Arbeitsmarktservice
Österreich

Update der Clustereinteilung österreichischer Arbeitsmarksprengel

Endbericht

Projektleitung AMS:
Maria Hofstätter, Mag. Sabine Putz

Projektleitung Statistik Austria:
Mag. Manuela Lenk, Mag. Irene Bettel

STATISTIK AUSTRIA
Die Informationsmanager

Wien, März 2013

Impressum
Medieninhaber und Herausgeber:
Arbeitsmarktservice Österreich
Bundesgeschäftsstelle
ABI/Arbeitsmarktforschung und Berufsinformation
Maria Hofstätter, Mag. Sabine Putz
A-1200 Wien, Treustraße 35-43
Tel: (+43 1) 331 78–0

ISBN 978-3-85495-…

Inhalt

Inhalt 3
1. Einleitung 4
2. Modell für das AMS-Projekt 2013 5
2.1 Methodische Vorgehensweise 5
3. Durchführung 9
3.1 Regressionsanalyse 9
3.1.1 Schrittweise Regressionsanalyse 9
3.1.2 Prüfung der Regressionsprämissen 10
3.1.3 Deskriptive Statistik der Clustermerkmale 13
3.2 Clusteranalyse 15
3.2.1 Clusteranalyse nach Ward 15
3.2.2 Clusteranalyse nach k-means 17
3.3 Güte des Clustermodells 19
4. Ergebnisse 20
4.1 Clusterzugehörigkeit 20
4.2 Typisierung der Cluster 22
4.3 Regionale Übersicht 25
Tabellenverzeichnis 27
Abbildungsverzeichnis 28
Literaturverzeichnis 29
Anhang 30

1. Einleitung

Unter Clusteranalyse versteht man strukturentdeckende, multivariate Analyseverfahren zur Ermittlung von Gruppen (Clustern) von Objekten, deren Eigenschaften oder Eigenschaftsausprägungen bestimmte Ähnlichkeiten aufweisen.
(Vgl.: Backhaus 2006)

Das Update des Clusterprojekts vom August 2011 wurde im Frühjahr 2013 mit den aktuellsten Regionaldaten aus der Abgestimmten Erwerbsstatistik erneut durchgeführt.
Das Grundkonzept der Clusteranalyse wird im Endbericht von 2009 (http://www.ams-forschungsnetzwerk.at/deutsch/publikationen/BibShow.asp?id=5664&sid=119669819&look=2&jahr=2009) ausführlich diskutiert.

Der vorliegende Bericht beschreibt das Update des Clustermodells, das analog zum Update von 2011 nach den Vorgaben der AMS-Projektleitung durchgeführt wurde:

1. Verwendung der aktuellsten Daten: Das Clusterprojekt von 2009 wurde mit den Daten der Probezählung 2006 und der Volkszählung 2001 durchgeführt. Das aktuelle Update nutzt die Daten der Abgestimmten Erwerbsstatistik zum Stichtag 31.10.2010 und Teile der Volkszählung 2001 für Merkmale, die nur bei der Volkszählung erhoben werden.
2. Berücksichtigung der Erwerbsbevölkerung: Die Vorgaben legen ihren Fokus auf die Erwerbsbevölkerung. D.h. dass die regionalen Geschäftsstellen (RGS) in erster Linie nach deren Größe in Cluster eingeteilt werden und erst in zweiter Linie nach regionalen Merkmalen.
3. Festgesetzte Einteilung von bestimmten RGS: Mit den Landesgeschäftsführern wurde vereinbart, dass einige RGS vorab in zwei Cluster eingeteilt werden sollen. Diese RGS fließen somit nicht mehr in die Clusteranalyse ein und werden in diesem Bericht nicht berücksichtigt.

Großstadtcluster:

409	Linz
504	Salzburg
630	Graz West
631	Graz Ost
958	Wien Johnstraße
959	Wien Hauffgasse
960	Wien Esteplatz Bank/Versich./EDV
961	Wien Dresdner Str. Metall/Chemie/Energie/Wasser
962	Wien Redergasse Einzelhandel
963	Wien Währinger Gürtel Gesundheit/Soziales/öffentl. Dienst
964	Wien Geiselbergstr. Verkehr/Agrar/Nahrung
965	Wien Schönbrunner Str. Personalbereitst./ REHA
966	Wien Hietzinger Kai Fremdenverkehr
967	Wien Huttengasse Bauwesen
968	Wien Schloßhofer Str. Großhandel/Werbung/Druck/Foto
969	Wien Prandaugasse Persönliche Dienste
970	Wien Jugendliche

Tourismuscluster:

202	Hermagor
501	Bischofshofen
505	Tamsweg
506	Zell am See
620	Gröbming
701	Imst
704	Kitzbühel
706	Landeck
708	Reutte
709	Schwaz
801	Bludenz

Neu bei diesem Update ist die vorab Einteilung der RGS Salzburg. Diese RGS wurde im Update von 2011 in die Clusterung mit einbezogen, im aktuellen Projekt jedoch ausgeklammert.
Ein Test der Regressionsanalyse und Clusterung sowohl nach Ward als auch nach k-means ergab dennoch eine starke Stabilität des Verfahrens, da sich nur zwei RGS in einem anderen Custer finden würden, als nach den aktuellen Vorgaben.

Die Berechnung des Modells wird in den angeschlossenen Kapiteln beschrieben und dessen Güte getestet. Am Ende des Modells steht die Zuteilung jeder österreichischen regionalen Geschäftsstelle (RGS) zu einem Cluster. Die Beschreibung der Unterschiede der einzelnen Cluster vervollständigt das Modell im letzten Schritt.

2. Modell für das AMS-Projekt 2013

2.1 Methodische Vorgehensweise

Die Methodik des Updates von 2011 wird 1:1 angewendet. Jedoch werden die Merkmale mit den Daten der Abgestimmten Erwerbsstatistik 2010 berechnet.

Im Folgenden wird ein kurzer Abriß über die Methode gegeben.

1. **Schritt: Berechnen der Merkmale**

Die Daten für die spätere Clusteranalyse beziehen sich in erster Linie auf die Daten der Abgestimmten Erwerbsstatistik der Statistik Austria zum Stichtag 31.10.2010. Diese Daten wurden aus verschiedensten Administrativdatenquellen und Registern gewonnen.

Zusätzlich werden die Volkszählungsdaten 2001 aus dem Hause der Statistik Austria auf aggregierter Gemeindeebene verwendet.

Jegliche Daten sind auf Gemeindeebene abgebildet und bilden somit eine solide Basis für die weiteren Berechnungsschritte. Aus den Gemeindedaten zu den jeweiligen Merkmalen wird ein Wert für die Regionalstellen (RGS) berechnet. Das bedeutet, dass die Absolutwerte der einzelnen Gemeinden summiert und die Anteilsberechnungen auf Basis der RGS durchgeführt werden. Diese RGS dienen als Grundstock für die Clusteranalyse, indem jede RGS ein abgegrenztes Objekt bildet.

Demografie:

- Anteil Erwerbsbevölkerung (ANTEIL_ERW):
Die Summe aller Personen im erwerbsfähigen Alter (zwischen 15 und 64) einer Region wird in Bezug zur gesamtösterreichischen Bevölkerung in dieser Altersgruppe gesetzt. Die Daten stammen aus der Abgestimmten Erwerbsstatistik 2010.

- Anteil Personen mit einem nicht-österreichischen Geburtsland der EU-15 (EU15):
Der Anteil dieser Personengruppe im Alter zwischen 15 und 64 Jahren wird in Bezug auf die Bevölkerung im erwerbsfähigen Alter (zwischen 15 und 64 Jahren) gesetzt. Die Daten stammen aus der Abgestimmten Erwerbsstatistik 2010.

- Bevölkerungsentwicklung (BEV_ENTW):
Dieses Maß wird aus der Veränderung der Bevölkerungszahl einer Region im erwerbsfähigen Alter (15 bis 64 Jahre) zwischen den Jahren 2001 (Jahr der Volkszählung) und 2010 (Jahr der Abgestimmten Erwerbsstatistik) berechnet. Der

Bevölkerungsstand 2001 gilt hierbei als Bezugsgröße. Die Bevölkerungsentwicklung wird in der prozentuellen Veränderung dieser beiden speziellen Bevölkerungsstände angegeben.

Einkommen:

- Einkommensmedian (EINKOMMEN_Q2):

Für diese Berechnung werden die personenbezogenen Lohnsteuerdaten 2010 verwendet. Als Grundgesamtheit gelten alle unselbständig Erwerbstätigen, die im Referenzjahr lohnsteuerpflichtig waren. Ausgenommen sind somit Pensionisten, Selbständige, Mithelfende und im Haushalt Tätige. Es wird der Median des Bruttojahreseinkommens aus eigener Erwerbstätigkeit pro RGS berechnet.

Gender-Mainstreaming:

- Einkommensdifferenzen bei Männern und Frauen (DIFF_EINKOMMEN_VZ):

Für die Berechnung werden aus den Lohnsteuerdaten 2010 alle Personen extrahiert, die im Jahr 2010 vollzeit und ganzjährig unselbständig beschäftigt waren. Ausgenommen sind somit Pensionisten, Selbständige, Mithelfende und im Haushalt Tätige. Der durchschnittliche Bruttojahreslohn der Männer wird in Bezug zum durchschnittlichen Bruttojahreslohn der Frauen gesetzt, wobei der Bezug der Männer als 100% betrachtet wird. Die Einkommensdifferenz gibt den Prozentsatz an, mit dem sich das Durchschnittseinkommen von Frauen und Männern unterscheidet.

Arbeitsmarkt:

- Saisonspanne (SAISONSPANNE):

Die Saisonspanne bildet den relativen Ausschlag des Arbeitslosenbestandes einer Region ab. Das Maximum des monatlichen Arbeitslosenbestandes bildet 100%, während das Minimum mit einem Prozentsatz davon angegeben wird. Diese Daten wurden vom AMS für jeden Monat des Jahres 2010 bereitgestellt.

Die Berechnung der Saisonspanne wurde in mehreren Variationen getestet. Da sich durch die unterschiedlichen Berechnungen keine gravierenden Änderungen in den Clusterzuordnungen ergeben haben, wurde die oben beschriebene Berechnungsart beibehalten.

- Verhältnis Vollzeit-Teilzeit Beschäftigungen (VZ_TZ):

Die Anzahl von unselbständigen Teilzeitbeschäftigten wird in Bezug zur Anzahl von Vollzeitbeschäftigten einer Region gesetzt. Dabei gelten die Vollzeitbeschäftigten als 100% und die Teilzeitbeschäftigten bilden einen gewissen Prozentsatz davon. Die Daten hierzu stammen aus der Abgestimmten Erwerbsstatistik 2010.

- Erwerbstätigenquote (ERWERBST):

Die Erwerbstätigen einer Region werden als Anteil an deren Bevölkerung im Alter zwischen 15 und 64 berechnet. Zu den Erwerbstätigen zählen Selbständige, Unselbständige, Mithelfende und temporär von der Arbeit Abwesende (z.B. durch Krankenstand). Hierzu werden die Daten der Abgestimmten Erwerbsstatistik 2010 verwendet.

Pendler:

- Anteil Arbeitsmarktbezirkspendler (PEND_RGS):

Mit Hilfe der Volkszählungsdaten 2001 wird der Anteil der Erwerbspendler an den Personen im erwerbsfähigen Alter (zwischen 15 und 64) berechnet, die in einem anderen Arbeitsmarktbezirk arbeiten als sie wohnen.

2. Schritt: Regressionsanalyse für die Gewichtung der Merkmale

Da die Auswahl der Merkmale bereits vorab getroffen wurde, entfällt die Analyse der Korrelationsmatrix. Außerdem werden durch die schrittweise Regressionsanalyse hochkorrelierte Merkmale ohnedies ausgeschlossen.

Eine multiple Regressionsanalyse wird durchgeführt, um die Wirkung mehrerer unabhängiger Variablen auf eine abhängige Variable zu messen. Diese Wirkung ist nur in eine Richtung denkbar. Somit darf die abhängige Variable keinen Einfluss auf die unabhängigen ausüben.
Eine Regressionsanalyse baut auf einer Kausalbeziehung (je-desto) auf.
Es wird getestet, ob die Null-Hypothese, die jegliche Wirkung der Merkmale auf die Zielvariable verneint, verworfen werden kann. Mit dem Verwerfen der Null-Hypothese gilt ein Zusammenhang als erwiesen.

Alle Merkmale werden in die Regressionsanalyse miteinbezogen. Es werden standardisierte Werte verwendet, um die Vergleichbarkeit zu gewährleisten.
Nach der Berechnung der Regressionskoeffizienten kann für jedes Merkmal die Aussage getroffen werden, wie hoch ihr Einfluss auf die abhängige Variable, also die Arbeitslosigkeit in der betreffenden Region, ist. Das bedeutet, je stärker sich die unabhängige Variable, z.B. die Erwerbsquote, ändert, umso stärker ändert sich die abhängige Variable, die Arbeitslosigkeit. Das Signifikanzniveau gibt zudem Auskunft darüber, ob die Null-Hypothese verworfen werden kann, oder nicht.

