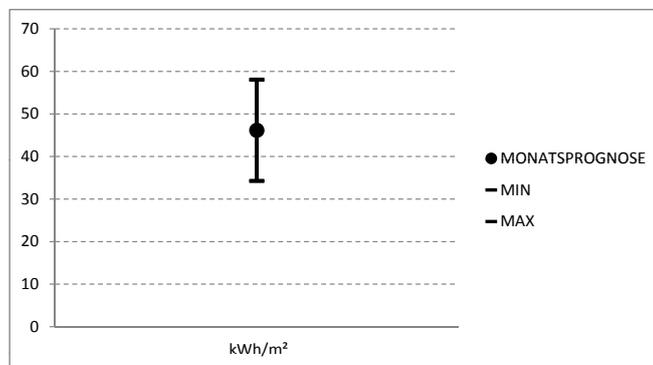
AUSGABE	
TAGESPROGNOSE	1,90 kWh/m ² 1,07 2,73
TAGESPROGNOSE	2.357 kWh 1.325 3.389



A.4 PROGNOSE_Monat_Graz

Monatsprognose für Solaranlagen in Graz / Graz-Umgebung

EINGABE													
ANLAGE	Kollektorfläche (m ²) Stadion Liebenau 1.242,7 Eingabe löschen												
WETTERPROGNOSE	<table border="1"> <thead> <tr> <th></th> <th>MIN</th> <th>MAX</th> </tr> </thead> <tbody> <tr> <td>Globalstrahlung [W/m²]</td> <td>17.803</td> <td>196.367</td> </tr> <tr> <td>Diffusanteil [%]</td> <td>33</td> <td>90</td> </tr> <tr> <td>Lufttemperatur [°C]</td> <td>-3,3</td> <td>22,8</td> </tr> </tbody> </table>		MIN	MAX	Globalstrahlung [W/m ²]	17.803	196.367	Diffusanteil [%]	33	90	Lufttemperatur [°C]	-3,3	22,8
	MIN	MAX											
Globalstrahlung [W/m ²]	17.803	196.367											
Diffusanteil [%]	33	90											
Lufttemperatur [°C]	-3,3	22,8											
ANLAGEDATEN	geographische Breite [°] 47,05 Leitungslänge außen [m] 125 Glykolgehalt [%] 33 Wärmekapazität [J/m ² K] 6040												
AUSGABE													
MONATSPROGNOSE	<table border="1"> <tbody> <tr> <td>46,17</td> <td>kWh/m²</td> <td>34,28</td> <td>58,06</td> </tr> <tr> <td></td> <td></td> <td>MIN</td> <td>MAX</td> </tr> </tbody> </table>	46,17	kWh/m ²	34,28	58,06			MIN	MAX				
46,17	kWh/m ²	34,28	58,06										
		MIN	MAX										
MONATSPROGNOSE	<table border="1"> <tbody> <tr> <td>57.371</td> <td>kWh</td> <td>42.595</td> <td>72.146</td> </tr> </tbody> </table>	57.371	kWh	42.595	72.146								
57.371	kWh	42.595	72.146										



A.5 Weitere Tages- und Monats-Prognosefiles

- PROGNOSE_Tag_Koordinaten
- PROGNOSE_Tag_geneigt_Koordinaten
- PROGNOSE_Tag_Graz_geneigt
- PROGNOSE_Monat_Koordinaten
- PROGNOSE_Monat_geneigt
- PROGNOSE_Monat_Graz_geneigt

Literaturverzeichnis

- [1] **Brosius F.:** *SPSS 8.0: Professionelle Statistik unter Windows*. Bonn: MITP-Verlag GmbH, 1998, ISBN 3-8266-0437-7.
- [2] **Bühl A., Zöfel P.:** *SPSS für Windows Version 6.1: Praxisorientierte Einführung in die moderne Datenanalyse*. 3. Auflage, Bonn [u.a.]: Addison-Wesley-Longmann, 1996, ISBN 3-8273-1134-9.
- [3] **Fahrmeir L., Künstler R., Pigeot I., Tutz G.:** *Statistik: Der Weg zur Datenanalyse*. 2. Auflage, Springer-Verlag, Berlin, Heidelberg, New York, 1999, ISBN 3-540-65053-9.
- [4] **Friedl H.:** *Computerunterstützte Statistik*. Vorlesungsskriptum, Institut für Statistik, TU Graz, 1993.
- [5] **Hussain M.:** *Wirtschaftsstatistik*. Vorlesungsskriptum, Institut für Angewandte Statistik und Systemanalyse, Joanneum Research, 1998.
- [6] **Müller W.:** *Statistik für Maschinenbauer (Teil II)*. Vorlesungsskriptum, 3. Auflage, Institut für Statistik, TU Graz, 2010.
- [7] **Naredi-Rainer H.:** *Residuenanalyse in der multiplen linearen Regression*. Diplomarbeit, Technisch-Naturwissenschaftliche Fakultät, TU Graz, 1986.
- [8] **Sponer B.:** *Modellierung von Projektparametern mittels multipler Regression*. Diplomarbeit, Institut für Statistik, TU Graz, 2009.
- [9] **Stadlober E.:** *Angewandte Statistik*. Vorlesungsskriptum, Institut für Statistik, TU Graz, 2001-2011.
- [10] **Stadlober E.:** *Statistik für Informatikstudien*. Vorlesungsskriptum, Institut für Statistik, TU Graz, 2011.

- [11] **Stadlober E.:** *Wahrscheinlichkeitstheorie für Informatikstudien*. Vorlesungsskriptum, Institut für Statistik, TU Graz, 2011.
- [12] **Stadlober E., Schauer J.:** *Statistik: Bakkalaureat Technische Mathematik*. Vorlesungsskriptum, Institut für Statistik, TU Graz, 2007-2011.