Mit Hilfe des Regressionskoeffizienten kann der t-Wert errechnet werden, indem er durch dessen Standardabweichung dividiert wird. Dieser Wert wird in der folgenden Clusteranalyse als Gewichtung zur Clusterbildung herangezogen. Der t-Wert gilt als Maß für die Einflusshöhe jedes Merkmals auf die Zielvariable.
Die Gewichtung ist in der Clusteranalyse von hohem Wert, da ansonsten jedem Merkmal dieselbe Aussagekraft auf die Clusterbildung unterstellt wird. Um diese fälschliche Aufteilung zu umgehen, werden die t-Werte jedes Merkmals als Gewichtung in den folgenden Schritten herangezogen.
(Vgl.: Backhaus 2006)

3. Schritt: Clusteranalyse mit gewichteten Merkmalen nach dem Ward-Verfahren

Bei der Clusteranalyse werden Objekte zu Gruppen zusammengefasst, die einander so ähnlich wie möglich sind, und andererseits sollen die Gruppen untereinander so unterschiedlich wie möglich sein.
In diesem Projekt gelten die RGS als Objekte, die es zusammenzufassen gilt.

Die Bildung von Clustern erfolgt über ausgewählte Merkmale. Die entstandenen Gruppen sollen eine Homogenität innerhalb und eine Heterogenität zueinander aufweisen.
Für die Bildung der Cluster werden wiederum standardisierte Werte der Merkmale verwendet, um einen Vergleich möglich zu machen. Wären die Werte nicht standardisiert, würden die unterschiedlichen Skalen einen Einfluss auf die Clusterbildung ausüben, der nicht erwünscht ist.
Die standardisierten Werte der Merkmale aller RGS werden mit den Gewichten (t-Werten aus der Regression) multipliziert und dienen als Ausgangspunkt für die Clusteranalyse.

Durch Berechnung der Distanzmaße (die quadrierte euklidische Distanz) kann die Distanz zwischen zwei Objekten bzw. deren Unähnlichkeit bestimmt werden.
Das Ward-Verfahren als hierarchisch-agglomeratives Verfahren ermittelt mit Hilfe des Distanzmaßes zwei Objekte, die einander am ähnlichsten sind und fasst diese zu einem Cluster zusammen. Im nächsten Schritt werden die Objekte vereint, die die kleinste Zunahme der Varianz verursachen, d.h. einander so ähnlich wie möglich sind und die

Homogenität innerhalb einer Gruppe aufrecht erhalten. Ab dem zweiten Schritt können als Objekte entweder bereits gebildete Cluster dienen, oder einzelne RGS, die noch in keinem Cluster enthalten sind.
Das bedeutet, dass am Anfang alle Objekte (RGS) einzeln bestehen und am Ende des Prozesses alle Objekte in einem Cluster zusammengefasst sind.

Es muss nun entschieden werden, welcher Zwischenschritt am besten erscheint, um die Clustereinteilung darzustellen. Dies geschieht mittels Prüfung der Heterogenitätsmaße, die bei jedem Schritt der Clusterbildung messbar sind. Es wird die Stufe der Clusterbildung gewählt, bei dem das Heterogenitätsmaß innerhalb der gebildeten Cluster (auch als Koeffizient in den Tabellen angeführt bzw. aus dem Dendogramm ersichtlich) möglichst klein ist. Dieses Maß gibt an, wie hoch die Fehlerwahrscheinlichkeit ist, wenn die Objekte zu einem Cluster zusammengeführt werden.
(Vgl.: Backhaus 2006)

Die so ermittelte Anzahl der Cluster wird im nächsten Clusterverfahren aufgenommen und die Zuordnung der Objekte verfeinert.

4. Schritt: Clusteranalyse mit gewichteten Merkmalen nach dem k-means-Verfahren

Da das Ward-Verfahren keine spätere Auflösung von gebildeten Clustern zulässt, ist die Zuordnung der Objekte zu den Clustern nicht immer optimal.
Um diesen Nachteil auszugleichen wird die Clusteranalyse mittels k-means-Verfahren erneut durchgeführt.

Das k-means-Verfahren – auch Clusterzentrenanalyse genannt - baut ebenfalls auf Distanzmaßen auf, jedoch wird die Distanz immer vom Clustermittelpunkt, dem Clusterzentroid aus, berechnet. Vorab muss jedoch die gewünschte Anzahl von Clustern angegeben werden.
Nach jedem Schritt der Clusterbildung werden die Zentroide neu berechnet und können auch zu einer Umschichtung der Objekte, von einem Cluster in einen anderen, führen. Dabei wird ebenfalls auf die Minimalhaltung der Varianz innerhalb der Cluster geachtet.
(Vgl.: Rüb 2007)

Als Hilfestellung für die feine Zuordnung wird die gewünschte Anzahl von Clustern nach dem Ward-Verfahren festgelegt und zusätzlich die Clusterzentren des Ward-Verfahrens übernommen. D.h. dass die Mittelwerte der Ward-Cluster berechnet werden und für das k-means-Verfahren als anfängliche Clusterzentren herangezogen werden. Dies unterstützt die Zuordnung zu den angegeben Clustern und festigt das Verfahren der Clusterbildung.

Die daraus entstandenen Cluster bilden das Endprodukt des Forschungsprojektes. Alle regionalen Geschäftsstellen des AMS finden sich in einem der berechneten Cluster wieder.

5. Schritt: Typologie der gebildeten Cluster

Es werden alle RGS betrachtet, die sich innerhalb eines Clusters befinden. Zum einen können „Problemfälle“ entdeckt werden, die ein großes Distanzmaß zum eigenen Clusterzentrum aufweisen, zum anderen können die Cluster anhand der Merkmale der einzelnen enthaltenen RGS beschrieben werden. Die Merkmale der einzelnen RGS werden in eine ähnliche Richtung weisen und können thematisch zusammengefasst werden. Es ist somit möglich jedem Cluster einen Typ zuzuweisen, der den Cluster beschreibt.

3. Durchführung

Dieses Kapitel beschäftigt sich mit der technischen Umsetzung des Modells, mit der Testung der Ergebnisse und Prüfung der statistisch mathematischen Korrektheit der Berechnungen. Hier sei noch einmal darauf hingewiesen, dass sich das gesamte Modell, also auch die Regressionsanalyse, nur auf jene 73 RGS bezieht, die im Modell verbleiben.
Die speziell eingeteilten RGS bleiben in allen folgenden Schritten unberücksichtigt.

3.1 Regressionsanalyse

Die Regressionsanalyse wird in zwei Schritten vollzogen, um die Merkmale und deren Wirkung auf die Zielvariable besser einschätzen zu können. Im ersten Schritt werden die Auswirkungen aller Merkmale gemeinsam auf die Zielvariable (Arbeitslosenquote 2010) gemessen und im zweiten Schritt werden die Merkmale ausgefiltert, die den stärksten Einfluss ausüben und voneinander unabhängig sind.

3.1.1 Schrittweise Regressionsanalyse

Hierbei wird in einem ersten Schritt mit einem Merkmal als Regressor begonnen. Im nächsten Schritt wird das Merkmal dazugenommen, das den höchsten t-Wert hat, d.h. den höchsten Einfluss auf die unabhängige Zielvariable. Dabei muss jedoch der p-Wert des Merkmals unter einer vorgegebenen Größe bleiben und somit ein bestimmtes Signifikanzniveau erreichen.
Sobald alle Merkmale geprüft wurden und die beste Modellgüte erreicht ist, ist das Modell abgeschlossen.
Die Güte des Modells wird anhand des R^2-Wertes angegeben. Ein R^2-Wert von 1 bedeutet, dass die gesamte Varianz der Zielvariablen durch die Regressoren (gewählten Merkmale) erklärt wird. Je näher der R^2-Wert bei 0 ist, umso weniger aussagekräftig ist das Modell.

Für das vorliegende Modell wurde für die Aufnahme in dieses das Signifikanzniveau von 0,5 und für das Verbleiben das Niveau von 0,2 gewählt.

Aus der folgenden Tabelle sind die Ergebnisse der stufenweisen Regression abzulesen, die Merkmale und deren t-Werte.

Tabelle 1: Schrittweise Regression

R-Quadrat 0.7908

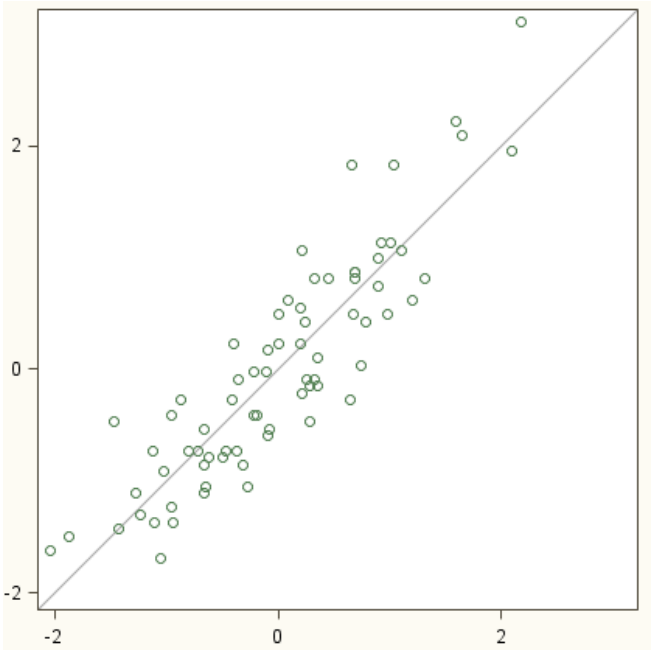
Variable	t-Wert	Pr > |t|
EU15	-2.17	0.0333
VZ_TZ	-4.40	<.0001
ERWERBST	-6.25	<.0001
DIFF_EINKOMMEN_VZ	-1.63	0.1077
EINKOMMEN_Q2	-5.89	<.0001
SAISONSPANNE	3.00	0.0039

Quelle: Berechnungen Statistik Austria mit Daten der Statistik Austria und des AMS Österreich

Es zeigt sich, dass die Merkmale VZ_TZ, ERWERBST und EINKOMMEN_Q2 die stärkste Wirkung auf die Zielvariable ausüben. Jedoch sind die Wirkungen aller Merkmale auf einem Signifikanzniveau von 0,2 erwiesen und können demnach in die Clusterbildung einfließen. Aus dem Modell herausgefallen sind durch das Nichterreichen des Signifikanzniveaus die Merkmale PEND_RGS, BEV_ENTW sowie das Merkmal ANTEIL_ERW.

Die hier vorgeschlagenen Variablen erklären bereits über 79% der Streuung der Arbeitslosenquote, wie aus dem R^2-Wert hervorgeht.
Die Güte des Modells zeigt sich sowohl im hohen R^2-Wert als auch in der Punktwolke (Streudiagramm) der Regression.

Grafik 1: Beobachtete ALQ nach vorhergesagter ALQ



Quelle: Berechnungen Statistik Austria mit Daten der Statistik Austria und des AMS Österreich

Mittels robuster Regression wurden die Effekte der Merkmale auf die abhängige Variable zusätzlich überprüft. Dabei werden die verzerrenden Auswirkungen von Ausreißern in den Daten ausgeschalten.
Nach der robusten Regression zeigen sich die t-Werte der Merkmale mit denselben Vorzeichen und denselben Größenverhältnissen. Somit kann das Ergebnis der stufenweisen Regression als robustes Modell gegenüber Ausreißern bestätigt werden.

Die sechs Merkmale, die aus der stufenweisen Regression hervorgehen werden als die Merkmale festgelegt, die die Clusteranalyse definieren sollen. In den Clusterbildungsschritten werden somit diese Merkmale (EU15, VZ_TZ, ERWERBST, DIFF_EINKOMMEN_VZ, EINKOMMEN_Q2 und SAISONSPANNE) mit ihren Gewichtungen ausschlaggebend für die Einteilung der RGS in Cluster sein.

3.1.2 Prüfung der Regressionsprämissen

Durch eine Prüfung der Modellprämissen einer Regressionsanalyse kann abschließend die Beobachtung gemacht werden, ob das Regressionsmodell stabil ist und die Ergebnisse für die Clusteranalyse herangezogen werden können.

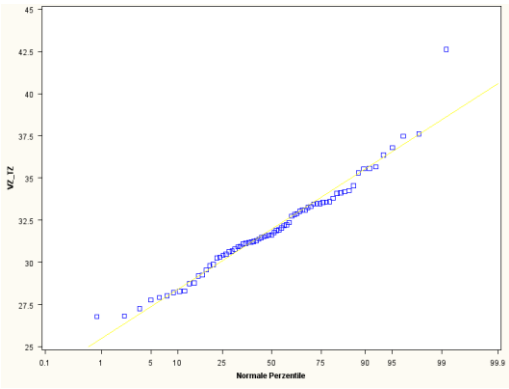
Diese Prämissenüberprüfung wird ausschließlich für die Merkmale gemacht, die am Ende der Regressionsanalyse stehen. Somit wird überprüft, ob das Ergebnis der schrittweisen Regression plausibel ist und die Merkmale, die später in die Clusteranalyse einfließen, auch dieser Prüfung standhalten können. Das Merkmal ANTEIL_ERW wird nicht dieser Prüfung unterzogen, da dieses Merkmal nicht im Ergebnis der Regressionsanalyse vorkommt.

Diese Modellprämissen zeigen sich wie folgt:

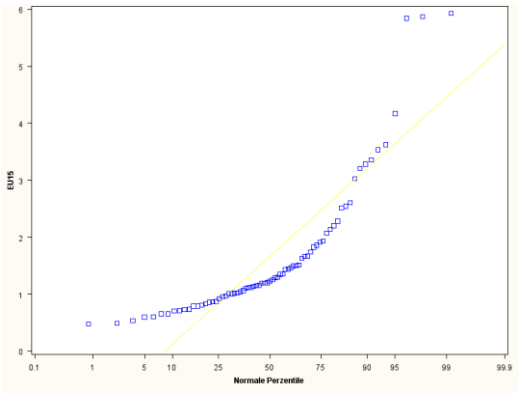
- Linearität in den Parametern: Alle unabhängigen Variablen werden daraufhin überprüft, ob sie linear verteilt sind. Dazu dient ein Plot über die beobachteten Werte. Im Folgenden finden sich die Plots aller fünf Merkmale wieder, die als Ergebnis der Schrittweisen Regressionsanalyse stehen.

Grafik 2: Verteilung der unabhängigen Variablen

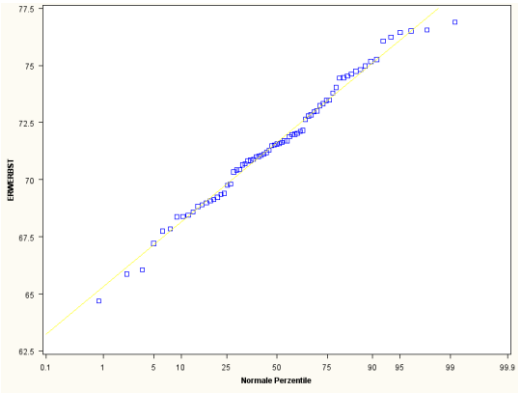
Variable: VZ_TZ



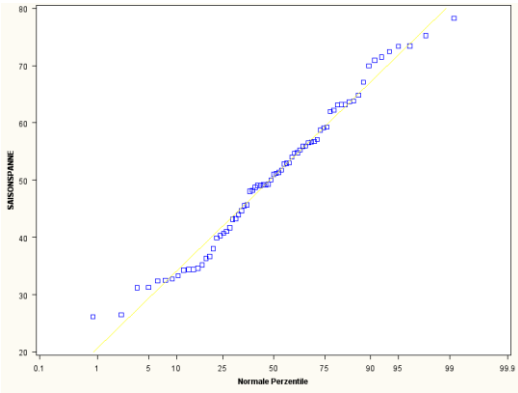
Variable: EU15



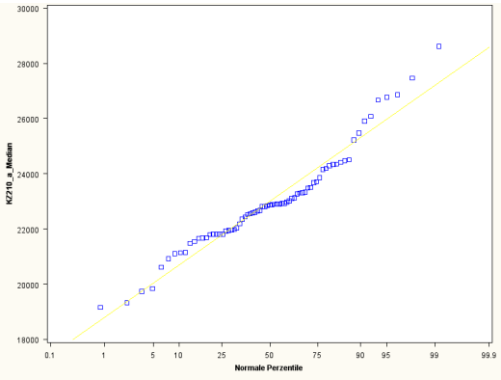
Variable: ERWERBST



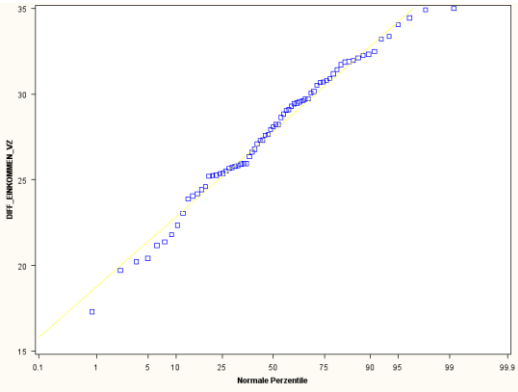
Variable: SAISONSPANNE



Variable: EINKOMMEN_Q2



Variable: DIFF_EINKOMMEN_VZ



Quelle: Berechnungen Statistik Austria mit Daten der Statistik Austria und des AMS Österreich (siehe Tabelle 1)

Aus diesen Grafiken geht hervor, dass alle Merkmale annähernd linear verteilt sind und somit die Prämisse erfüllt ist.

- Vollständigkeit des Modells: die Vollständigkeit des Modells kann nicht allgemein getestet werden. Für das Regressionsmodell wird unterstellt, dass alle Merkmale, die die abhängige Variable erklären im Modell vorhanden sind. Jedoch gibt der R^2-Wert von 0,7908 Auskunft darüber, wieviel Varianz der Zielvariable durch die unabhängigen Variablen erklärt wird. Über 79 % erklärte Varianz deuten auf eine relativ gute Auswahl von Merkmalen hin.

- Homoskedastizität der Störgrößen: Ob bei diesem Modell Heteroskedastizität besteht, d.h. die Streuung der Residuen in einer Reihe von geschätzten Werten nicht konstant ist, wird mit einem Test überprüft. Die Durchführung eines White-Tests ergibt bei einem Signifikanzniveau von 5 % eine Bestätigung der Annahme von Homoskedastizität und erfüllt somit diese Regressionsprämisse.

- Unabhängigkeit der Störgrößen: Um eine eventuell vorhandene Autokorrelation zu erkennen, die zu ineffizienten Schätzungen führen kann, wird ein Durbin/Watson-Test durchgeführt. Der d-Wert von 1,805 weist auf keine starke Autokorrelation hin. Der geschätzte Autokorrelationskoeffizient erster Ordnung weist einen Wert von 0,078 auf und deutet somit auf eine sehr schwache positive Autokorrelation hin, da sich die Werte in einem Bereich von -1 bis 1 bewegen können. Durch diese niedrigen Werte kann die Prämisse der Unabhängigkeit der Störgrößen als erfüllt betrachtet werden.

- Keine lineare Abhängigkeit zwischen den unabhängigen Variablen: Im Zuge der schrittweisen Regression wurden alle Merkmale entfernt, die einander stark beeinflussen und somit keinen zusätzlichen Nutzen für die Varianzerklärung der Zielvariablen haben.

  Zur zusätzlichen Überprüfung der Kollinearität wird der Varianzinflationsfaktor für jedes Merkmal der Regression berechnet. Dieser Faktor wird berechnet durch $1/1-R^2$ und zeigt eine starke Korrelation eines Merkmals mit einem Wert von über 10 auf.

  Tabelle 2: Kollinearitätstest

Variable	Varianz Inflation
EU15	1.60511
VZ_TZ	1.60710
ERWERBST	2.31504
DIFF_EINKOMMEN_VZ	1.54540
EINKOMMEN_Q2	2.75724
SAISONSPANNE	4.10000

Quelle: Berechnungen Statistik Austria mit Daten der Statistik Austria und des AMS Österreich

  Aus der Tabelle geht hervor, dass kein Merkmal des Regressionsmodells eine Multikorrelation aufweist und diese Prämisse somit ebenfalls erfüllt ist.

- Normalverteilung der Störgrößen: Mit einer Stichprobengröße kleiner 40 ist es möglich, dass durch die nicht normalverteilten Störgrößen alle Signifikanztests ungültig werden. Durch die hohe Anzahl der Beobachtungen wird diese Annahme nicht getestet.

Abschließend kann festgehalten werden, dass die durchgeführte Regressionsanalyse alle geforderten Prämissen einhält und somit als stabiles Modell angesehen werden kann. (Vgl. Backhaus 2006, Krämer 2005)

3.1.3 Deskriptive Statistik der Clustermerkmale

Zusätzlich zu den sechs Merkmalen, die sich aus der Regressionsanalyse ergeben, wird das Merkmal ANTEIL_ERW in die Clusterbildung aufgenommen. Dieses Merkmal wird auf Wunsch der AMS-Projektleitung mit einem starken Gewicht versehen und wird – auch wenn in der Regressionsanalyse die Wirkung auf die Arbeitslosenquote nicht signifikant war – gewichtet in das Clustermodell einfließen: t-Wert = 10

Alle Clustermerkmale und die Zielvariable wurden unstandardisiert und ungewichtet einer deskriptiven Statistik unterzogen, um deren Eigenschaften aufzuzeigen.

Tabelle 3: Deskriptive Statistik der Clustermerkmale

Variable	Spannweite	Minimum	Maximum	Mittelwert	Standard-abweichung	Varianz	Schiefe	Kurtosis
EU15	5,45	,48	5,93	1,6616	1,20534	1,453	2,068	4,498
VZ_TZ	15,83	26,79	42,62	31,9785	2,79186	7,794	,733	1,907
ERWERBST	12,21	64,69	76,91	71,6254	2,71660	7,380	-,098	-,265
ANTEIL_ERW	3,26	,21	3,47	,8358	,49826	,248	2,349	9,766
DIFF_EINKOMMEN_VZ	17,69	17,30	35,00	27,8243	3,89585	15,178	-,363	-,217
EINKOMMEN_Q2	9454,91	19164,63	28619,54	23003,2671	1803,22441	3251618,282	,698	1,136
SAISONSPANNE	52,09	26,19	78,28	50,6860	12,90296	166,486	,112	-,727
ALQ	7,50	3,70	11,20	6,3384	1,56486	2,449	,647	,340

Quelle: Berechnungen Statistik Austria mit Daten der Statistik Austria und des AMS Österreich

Lagemaße:

Nahezu alle Merkmale werden in einem Prozentanteil gemessen und sind somit auf die Spanne zwischen 0 und 100 begrenzt.
Für das Merkmal EINKOMMEN_Q2 trifft dies jedoch nicht zu, da es sich um den Bruttojahreslohn in Euro handelt. Aus diesem Grund liegt der Mittelwert dieses Merkmals auch deutlich höher als die der anderen Merkmale.

Die Werte des Minimums und Maximums sind bei fast allen Merkmalen nicht allzu weit vom Mittelwert entfernt, was auf nicht allzu extreme „Ausreißerwerte“ hinweist. Die Spannweite des Merkmals SAISONSPANNE ist um einiges höher als bei allen anderen Merkmalen und zeigt damit die großen Unterschiede zwischen den RGS in diesem Merkmal auf.

Streuungsmaße:

Die Standardabweichung einiger Merkmale ist höher als bei anderen. Die Merkmale EINKOMMEN_Q2 und SAISONSPANNE weisen hohe Schwankungen auf. Die durchschnittlichen Abweichungen der Werte der einzelnen RGS zum Mittelwert dieser beiden Merkmale sind demnach hoch. Alle anderen Merkmale weisen eine eher geringe Standardabweichung auf.

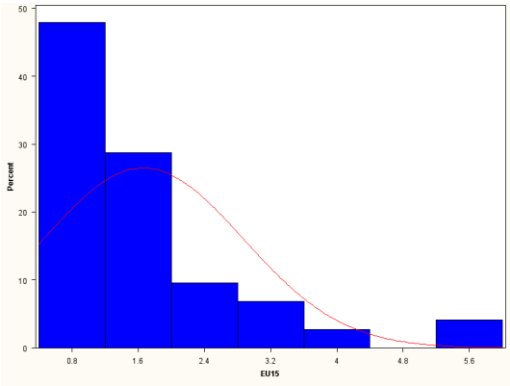
Verteilung:

Die Kurtosis (auch Exzess oder Wölbung) ist ein Maß für die relative "Flachheit" einer Verteilung. Bei einer Normalverteilung einer Variablen ergibt sich eine Schiefe von Null.
Bei einer Kurtosis zwischen -0,3 und 0,3 spricht man von einer mittelgewölbten Verteilung (Normalverteilung), bei negativen Werten von einer breitgewölbten Verteilung und bei positiven Werten von einer schmalgewölbten Verteilung.

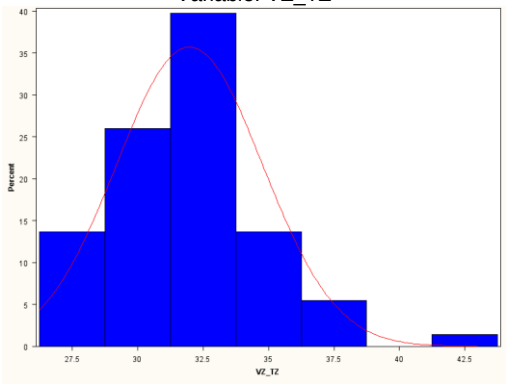
Das Merkmal SAISONSPANNE zeigt eine breitgewölbten Verteilung, während die Merkmale EU15, ANTEIL_ERW, EINKOMMEN_Q2, VZ_TZ und ALQ eine Schmalwölbung aufweisen. Die restlichen Merkmale erscheinen mittelgewölbt und somit annähernd normalverteilt.

Grafik 3: Verteilung der Clustermerkmale und ALQ

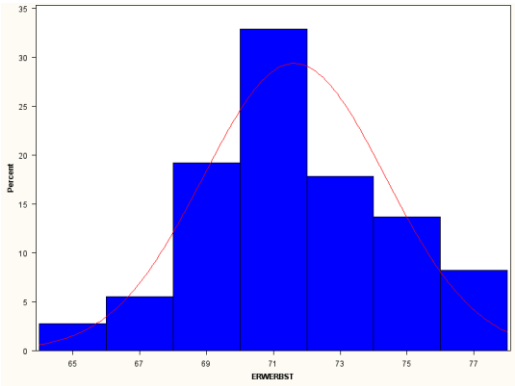
Variable: EU15



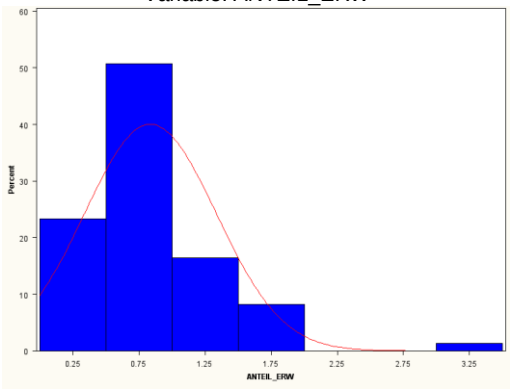
Variable: VZ_TZ



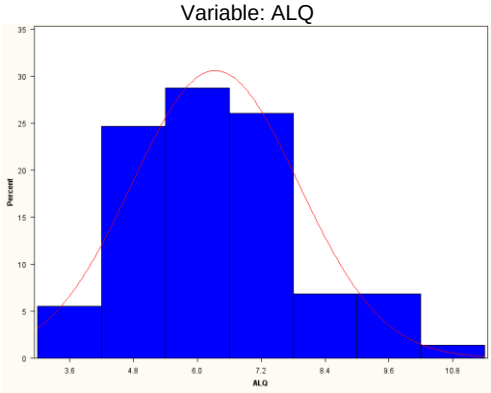
Variable: ERWERBST



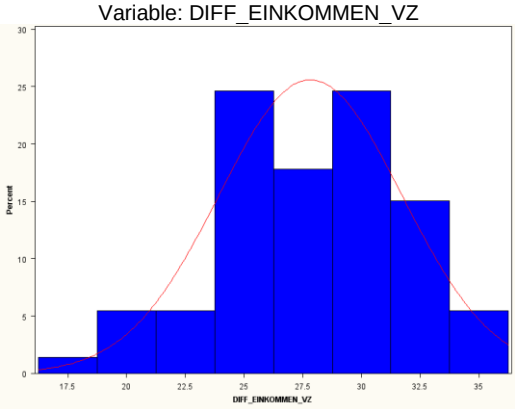
Variable: ANTEIL_ERW



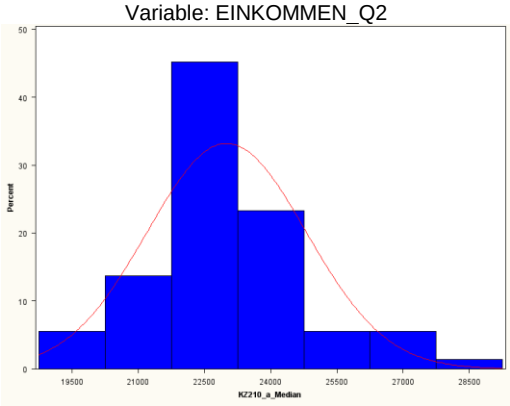
Variable: ALQ



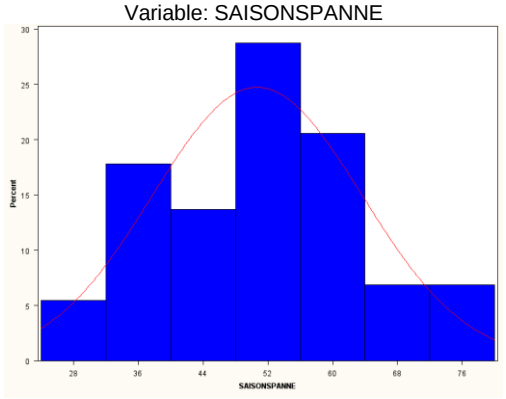
Variable: DIFF_EINKOMMEN_VZ



Variable: EINKOMMEN_Q2



Variable: SAISONSPANNE



Quelle: Berechnungen Statistik Austria mit Daten der Statistik Austria und des AMS Österreich

Die Schiefe einer Verteilung gibt an, ob die meisten Beobachtungen eher im unteren Bereich oder im oberen liegen. In einer Normalverteilung liegen die meisten Beobachtungen um den Mittelwert und weisen somit eine Schiefe von Null bzw. eine annähernde Normalverteilung eine Schiefe von -0,3 bis 0,3 auf. Eine positive Schiefe weist auf eine linkszentrierte Verteilung hin, d.h. eher im unteren Bereich der Werteskala, während eine negative Schiefe auf eine rechtszentrierte Verteilung weist.
Die Merkmale EU15, VZ_TZ, ANTEIL_ERW, EINKOMMEN_Q2 und ALQ weisen eine linksschiefe Verteilung auf, während alle anderen Merkmale annähernd normalverteilt sind.

3.2 Clusteranalyse

Die Clusteranalyse wird in zwei Schritten durchgeführt. In einem ersten Schritt wird die optimale Clusterzahl mit dem Ward-Verfahren ermittelt, während in einem zweiten Schritt mittels k-means-Verfahren alle RGS einem Cluster zugeordnet werden.

3.2.1 Clusteranalyse nach Ward

Alle 73 RGS (also wiederum nur die RGS, die nicht vorab bereits vom AMS eingeteilt wurden) werden in die Clusteranalyse nach der Ward-Methode als Objekte eingelesen. Bei der Analyse werden die Merkmale herangezogen, die nach der schrittweisen Regression den signifikant höchsten Einfluss auf die Zielvariable aufweisen, und zusätzlich Merkmal ANTEIL_ERW, das durch die AMS-Projektleitung bestimmt wurde.
Die Ausprägungen der Merkmale werden einer Z-Transformation unterzogen und mit den Gewichten (t-Wert) aus der Regressionsanalyse multipliziert:
EU15*-2,17
VZ_TZ*-4,40
ERWERBST*-6,25
DIFF_EINKOMMEN_VZ*1,63
EINKOMMEN_Q2*-5,89
SAISONSPANNE*3,00
ANTEIL_ERW*10

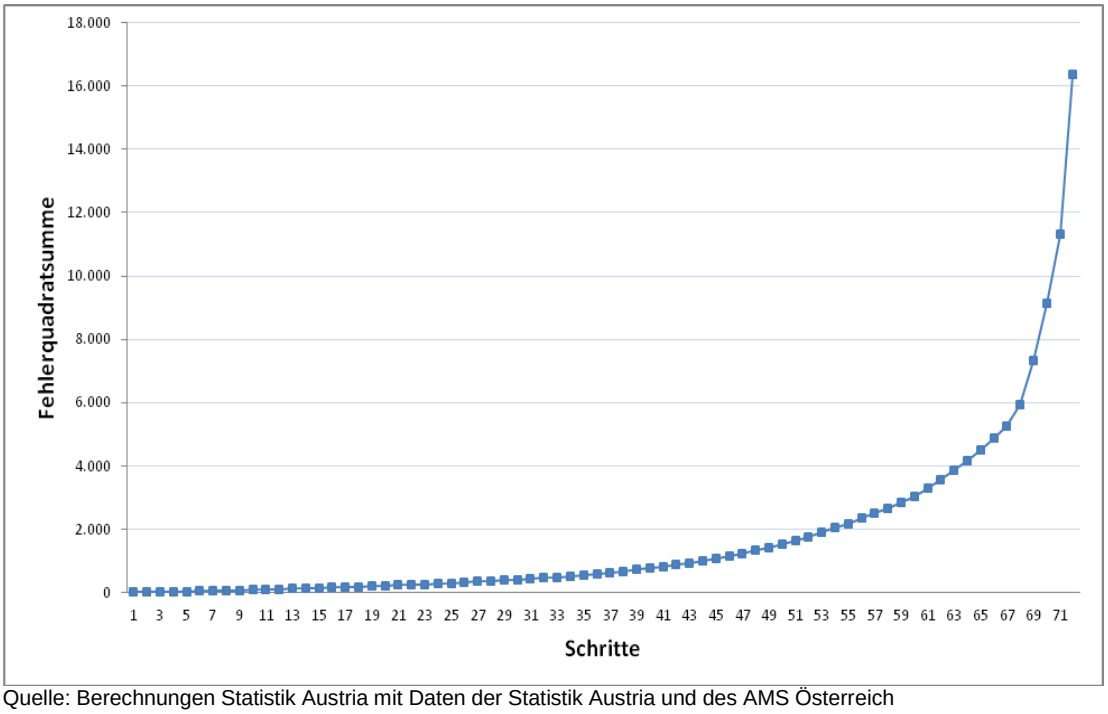
Das Ward-Verfahren entwickelt Lösungen für alle möglichen Clusteranzahlen. Als Distanzmaß wird die quadrierte euklidische Distanz (d) ausgewählt. Diese berechnet sich aus der Summe der quadrierten Abweichungen aller Merkmale zwischen den Objekten, mittels folgender Formel:

$$d(X,Y) = \sqrt{\sum_{i=1}^{n}(X_i - Y_i)^2}$$

Zur Entscheidung, welche Anzahl von Clustern die statistisch Beste ist, kommt man über folgende Analyseschritte:

1. Elbow-Kriterium: Hierbei werden die Koeffizienten der Ward-Clusteranalyse in ein Diagramm eingetragen. Die so sichtbar gemachte Heterogenitätsentwicklung weist einen „Ellbogen“ auf, bei dem die Kurve der Unähnlichkeit steil nach oben geht. Dieser Punkt zeigt die Clusterstufe an, über der verhältnismäßig große Fehler bei der Clusterbildung begangen werden.

Grafik 4: Heterogenitätsentwicklung nach der Ward-Clusteranalyse



Quelle: Berechnungen Statistik Austria mit Daten der Statistik Austria und des AMS Österreich

In der Grafik zeigt sich, dass bei Schritt 67 ein „Ellbogen" ersichtlich ist, d.h. dass ab diesem Punkt die Fehlerquadratsumme stark ansteigt und somit die Clusterzusammenführung nur mit großen Fehlern durchgeführt werden kann. Unterhalb dieses Ellbogens zeigen sich 6 gebildete Cluster.

2. Zuordnungsübersicht: Aus der Zuordnungsübersicht der Objekte zu Clustern nach jedem Schritt kann das Heterogenitätsmaß jeder Clusterzusammenführung gelesen werden. Der angegebene Koeffizient bezeichnet somit die Unähnlichkeit der beiden Objekte, die in diesem Schritt zusammengefasst werden.

Tabelle 4: Zuordnungsübersicht nach der Ward-Clusteranalyse

Schritt	Zusammengeführte Cluster			Erstes Vorkommen des Clusters		
	Cluster 1	Cluster 2	Koeffizienten	Cluster 1	Cluster 2	Nächster Schritt
63	54	68	3860,595	31	0	66
64	10	37	4164,784	46	53	66
65	9	72	4503,828	61	32	69
66	**10**	**54**	**4874,970**	**64**	**63**	**71**
67	**3**	**8**	**5258,764**	**56**	**24**	**70**
68	1	4	5928,884	62	57	70
69	9	16	7324,002	65	52	72

Quelle: Berechnungen Statistik Austria mit Daten der Statistik Austria und des AMS Österreich

Auch aus dieser Tabelle ist der Sprung des Unähnlichkeitsmaßes zwischen den Schritten 66 und 67 zu erkennen. Bis zu diesem Anstieg wurden bereits 6 Cluster gebildet.

3. Dendogramm: Ein Dendogramm (siehe Anhang) verbildlicht die Clusterzusammenführungen mit den zugehörigen Heterogenitätsmaßen. Die Heterogenitätsentwicklung wird auf einer Skala von 0 bis 25 normiert. Wobei der Wert 25 immer das Endstadium bezeichnet, in dem alle Objekte in einem einzigen Cluster zusammengeführt wurden.

   Es zeigt sich, dass ab einem normierten Heterogenitätsmaß von 5 die Clusterzusammenführungen nur noch mit großen Unähnlichkeiten von statten gehen

können. Darunter wurden bereits 6 Cluster gebildet, deren Unähnlichkeiten um vieles geringer sind, als in den darauffolgenden Schritten.

Aus den eben erwähnten Analysen kann die Schlussfolgerung gezogen werden, dass eine Clustereinteilung von 6 Clustern als optimal angesehen werden kann.

(Vgl.: Backhaus 2006)

3.2.2 Clusteranalyse nach k-means

Die nach dem Ward-Verfahren ermittelte Zahl von Clustern wird in einer neuerlichen Clusteranalyse nach dem k-means Verfahren (Clusterzentrenanalyse) festgelegt. Zusätzlich werden die Clustermittelwerte des ersten Verfahrens mit in die Analyse aufgenommen. Diese Zentroide bilden die Ausgangswerte für das k-means Verfahren. Als Distanzmaß wird hier ebenfalls das quadrierte euklidische Distanzmaß eingesetzt.

Im Laufe der Clusterzuordnungen werden die Clustermittelwerte neu berechnet, da die Distanzen zwischen den Objekten innerhalb eines Clusters und des Clusterzentroids nicht zu groß werden dürfen.

Am Ende der Clusteranalyse nach k-means stehen die 6 gebildeten Cluster, deren zugehörige Objekte, Clusterzentren und Distanzen zu den Clustermittelpunkten fest.

In einem ersten Schritt werden die Eckdaten des Clusterergebnisses analysiert und in einem zweiten Schritt wird die Güte des Modells getestet:

1. ANOVA-Test: Alle verwendeten Merkmale für die Clusterbildung werden einem ANOVA-Test unterzogen. Hierbei wird ermittelt, ob die Werte innerhalb der Cluster einander stark ähnlich sind, während sie sich zwischen den Clustern stark unterscheiden. Dieser Test ist ein Hinweis dafür, ob die Heterogenität zwischen den Clustern und die Homogenität innerhalb der Cluster gegeben ist. Der Quotient aus der Quadratsumme zwischen den Clustern bzw. innerhalb der Cluster wird in der Spalte F ausgegeben. Wenn der F-Wert hoch ist, während das Signifikanzniveau niedrig ist, deutet das auf eine qualitativ hochwertige Clusterbildung hin.

Tabelle 5: ANOVA-Test nach der k-means-Clusteranalyse

	Cluster		Fehler			
	Mittel der Quadrate	df	Mittel der Quadrate	df	F	Sig.
EU15	31,989	5	2,673	67	11,967	,000
VZ_TZ	126,970	5	11,329	67	11,207	,000
ERWERBST	384,498	5	13,284	67	28,945	,000
DIFF_EINKOMMEN_VZ	14,399	5	1,781	67	8,087	,000
EINKOMMEN_Q2	288,733	5	15,734	67	18,351	,000
SAISONSPANNE	70,530	5	4,408	67	16,000	,000
ANTEIL_ERW	1202,276	5	17,741	67	67,770	,000

Quelle: Berechnungen Statistik Austria mit Daten der Statistik Austria und des AMS Österreich

Das Mittel der Quadrate der Cluster gibt die Distanz zwischen den Clustern an, während das Mittel der Quadrate der Fehler die Distanz zwischen den RGS innerhalb eines Clusters beschreibt. Aus der Tabelle ist zu lesen, dass das Merkmal ANTEIL_ERW die höchsten Distanzen innerhalb des Clusters aufweist, während die Merkmale EU15 und S DIFF_EINKOMMEN_VZ sehr kleine Distanzen zwischen den Clustern haben.

Allgemein liegt das Signifikanzniveau bei jedem Merkmal bei Null und ist als Hinweis auf eine gute Homogenität innerhalb der Cluster und eine hohe Heterogenität zwischen den Clustern zu sehen.

2. Distanz zwischen Clusterzentren: Die Distanz zwischen den Mittelpunkten der Cluster sollte nach der Clusterbildung relativ hoch sein.

Tabelle 6: Clusterzentrendistanz nach der k-means-Clusteranalyse

Cluster	1	2	3	4	5	6
1		16,307	18,716	15,300	16,652	56,970
2	16,307		11,184	21,119	10,340	**65,715**
3	18,716	11,184		16,528	15,575	59,695
4	15,300	21,119	16,528		18,148	45,205
5	16,652	10,340	15,575	18,148		60,264
6	56,970	65,715	59,695	45,205	60,264	

Quelle: Berechnungen Statistik Austria mit Daten der Statistik Austria und des AMS Österreich

In der Tabelle werden die quadrierten euklidischen Distanzen zwischen den 6 Clusterzentren veranschaulicht. Es zeigt sich, dass der größte Unterschied zwischen den Clustern 2 und 6 besteht, die stärkste Ähnlichkeit zwischen den Clustern 2 und 5. Allgemein sind die Distanzen zwischen den Clusterzentren zwischen 10,340 und 65,715 ausreichend hoch.

3. Aufteilung der Objekte: In der folgenden Tabelle ist die Anzahl der RGS je Cluster aufgelistet.

Tabelle 7: Anzahl Objekte nach der k-means-Clusteranalyse

Cluster	Anzahl RGS
1	10
2	15
3	13
4	14
5	20
6	1

Quelle: Berechnungen Statistik Austria mit Daten der Statistik Austria und des AMS Österreich

Aus der Tabelle geht hervor, dass Cluster 6 mit nur einer einzigen RGS befüllt ist, während Cluster 5 die meisten Objekte vereint.

Im folgenden Kapitel wird die Güte und Stabilität des Clustermodells getestet.

(Vgl.: Backhaus 2006)

3.3 Güte des Clustermodells

Die Güte des Clustermodells der 73 RGS wird in zwei Varianten getestet. Es sei wiederum ausdrücklich darauf hingewiesen, dass die Modellgüte nur für die verbliebenen 73 RGS berechnet werden kann, jedoch keine Aussagekraft für die zwei vorabgewählten Cluster haben.

1. Varianztest

Ist die Verteilung eines Merkmals innerhalb eines Clusters sehr heterogen im Vergleich mit dem Gesamtmodell, deutet das auf ein instabiles Merkmal in einem Cluster hin. Liegt der Quotient der Clustervarianz zur Gesamtvarianz über 1, ist das Merkmal in dem entsprechenden Cluster nicht besser verteilt, als insgesamt. Bei diesen Merkmalen kann man demnach nicht von einem homogenen Cluster sprechen.

Tabelle 8: Varianztest der k-means-Clusteranalyse

	EU15	VZ_TZ	ERWERBST	DIFF_EINKOMMEN_VZ	EINKOMMEN_Q2	SAISONSPANNE	ANTEIL_ERW
1	0,290	0,493	0,304	0,581	0,570	0,421	0,371
2	0,079	0,458	0,220	0,692	0,477	0,510	0,100
3	0,291	0,803	0,328	0,516	0,710	0,512	0,126
4	**1,739**	0,645	0,494	0,881	0,467	0,505	0,255
5	0,433	0,544	0,348	0,649	0,210	0,483	0,122
6	0,000	0,000	0,000	0,000	0,000	0,000	0,000

Quelle: Berechnungen Statistik Austria mit Daten der Statistik Austria und des AMS Österreich

Aus der Tabelle ist zu erkennen, dass Cluster 4 im Merkmal EU15 heterogen ist. D.h. in diesem Merkmal können sich die einzelnen RGS innerhalb des Clusters stärker unterscheiden. Da jedoch das Merkmal EU15 mit einem geringen Gewicht in die Clusterbildung eingegangen ist, wurde der Cluster dennoch gebildet.

Cluster 6 weist eine Varianz von Null auf, da dieser Cluster nur aus einem einzigen Objekt besteht.

Alle anderen Cluster und Merkmale zeigen eine geringere Varianz innerhalb der Cluster als in der Gesamtheit. Das bedeutet, dass die Clustereinteilung die Merkmalsverteilungen gut zusammenfassen kann.

2. Test der Distanzmaße

Als zusätzlicher Hinweis auf die Güte des Clustermodells kann die Verwendung des Distanzmaßes in Frage gestellt werden. Bei dem vorgestellten Modell wurde für die Clusterbildung das quadrierte euklidische Distanzmaß herangezogen.

Da dieser Test bereits im Modell aus 2009 durchgeführt wurde sei hier nur erwähnt, dass aus inhaltlichen Gründen die Verwendung der quadrierten euklidischen Distanz als bestmögliches Distanzmaß festgestellt werden konnte (siehe http://www.ams-forschungsnetzwerk.at/deutsch/publikationen/BibShow.asp?id=5664&sid=119669819&look=2&jahr=2009).

4. Ergebnisse

Die Ergebnisse der Clusteranalyse nach k-means werden im Folgenden aus einem fachlichen Blickwinkel betrachtet.

4.1 Clusterzugehörigkeit

Die 73 RGS wurden im oben beschriebenen Clusterverfahren zu 6 Clustern zusammengeführt.
Die folgende Tabelle zeigt die Clusterzugehörigkeit jeder dieser RGS an und deren Distanz zum Clusterzentrum. Hohe Distanzen weisen auf eine Randlage der betroffenen RGS hin, während geringe Distanzen RGS kennzeichnen, die typisch für diesen Cluster sind.

Wiederum können die vorabgewählten RGS des Tourismus- und Großstadtclusters hier nicht behandelt werden, da diese nicht in die Clusteranalyse eingegangen sind.

Tabelle 9: Clusterzugehörigkeit RGS und Distanz zum Clusterzentrum

RGS_NR	RGS_NAME	CLUSTER	DISTANZ
101	Eisenstadt	1	7,837
102	Mattersburg	2	6,604
103	Neusiedl am See	2	8,849
104	Oberpullendorf	2	6,168
105	Oberwart	3	3,595
106	Stegersbach	2	7,238
107	Jennersdorf	2	7,797
201	Feldkirchen	3	10,127
203	Klagenfurt	4	11,419
204	Spittal/Drau	3	13,218
205	St. Veit/Glan	3	3,265
206	Villach	4	11,733
207	Völkermarkt	3	12,187
208	Wolfsberg	2	7,358
301	Amstetten	1	9,873
304	Baden neu	4	11,319
306	Bruck/Leitha	1	9,322
308	Gänserndorf	1	5,495
311	Gmünd	2	3,112
312	Hollabrunn	2	7,255
313	Horn	2	4,932
314	Korneuburg	1	5,948
315	Krems	5	8,983
316	Lilienfeld	2	4,828
317	Melk	5	6,121
319	Mistelbach	1	2,862
321	Mödling	1	12,784
323	Neunkirchen	1	8,924
326	St. Pölten	4	9,395
328	Scheibbs	5	6,748
329	Schwechat	1	8,820
331	Tulln	1	11,177
332	Waidhofen/Thaya	2	5,352
333	Waidhofen/Ybbs	5	7,005
334	Wr. Neustadt	4	9,934
335	Zwettl	5	8,752
401	Braunau	4	9,054
402	Eferding	5	6,422
403	Freistadt	5	8,860
404	Gmunden	4	8,539
406	Grieskirchen	5	3,000
407	Kirchdorf/Krems	5	4,131
411	Perg	5	6,608
412	Ried im Innkreis	5	2,382
413	Rohrbach	5	7,483

414	Schärding	5	6,453
415	Steyr	4	6,708
418	Vöcklabruck	4	7,375
419	Wels	4	4,892
421	Traun	5	8,575
503	Hallein	5	8,916
601	Bruck/Mur	3	7,047
603	Deutschlandsberg	2	6,823
604	Feldbach	5	7,186
605	Fürstenfeld	2	6,936
606	Gleisdorf	5	7,107
609	Hartberg	5	5,884
610	Judenburg	3	8,497
611	Murau	2	9,528
613	Knittelfeld	3	5,306
614	Leibnitz	5	9,092
615	Mureck	2	7,781
616	Leoben	3	8,178
618	Liezen	3	7,040
621	Mürzzuschlag	3	8,871
622	Voitsberg	3	5,861
623	Weiz	5	5,226
702	Innsbruck	6	0,000
705	Kufstein	4	9,306
707	Lienz	3	6,764
802	Bregenz	4	7,254
804	Dornbirn	4	9,633
805	Feldkirch	4	7,649

Quelle: Berechnungen Statistik Austria mit Daten der Statistik Austria und des AMS Österreich

An der Distanz zum Clusterzentrum (quadrierte euklidische Distanz) kann erkannt werden, welche RGS eher am Rand ihres Clusters liegen. In der folgenden Tabelle scheinen die Ergebnisse der Analyse der Randlagen auf. Es wurden alle RGS analysiert, deren Distanz zum Clusterzentrum über 10 liegt.

Die drei „Nächsten-Nachbarn“ der RGS mit großer Distanz wurden überprüft, um deren Randlage einschätzen zu können. Dabei wurden aus der Unähnlichkeitsmatrix die Paare gefiltert, die die niedrigste Distanz zueinander aufweisen.

Tabelle 10: Nächste Nachbarn der RGS mit Randlage

			nächste Nachbarn			quadrierte euklidische Distanz		
RGS	**Cluster**	**Distanz**	**1. Nachbar**	**2. Nachbar**	**3. Nachbar**	**1. Nachbar**	**2. Nachbar**	**3. Nachbar**
201-Feldkirchen	3	10,1	207-Völkermarkt (Cluster 3)	208-Wolfsberg (Cluster 2)	107-Jennersdorf (Cluster 2)	37,9	121,4	123,8
203-Klagenfurt	4	11,4	316-Lilienfeld (Cluster 2)	333-Waidhofen/Ybbs (Cluster 5)	622-Voitsberg (Cluster 3)	22,2	67,8	78,8
204-Spittal/ Drau	3	13,2	414-Schärding (Cluster 5)	611-Murau (Cluster 2)	205-St. Veit/Glan (Cluster 3)	104,4	154,4	158,3
206-Villach	4	11,7	419-Wels (Cluster 4)	621-Mürzzuschlag (Cluster 3)	205-St. Veit/Glan (Cluster 3)	20,7	34,6	63,5
207-Völkermarkt	3	12,2	201-Feldkirchen (Cluster 3)	208-Wolfsberg (Cluster 2)	205-St. Veit/Glan (Cluster 3)	37,9	134,3	227,5
304-Baden neu	4	11,3	315-Krems (Cluster 5)	331-Tulln (Cluster 1)	329-Schwechat (Cluster 1)	88,7	90,1	91,5
321-Mödling	1	12,8	329-Schwechat (Cluster 1)	331-Tulln (Cluster 1)	304-Baden neu (Cluster 4)	154,4	174,2	224,7
331-Tulln	1	11,2	314-Korneuburg (Cluster 1)	306-Bruck/Leitha (Cluster 1)	304-Baden neu (Cluster 4)	19,9	80,6	90,1

Quelle: Berechnungen Statistik Austria mit Daten der Statistik Austria und des AMS Österreich

An den Distanzen kann für einige RGS festgestellt werden, dass ihre nächsten Nachbarn bereits sehr unähnlich zu ihnen selbst sind. Alle Nachbarn der 8 betrachteten RGS weisen eine minimale Distanz von knapp 20 auf und sind somit ihren Nachbarn bei weitem unähnlicher, als ihrem Clusterzentrum. Es kann somit angenommen werden, dass die Clusterzuteilung dieser Randfälle dennoch korrekt ist.

4.2 Typisierung der Cluster

Die Typisierung der Cluster hängt stark von deren Merkmalsausprägungen ab. Dafür werden die Clustermittelpunkte verglichen, um die Unterschiede zwischen den Clustern erkennen zu können.

Nach der Clusterbildung im k-means Verfahren zeigen sich folgende Clustermittelpunkte der 6 Cluster:

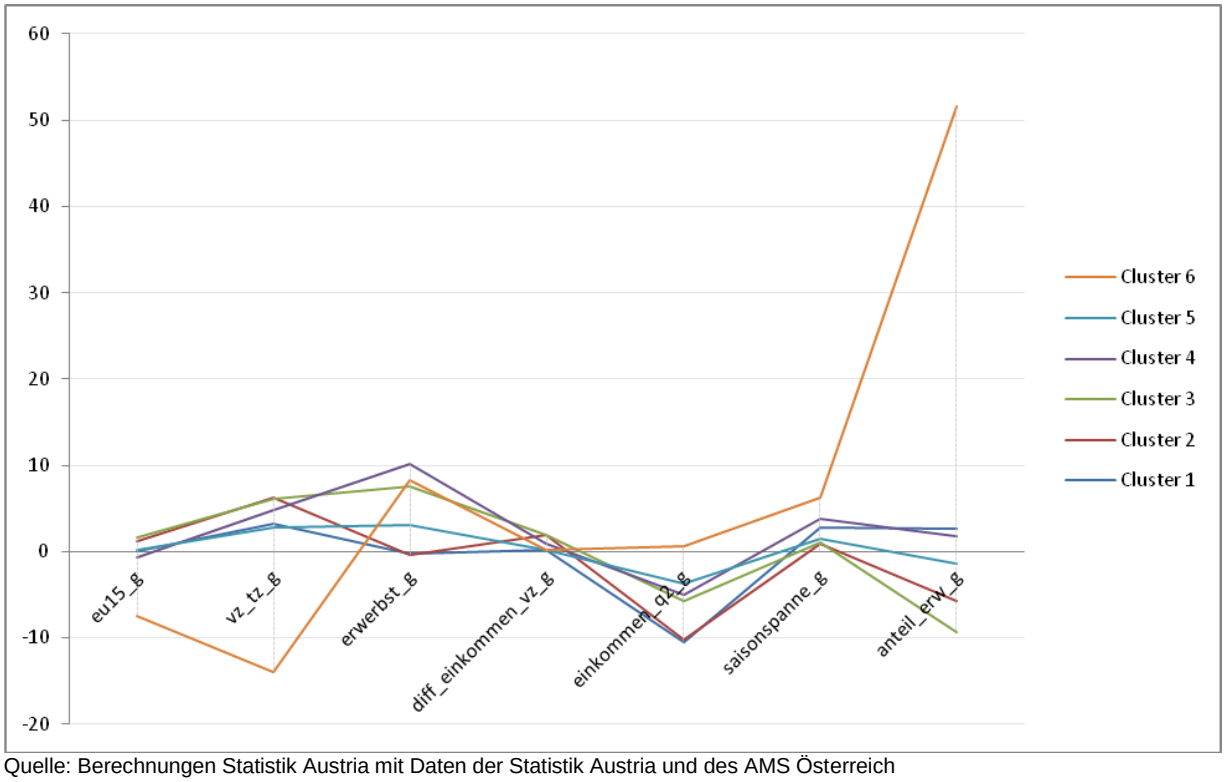
Tabelle 11: Standardisierte und gewichtete Clusterzentren als Ergebnis der k-means-Clusteranalyse

Variable	Cluster					
	1	2	3	4	5	6
eu15_g	,11	1,15	,33	-2,21	,79	-7,58
vz_tz_g	3,29	3,02	-,18	-1,28	-2,06	-16,77
erwerbst_g	-,24	-,22	7,95	2,64	-6,99	5,10
diff_einkommen_vz_g	,12	1,81	-,06	-,98	-,69	-,05
einkommen_q2_g	-10,57	,35	4,46	,68	1,43	4,29
saisonspanne_g	2,75	-1,87	,21	2,70	-2,24	4,72
anteil_erw_g	2,63	-8,37	-3,60	11,10	-3,11	52,94

Quelle: Berechnungen Statistik Austria mit Daten der Statistik Austria und des AMS Österreich

In der folgenden Grafik werden die Unterschiede in den 6 Clusterzentren aufbereitet. Dafür wurden die z-standardisierten und gewichteten Mittelwerte der Cluster verwendet.

Grafik 5: Graphisch dargestellte Unterschiede in den Clusterzentren



Quelle: Berechnungen Statistik Austria mit Daten der Statistik Austria und des AMS Österreich

Aus der Grafik können die gröbsten Unterschiede zwischen den einzelnen Clusterzentren gelesen werden. Anhand der einzelnen Merkmale kann erkannt werden, welche Cluster sich ähnlich verhalten und welche sehr stark voneinander unterschiedlich sind.
Je geringer das Gewicht der Merkmale ist, umso weniger eindeutige Unterschiede zeigen sich in den Clustern. Je höher das Gewicht der Merkmale, umso deutlicher sind die Unterschiede zu erkennen.

Dies zeigt am eindrucksvollsten das Merkmal ANTEIL_ERW, bei dem sich die Cluster voneinander stark unterscheiden. Dieses Merkmal hat mit Abstand das höchste Gewicht, da dies von der AMS-Projektleitung festgelegt wurde.
Sehr ähnlich scheinen sich alle Cluster beim Differenzeinkommen zu sein, während bei der Erwerbstätigenquote die Unterschiede sichtbar sind.
Cluster 6 – mit Innsbruck – verhält sich in nahezu allen Clustermerkmalen konträr zu den übrigen Beobachtungen. Auch in diesem Vergleich kann wieder erkannt werden, dass die RGS Innsbruck nicht sinnvoll in einem Cluster mit anderen RGS zusammengeführt werden kann.

Für eine sinnvolle inhaltliche Analyse der Cluster, wurden die Mittelwerte aller 6 Cluster mit den Originaldaten vor der Z-Transformation und Gewichtung in eine Übersicht gestellt.

Tabelle 12: Clusterzentren ohne Standardisierung und Gewichtung

	EU15_ Mean	VZ_TZ_M ean	ERWERBST _Mean	DIFF_EINKOMMEN _VZ_Mean	EINKOMMEN_Q2_ Mean	SAISONSPANNE _Mean	ANTEIL_ERW _Mean
1	1,60	**29,89**	71,73	27,54	**26239,91**	62,52	0,97
2	**1,02**	30,06	71,72	**23,51**	22894,76	42,65	**0,42**
3	1,48	32,09	**68,17**	27,97	**21638,84**	51,61	0,66
4	2,89	32,79	70,48	**30,17**	22794,50	62,28	1,39
5	1,22	33,29	**74,66**	29,46	22564,95	**41,07**	0,68
6	**5,87**	**42,62**	69,41	27,95	21690,95	**70,98**	**3,47**

Quelle: Berechnungen Statistik Austria mit Daten der Statistik Austria und des AMS Österreich

Aus der Tabelle können die unterschiedlichen Ausprägungen der Merkmale aller Cluster gelesen werden. Die Unterschiede in den Ausprägungen definieren die Cluster und sind somit die Grundlage für die Typisierung der Cluster. Die Werte, die sich eindeutig vom Durchschnitt abheben wurden in der Tabelle fett markiert. Dabei kann es sich um besonders hohe oder besonders niedrige Werte handeln.

Folgende Typisierung kann aufgrund der Unterschiede in den Clustermittelwerten aller Merkmale vorgenommen werden:

Cluster 1: „Einkommensstarker Cluster “
Der erste Cluster weist das höchste Medianeinkommen aller 6 Cluster auf, aber ebenso den niedrigsten Wert im Merkmal VZ_TZ. Somit ist das hohe Einkommen durch die niedrige Anzahl an Teilzeitbeschäftigungen erklärbar.
Die Erwerbstätigenquote und die Saisonspanne ist die zweithöchste aller Cluster, während alle anderen Clustermerkmale eher im Mittelfeld angesiedelt sind.

Cluster 2: „Bevölkerungsschwacher Cluster mit minimalen Einkommensdifferenzen zwischen Männern und Frauen“
Cluster zwei weist gleich drei Spitzenreiter in den Clustermittelwerten auf: Die geringste Bevölkerungsdichte (ANTEIL_ERW), den geringsten Anteil an Personen mit einem nicht-österreichischen Geburtsland der EU-15 und den geringsten Unterschied beim Einkommen zwischen Männern und Frauen. Die Einkommensdifferenz wird aber auch durch den sehr geringen Anteil an Teilzeitbeschäftigungen erklärt. Alle anderen Merkmale liegen im Durchschnitt.

Cluster 3: „Einkommens- und Erwerbsquotenschwacher Cluster“
Der dritte Cluster zeichnet sich vor allem durch sein sehr geringes Medianeinkommen aus in Kombination mit einer sehr geringen Erwerbstätigenquote. Alle anderen Merkmale liegen eher im Mittefeld aller betrachteten Cluster.

Cluster 4: „Bevölkerungsstarker Cluster mit hohen Einkommensdifferenzen zwischen Männern und Frauen“

Im Vergleich zu allen anderen Clustern, weist Cluster 4 einen starken Unterschied zwischen den Einkommen von Männern und Frauen auf, wobei die Teilzeitquote eher im Mittefeld angesiedelt ist.
Der Anteil an Personen mit einem nicht-österreichischen Geburtsland der EU-15 und die Bevölkerungsdichte liegt an zweithöchster Stelle, während sich die anderen Merkmale unauffällig verhalten.

Cluster 5: „Erwerbsstarker und saisonal stabiler Cluster“

Cluster 5 nimmt in zwei Clustermerkmalen eine spezielle Stellung ein: Er weist die niedrigste Saisonspanne auf und die höchste Erwerbstätigenquote.
Die Bevölkerungsdichte ist eher gering, weist also eher auf einen ländlichen Raum hin, wobei auch der Anteil an Personen mit einem nicht-österreichischen Geburtsland der EU-15 gering ist.

Cluster 6: „Innsbruck“

Da der sechste Cluster nur aus einer RGS besteht, nämlich Innsbruck, sind fast alle Clustermerkmale stark unterschiedlich zu den anderen Clustern.
So weist Innsbruck eine hohe Bevölkerungsdichte auf, ein sehr geringes Medianeinkommen, eine geringe Erwerbstätigenquote, einen hohen Anteil an Teilzeitbeschäftigungen und einen sehr hohen Anteil an Personen mit einem nicht-österreichischen Geburtsland der EU-15 – im Vergleich mit den anderen 5 Clustern.

Die Cluster 7 und 8 wurden vorab vom AMS-Projektteam definiert und tragen bereits einen sprechenden Namen:

Cluster 7: „Großstadtcluster“

Cluster 8: „Tourismuscluster“

Die Merkmale der gewählten Cluster 7 und 8 werden jedoch nicht analysiert und auf deren Clustertypisierung geprüft.

4.3 Regionale Übersicht

In der folgenden Tabelle werden alle RGS zu deren Cluster gelistet. Die Farbkennzeichnung der Cluster findet sich auf der anschließenden Österreichkarte wieder.

Tabelle 13: Clustereinteilung RGS

Cluster 1			
101	Eisenstadt	319	Mistelbach
301	Amstetten	321	Mödling
306	Bruck/Leitha	323	Neunkirchen
308	Gänserndorf	329	Schwechat
314	Korneuburg	331	Tulln
Cluster 2			
102	Mattersburg	313	Horn
103	Neusiedl am See	316	Lilienfeld
104	Oberpullendorf	332	Waidhofen/Thaya
106	Stegersbach	603	Deutschlandsberg
107	Jennersdorf	605	Fürstenfeld
208	Wolfsberg	611	Murau
311	Gmünd	615	Mureck
312	Hollabrunn		
Cluster 3			
105	Oberwart	613	Knittelfeld
201	Feldkirchen	616	Leoben
204	Spittal/Drau	618	Liezen
205	St. Veit/Glan	621	Mürzzuschlag
207	Völkermarkt	622	Voitsberg
601	Bruck/Mur	707	Lienz
610	Judenburg		
Cluster 4			
203	Klagenfurt	415	Steyr
206	Villach	418	Vöcklabruck
304	Baden neu	419	Wels
326	St. Pölten	705	Kufstein
334	Wr. Neustadt	802	Bregenz
401	Braunau	804	Dornbirn
404	Gmunden	805	Feldkirch
Cluster 5			
315	Krems	412	Ried im Innkreis
317	Melk	413	Rohrbach
328	Scheibbs	414	Schärding
333	Waidhofen/Ybbs	421	Traun
335	Zwettl	503	Hallein
402	Eferding	604	Feldbach
403	Freistadt	606	Gleisdorf
406	Grieskirchen	609	Hartberg
407	Kirchdorf/Krems	614	Leibnitz
411	Perg	623	Weiz
Cluster 6			
702	Innsbruck		

Quelle: Berechnungen Statistik Austria mit Daten der Statistik Austria und des AMS Österreich

Auf der Österreichkarte, eingeteilt nach Arbeitsmarktbezirken (RGS), zeigen sich die Cluster farblich getrennt. Anhand dieser Darstellung kann auch ein regionaler Bezug der Cluster erkannt werden, wie z.B. Gebiete mit starkem Tourismus oder Randgebiete um größere Städte.

Die beiden vordefinierten Cluster finden sich ebenfalls in der Karte wieder, auch wenn über dessen Clusterzuordnung nicht mit Hilfe des vorliegenden Modells entschieden wurde.

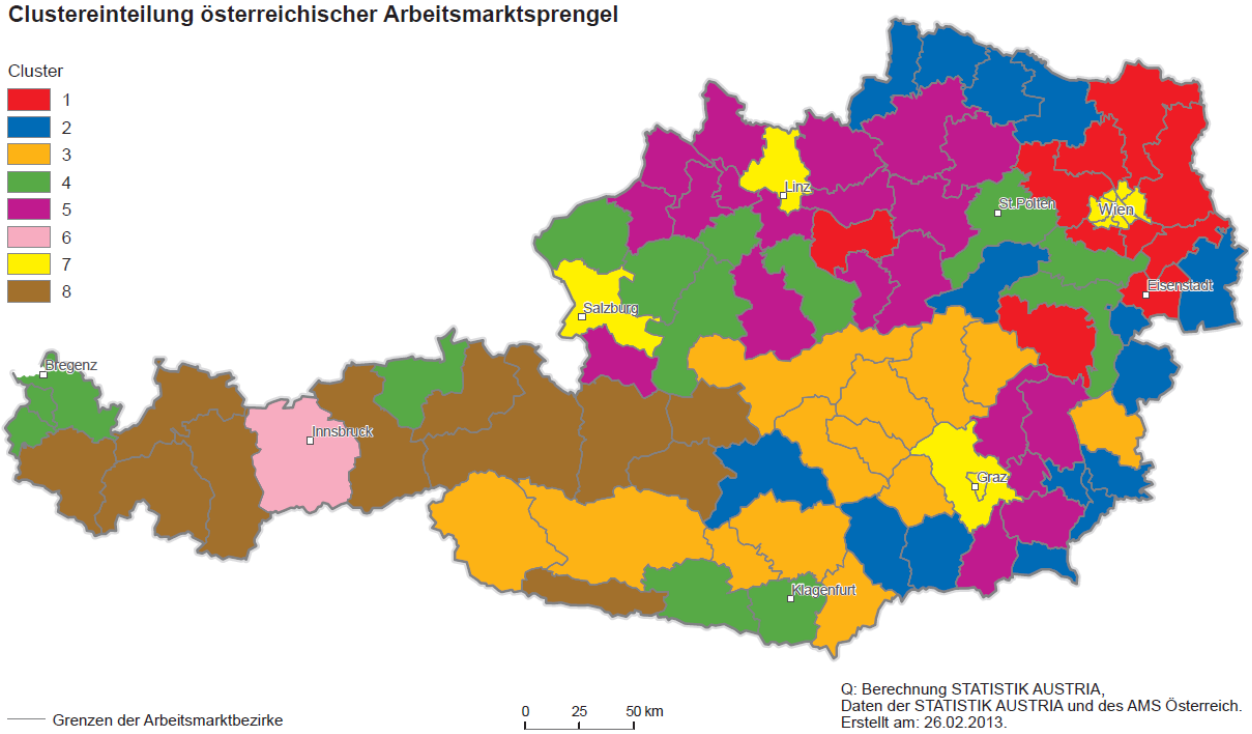
Tabelle 14: Clustereinteilung RGS durch das AMS

Cluster 7			
409	Linz neu	963	Wien Währinger Gürtel Gesundheit/Soziales/öffentl. Dienst
504	Salzburg	964	Wien Geiselbergstr. Verkehr/Agrar/Nahrung
630	Graz West	965	Wien Schönbrunner Str. Personalbereitst./ REHA
631	Graz Ost	966	Wien Hietzinger Kai Fremdenverkehr
958	Wien Johnstraße	967	Wien Huttengasse Bauwesen
959	Wien Hauffgasse	968	Wien Schloßhofer Str. Großhandel/Werbung/Druck/Foto
960	Wien Esteplatz Bank/Versich./EDV	969	Wien Prandaugasse Persönliche Dienste
961	Wien Dresdner Str. Metall/Chemie/Energie/Wasser	970	Wien Jugendliche
962	Wien Redergasse Einzelhandel		
Cluster 8			
202	Hermagor	704	Kitzbühel
501	Bischofshofen	706	Landeck
505	Tamsweg	708	Reutte
506	Zell am See	709	Schwaz
620	Gröbming	801	Bludenz
701	Imst		

Quelle: Informationen des AMS Österreich

In der folgenden Grafik wird regional die Clustereinteilung der RGS veranschaulicht. Die Grenzen der Arbeitsmarktbezirke gelten dabei als regionale Einheiten.

Grafik 6: Österreichkarte mit Clustereinteilung



Quelle: Berechnungen Statistik Austria mit Daten der Statistik Austria und des AMS Österreich

Tabellenverzeichnis

Tabelle 1: Schrittweise Regression..... 9
Tabelle 2: Kollinearitätstest.....12
Tabelle 3: Deskriptive Statistik der Clustermerkmale.....13
Tabelle 4: Zuordnungsübersicht nach der Ward-Clusteranalyse.....16
Tabelle 5: ANOVA-Test nach der k-means-Clusteranalyse 17
Tabelle 6: Clusterzentrendistanz nach der k-means-Clusteranalyse.....18
Tabelle 7: Anzahl Objekte nach der k-means-Clusteranalyse.....18
Tabelle 8: Varianztest der k-means-Clusteranalyse.....19
Tabelle 9: Clusterzugehörigkeit RGS und Distanz zum Clusterzentrum 20
Tabelle 10: Nächste Nachbarn der RGS mit Randlage.....21
Tabelle 11: Standardisierte und gewichtete Clusterzentren als Ergebnis der k-means-Clusteranalyse.....22
Tabelle 12: Clusterzentren ohne Standardisierung und Gewichtung.....23
Tabelle 13: Clustereinteilung RGS.....25
Tabelle 14: Clustereinteilung RGS durch das AMS 26
Tabelle 15: Merkmale der RGS aus Cluster 1 31
Tabelle 16: Merkmale der RGS aus Cluster 2 31
Tabelle 17: Merkmale der RGS aus Cluster 3 31
Tabelle 18: Merkmale der RGS aus Cluster 4 31
Tabelle 19: Merkmale der RGS aus Cluster 5 32
Tabelle 20: Merkmale der RGS aus Cluster 6 32

Abbildungsverzeichnis

Grafik 1: Beobachtete ALQ nach vorhergesagter ALQ 10
Grafik 2: Verteilung der unabhängigen Variablen 11
Grafik 3: Verteilung der Clustermerkmale und ALQ 14
Grafik 4: Heterogenitätsentwicklung nach der Ward-Clusteranalyse.....16
Grafik 5: Graphisch dargestellte Unterschiede in den Clusterzentren 22
Grafik 6: Österreichkarte mit Clustereinteilung 26
Grafik 7: Dendogramm nach der Ward-Clusteranalyse.....30
Grafik 8: Österreichkarte mit Clustereinteilung (schwarz/weiß).....32

Literaturverzeichnis

Backhaus, K.; Erichson, B.; Plinke, W.; Weiber, R.: Multivariate Analysemethoden. Eine anwenderorientierte Einführung. 2006; Springer; Berlin Heidelberg

Bender, S.; Hirschenauer, F.: Regionale Unterschiede in der Frauenerwerbstätigkeit – Eine Typisierung westdeutscher Arbeitsmarktregionen. Sonderdruck aus: Mitteilungen aus der Arbeitsmarkt- und Berufsforschung. 1993; W. Kohlhammer GmbH; Stuttgart

Blien, U.; Hirschenauer, F.: Vergleichstypen 2005. Neufassung der Regionaltypisierung für Vergleiche zwischen Agenturbezirken; 2005; IAB Forschungsbericht. Ergebnisse aus der Projektarbeit des Instituts für Arbeitsmarkt- und Berufsforschung; Bundesagentur für Arbeit.

Dauth, W.; Hirschenauer, F.; Rüb, F.: Vergleichstypen 2008. Neufassung der SBG – III – Typisierung; IAB-Forschungsbericht; Aktuelle Ergebnisse aus der Projektarbeit des Institus für Arbeitsmarkt- und Berufsforschung; August 2008; Bundesagentur für Arbeit.

Egger, M.; Lenz, C.: Wirkungsevaluation der öffentlichen Arbeitsvermittlung. Evaluationsbericht. Studie im Auftrag der Aufsichtskommission für den Ausgleichsfonds der Arbeitslosenversicherung. Oktober 2006; Staatssekretariat für Wirtschaft; Egger, Dreher & Partner AG; Bern

Egger, M.; Lenz, C.; Zürchner, B. A.: Die Vereinbarung zwischen Bund und Kantonen über den Vollzug des Arbeitslosenversicherungsgesetzes in der Schweiz. November 2001; Staatssekretariat für Wirtschaft; Bern

Krämer, W., Schoffer, O., Tschiersch, L.: Datenanalyse mit SAS. Statistische Verfahren und ihre grafischen Aspekte. 2005; Springer; Berlin Heidelberg

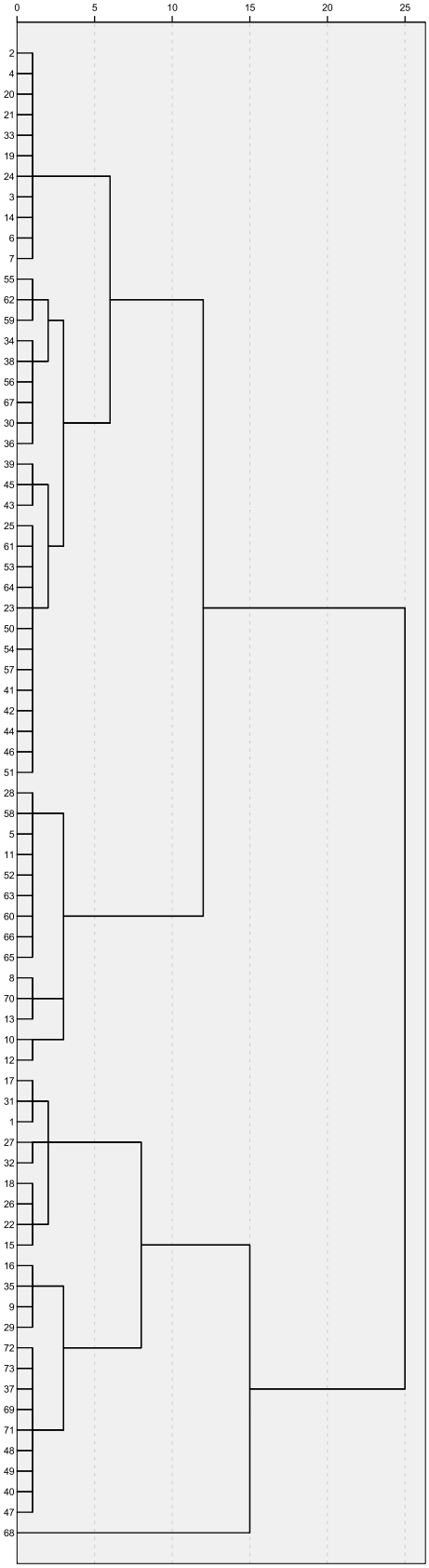
Rüb, F.; Werner, D.: Typisierung von SBG II – Trägern. IAB Forschungsbericht. Ergebnisse aus der Projektarbeit des Instituts für Arbeitsmarkt- und Berufsforschung. 2007; Bundesagentur für Arbeit.

Rüb, F.; Werner, D.; Kaufmann, K.; Wolf, K.; Blien, U.: Regionale Typisiering im SGB II – Bereich – Aktualisierung 2006. Fachliche Dokumentation (Stand 01.03.2008); 2008; Institut für Arbeitsmarkt- und Berufsforschung.

Templ, M.; Filzmoser, P.; Reinmann, C.: Cluster analysis applied to regional geochemical data. Problems and possibilities. Applied Geochemistry 23 (2008); S 2198 - 2213

Anhang

Grafik 7: Dendogramm nach der Ward-Clusteranalyse



Quelle: Berechnungen Statistik Austria mit Daten der Statistik Austria und des AMS Österreich

1. Merkmale aller RGS innerhalb eines Clusters

In den folgenden Tabellen sind alle RGS der 6 Cluster mit deren nicht-standardisierten und nicht-gewichteten Merkmalen angeführt.

Tabelle 15: Merkmale der RGS aus Cluster 1

RGS_NR	RGS_NAME	EU15	VZ_TZ	ERWERBST	DIFF_EINKOMMEN_VZ	EINKOMMEN_Q2	SAISONSPANNE	ANTEIL_ERW
101	Eisenstadt	1,51	27,92	71,02	25,36	26.089,93	49,20	0,66
301	Amstetten	0,87	30,49	74,63	30,16	24.341,52	56,00	1,11
306	Bruck/Leitha	1,44	28,02	71,06	28,23	26.672,32	63,24	0,54
308	Gänserndorf	1,25	28,77	70,42	26,36	25.479,86	63,71	1,13
314	Korneuburg	1,87	30,27	73,47	28,22	27.468,27	62,02	0,89
319	Mistelbach	1,07	29,89	72,13	24,44	25.912,05	54,78	1,00
321	Mödling	3,03	33,32	71,53	34,04	28.619,54	78,28	1,31
323	Neunkirchen	1,04	29,55	69,36	23,89	24.159,71	63,18	0,99
329	Schwechat	1,75	27,78	71,49	27,11	26.875,24	72,51	0,59
331	Tulln	2,20	32,91	72,16	27,59	26.780,71	62,25	1,46

Quelle: Berechnungen Statistik Austria mit Daten der Statistik Austria und des AMS Österreich

Tabelle 16: Merkmale der RGS aus Cluster 2

RGS_NR	RGS_NAME	EU15	VZ_TZ	ERWERBST	DIFF_EINKOMMEN_VZ	EINKOMMEN_Q2	SAISONSPANNE	ANTEIL_ERW
102	Mattersburg	0,80	29,26	71,03	27,31	24.516,74	54,73	0,46
103	Neusiedl am See	1,52	28,30	73,80	25,95	24.480,31	35,21	0,62
104	Oberpullendorf	0,66	28,74	70,70	25,84	24.425,00	48,26	0,44
106	Stegersbach	1,35	28,29	69,07	21,79	22.935,53	36,71	0,31
107	Jennersdorf	1,63	26,82	71,20	23,06	21.981,96	32,56	0,21
208	Wolfsberg	1,01	27,27	71,71	28,09	21.817,60	40,71	0,63
311	Gmünd	0,60	30,68	70,86	19,71	22.814,46	48,05	0,43
312	Hollabrunn	0,83	30,63	70,34	21,17	24.360,00	53,01	0,59
313	Horn	0,81	31,61	73,02	20,23	23.492,00	49,22	0,36
316	Lilienfeld	0,87	30,31	71,56	24,19	23.682,78	56,65	0,30
332	Waidhofen/Thaya	0,49	31,94	72,65	17,30	22.658,48	49,07	0,31
603	Deutschlandsberg	1,29	31,53	71,64	25,52	22.588,54	39,94	0,73
605	Fürstenfeld	1,20	31,60	72,85	26,79	21.489,22	34,43	0,27
611	Murau	1,15	33,49	71,90	25,26	20.619,30	34,64	0,35
615	Mureck	1,13	30,43	73,49	20,42	21.559,54	26,53	0,27

Quelle: Berechnungen Statistik Austria mit Daten der Statistik Austria und des AMS Österreich

Tabelle 17: Merkmale der RGS aus Cluster 3

RGS_NR	RGS_NAME	EU15	VZ_TZ	ERWERBST	DIFF_EINKOMMEN_VZ	EINKOMMEN_Q2	SAISONSPANNE	ANTEIL_ERW
105	Oberwart	0,96	31,13	68,45	24,60	22.441,45	49,09	0,64
201	Feldkirchen	2,55	29,80	68,91	25,72	19.852,66	36,37	0,36
204	Spittal/Drau	2,52	30,79	64,69	32,32	19.164,63	41,07	0,92
205	St. Veit/Glan	1,50	30,96	68,99	25,78	21.970,34	45,68	0,66
207	Völkermarkt	1,67	26,79	68,84	24,05	19.331,90	40,32	0,50
601	Bruck/Mur	1,10	34,55	67,21	29,44	22.932,80	64,88	0,72
610	Judenburg	1,02	30,92	68,59	29,29	22.676,80	55,33	1,02
613	Knittelfeld	0,79	33,81	68,40	29,72	22.900,55	56,56	0,68
616	Leoben	1,16	34,13	65,88	31,72	22.912,22	63,21	0,71
618	Liezen	1,94	32,36	70,89	25,66	22.529,64	51,22	0,67
621	Mürzzuschlag	0,74	35,55	67,75	30,71	23.027,82	63,85	0,46
622	Voitsberg	1,01	35,30	69,23	25,91	21.817,82	54,05	0,62
707	Lienz	2,28	31,09	68,38	28,64	19.746,36	49,31	0,58

Quelle: Berechnungen Statistik Austria mit Daten der Statistik Austria und des AMS Österreich

Tabelle 18: Merkmale der RGS aus Cluster 4

RGS_NR	RGS_NAME	EU15	VZ_TZ	ERWERBST	DIFF_EINKOMMEN_VZ	EINKOMMEN_Q2	SAISONSPANNE	ANTEIL_ERW
203	Klagenfurt	2,61	31,48	67,86	22,35	23.274,08	56,80	1,82
206	Villach	3,29	31,90	66,06	29,05	21.153,92	58,80	1,46
304	Baden neu	2,07	29,20	69,76	29,56	25.231,19	71,51	1,63
326	St. Pölten	1,14	33,46	70,82	25,26	24.193,47	67,13	1,74
334	Wr. Neustadt	1,36	28,22	69,13	26,61	24.292,86	70,00	1,37
401	Braunau	3,63	33,56	73,33	28,83	21.821,25	51,36	1,17
404	Gmunden	2,14	36,37	72,00	30,53	22.876,00	51,05	1,17
415	Steyr	1,23	33,25	70,65	31,44	23.717,68	59,18	1,14
418	Vöcklabruck	1,83	34,27	72,79	32,50	22.609,40	51,76	1,54
419	Wels	1,66	31,20	71,58	31,19	22.908,49	59,30	1,51
705	Kufstein	5,93	34,19	71,30	30,67	21.112,32	52,84	1,21
802	Bregenz	5,85	32,83	70,44	34,92	21.935,77	73,45	1,51
804	Dornbirn	3,54	33,58	69,82	34,46	22.043,98	75,31	0,97
805	Feldkirch	4,18	35,57	71,13	35,00	21.952,54	73,42	1,21

Quelle: Berechnungen Statistik Austria mit Daten der Statistik Austria und des AMS Österreich

Tabelle 19: Merkmale der RGS aus Cluster 5

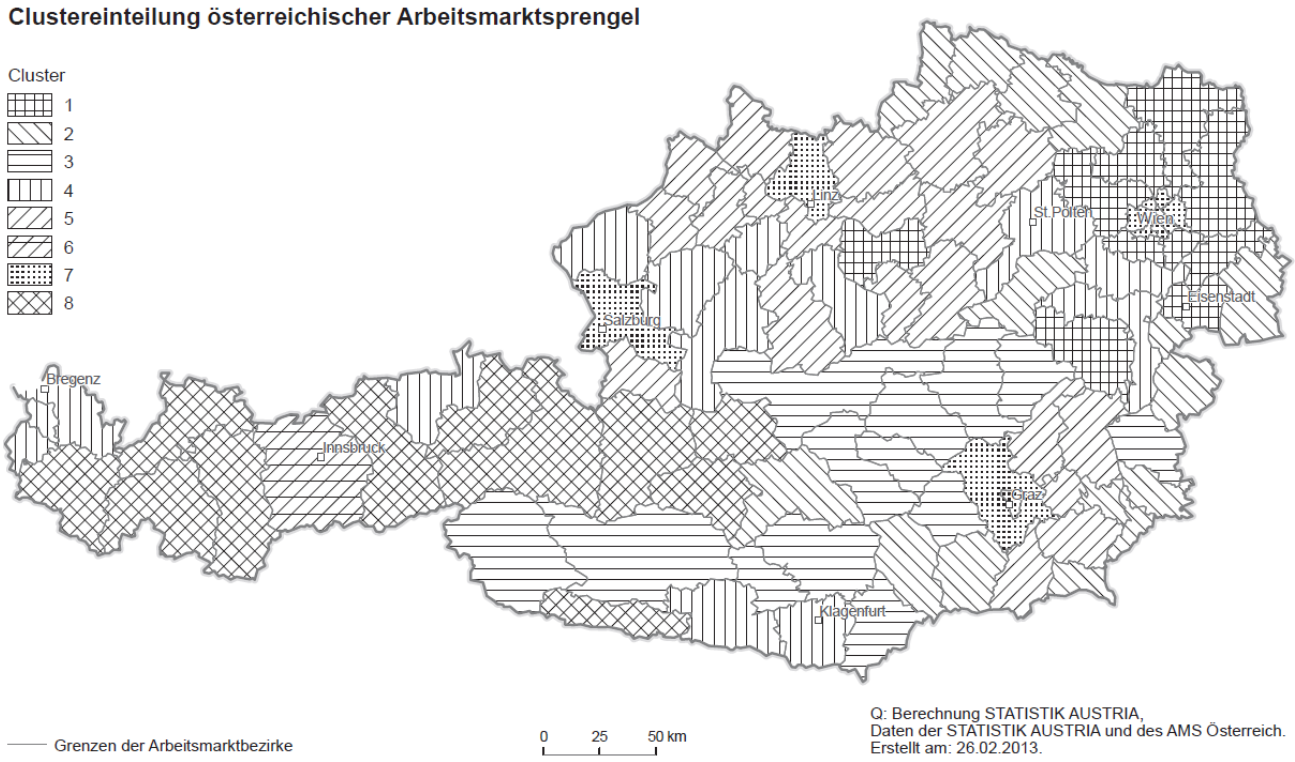
RGS_NR	RGS_NAME	EU15	VZ_TZ	ERWERBST	DIFF_EINKOMMEN_VZ	EINKOMMEN_Q2	SAISONSPANNE	ANTEIL_ERW
315	Krems	1,19	33,02	71,98	25,20	23.508,43	50,08	0,93
317	Melk	0,92	31,29	74,05	25,95	23.111,62	38,09	0,90
328	Scheibbs	0,73	31,25	76,08	25,37	22.822,22	32,74	0,48
333	Waidhofen/Ybbs	0,97	32,75	74,82	27,31	22.891,63	45,59	0,35
335	Zwettl	0,48	31,63	76,91	21,39	21.821,02	26,19	0,51
402	Eferding	0,86	33,49	75,18	32,28	22.849,59	43,28	0,38
403	Freistadt	0,60	37,62	76,45	33,22	23.339,40	44,68	0,78
406	Grieskirchen	1,20	33,11	74,75	31,96	23.321,14	43,16	0,74
407	Kirchdorf/Krems	1,46	32,04	74,47	32,11	23.142,52	52,96	0,65
411	Perg	0,71	35,67	74,98	33,36	23.866,25	48,78	0,78
412	Ried im Innkreis	1,91	32,18	74,54	29,71	22.363,96	43,98	0,70
413	Rohrbach	1,30	36,80	76,57	31,91	22.981,29	33,35	0,68
414	Schärding	3,36	33,59	72,97	30,07	21.799,57	31,31	0,67
421	Traun	1,44	34,09	71,71	31,88	23.307,01	57,12	0,79
503	Hallein	3,21	37,48	73,25	30,80	22.560,78	55,93	0,68
604	Feldbach	0,71	31,38	75,26	27,64	20.928,61	31,24	0,80
606	Gleisdorf	1,02	31,78	76,25	29,48	21.681,50	32,43	0,46
609	Hartberg	0,54	32,24	74,45	29,61	21.144,00	34,33	0,80
614	Leibnitz	1,12	31,19	72,04	29,09	21.665,42	41,72	0,94
623	Weiz	0,66	33,11	76,52	30,91	22.193,09	34,40	0,59

Quelle: Berechnungen Statistik Austria mit Daten der Statistik Austria und des AMS Österreich

Tabelle 20: Merkmale der RGS aus Cluster 6

RGS_NR	RGS_NAME	EU15	VZ_TZ	ERWERBST	DIFF_EINKOMMEN_VZ	EINKOMMEN_Q2	SAISONSPANNE	ANTEIL_ERW
702	Innsbruck	5,87	42,62	69,41	27,95	21.690,95	70,98	3,47

Quelle: Berechnungen Statistik Austria mit Daten der Statistik Austria und des AMS Österreich

2. Österreichkarte der Clustereinteilung mit Schraffur

Grafik 8: Österreichkarte mit Clustereinteilung (schwarz/weiß)