

Quadrat = 12002,3; df = 11769; Sig. = .065). Diese Feststellung ist die Voraussetzung für die verzerrungssarme Imputation fehlender Werte (Schafer/Graham 2002: 154). Darüber hinaus folgt die Hot-Deck Imputationen keinem parametrischen Modell und ist robust gegenüber Verletzungen der Normalverteilungsannahme. Auch ist es nicht auf metrische Daten<sup>145</sup> angewiesen. Beide Vorteile ermöglichen seine Verwendung im vorliegenden Fall – was zu zeigen sein wird. Schließlich können mit einem derart ergänzten Datensatz unmittelbar alle beliebigen Analysen vorgenommen werden, er ist nicht durch die begrenzte Verfügbarkeit von Analysetools beschränkt, wie dies bei komplexen Formen multipler Imputation der Fall ist. Ob seiner Einfachheit unterliegt das Hot Deck Verfahren auch deutlichen Limitationen (vgl. Myers 2011: 302).<sup>146</sup>

## 6.2 Analyseschritte und Methodik

Vor jeder Analyse wurden die Erfüllung der jeweils grundlegenden Annahmen hinsichtlich der Datenstruktur (Normalverteilungsannahme, Linearitätsannahme, Multikollinearität, Homoskedastizität und Autokorrelation) überprüft. Ergebnisse und ggf. vorliegende Einschränkungen werden im Kontext der Analysen berichtet. Mittels des statistischen Vergleichs der vier Modelle soll untersucht werden, ob eventuell zu beobachtende Unterschiede lediglich zufälliger oder systematischer Natur sind. Konkret werden hierzu die folgenden Analysen<sup>147</sup> durchgeführt:

scher Sicht am wünschenswertesten. Sie beinhaltet, dass der Prozess, welcher das Fehlen der Daten verursachte, in keinem statistischen Zusammenhang mit den in der Untersuchung interessierenden Variablen steht (Collins et al. 2001: 333). Die zweite, als *missing at random* (MAR) bezeichnete Annahme besagt, dass zwar ein Zusammenhang zwischen dem Fehlen von Daten und den in der Untersuchung interessierenden Variablen besteht, dieser jedoch über weitere in die Analyse einbezogene Kontrollvariablen erklärt werden kann (Acock 2005: 1014f.). Im Dritten, als *missing not at random* (MNAR) bezeichneten Fall liegt ein systematischer Ausfallmechanismus vor, der jedoch auch auf unbeobachteten Daten beruht. Dieses Ausfallmuster ist nicht durch Kontrollvariablen kontrollierbar und darf nicht ignoriert werden, die Ausfälle werden deshalb auch als *non-ignorable missing values* (NI) bezeichnet (Acock 2005: 1013, 1015).

- 145 Der vorliegende Datensatz umfasst mit lediglich zwei Ausnahmen bei den Kontrollvariablen ausschließlich latente kategoriale Daten und bietet damit keinen Ansatzpunkt für einen *Expectation Maximization* Ansatz.
- 146 So können einmalige Fälle, für die kein äquivalent im Datensatz vorliegt, nicht passgenau ersetzt werden, was insbesondere bei kleineren Datensätzen ein Problem darstellen kann. Auch führen Vervollständigungen mit einem Hot Deck Verfahren zu Verzerrungen von Korrelationen und anderen Beziehungsmaßen (vgl. Schafer/Graham 2002: 159).
- 147 Die statistischen Analysen wurden mit den Programmen *IBM SPSS Statistics 20* sowie *STATA 11* durchgeführt.

Auf Ebene der einzelnen Indikatoren:<sup>148</sup>

- Darstellung der Fallzahl, Errechnung von Mittelwert, Standardabweichung und Standardfehler für jedes der vier Organisationsmodelle,
- Analyse, ob zu beobachtende Unterschiede zwischen den vier Organisationmodellen zufälliger oder systematischer Natur sind (Ranggruppenanalyse),
- Analyse, welche der vier Organisationsmodelle sich signifikant voneinander unterscheiden (ANOVA-basierte post-hoc Tests),
- Analyse der inneren Konsistenz der additiven Indizes, auf deren Basis die theoretischen Konstrukte für die nachfolgenden komplexeren Untersuchungen gebildet wurden (Cronbach's Alpha).

Auf Ebene der theoretischen Konstrukte:

- Darstellung der Fallzahl, Errechnung von Mittelwert, Standardabweichung und Standardfehler für jedes der vier Organisationsmodelle,
- Analyse, ob zu beobachtende Unterschiede zwischen den vier Organisationmodellen zufälliger oder systematischer Natur sind (Ranggruppenanalyse),
- Analyse, welche der vier Organisationsmodelle sich signifikant voneinander unterscheiden (ANOVA-basierte post-hoc Tests).

Die Zielsetzung des ersten und dritten Analyseschrittes ist also, herauszufinden, ob sich die theoretisch postulierten Unterschiede in der Performanz realltypischer Organisationsmodelle in der Wahrnehmung ihrer Mitarbeiter wiederfinden lassen. Konkret bedeutet dies, dass untersucht werden soll, ob die Streuungen zwischen den Organisationsmodellen signifikant größer sind als deren Binnenvarianz. Da die untersuchten Gruppen als unabhängig voneinander betrachtet werden können und lediglich der Effekt der Gruppenzugehörigkeit analysiert wird, bietet sich methodisch die Verwendung einer einseitigen Varianzanalyse für unabhängige Stichproben an. Erste Analysen der Datenstruktur zeigen, dass gewisse Voraussetzungen zur Verwendung einer einseitigen Varianzanalyse (*ANOVA*) nicht erfüllt sein könnten. So erweisen sich erstens die Variablen trotz geringer Streuungen bei der visuellen Sichtung der *Q-Q-Normalverteilungsplots* bei der Analyse mit *Kolmogorov-Smirnov-* und *Shapiro-Wilk-Tests* durchgängig als nicht normalverteilt. Zweitens variieren, wie in Tabelle 8 gezeigt, die Größen der Stichproben zwischen den einzelnen Gruppen deutlich, so dass ein gegenüber dieser Gegebenheit insensibler Test gewählt werden muss. Drittens zeigt der *Levene's Test* bei manchen der Variablen heterogene Varianzen (Heteroskedastizi-

148 Die strukturellen Faktoren werden hierbei nicht eingeschlossen.

tät)<sup>149</sup> innerhalb der Gruppen, welches die Nutzungsmöglichkeit einer ANOVA weiter einschränkt. Angesichts der wenig differenzierten Kategorisierung der vier Behördentypen ist das Auftreten von Heteroskedastizität aufgrund einer defizitären Modellspezifikation wenig überraschend. Da die untersuchten Behördentypen jedoch keine (schlecht modellierten) Modelle, sondern in der wissenschaftlichen und praxisorientierten Debatte genutzte Realtypen sind, kommt eine Modifikation derselben nicht in Frage. Stattdessen wird nach technischen Wegen gesucht, die hier vorliegenden Beschränkungen in der Anwendung und Interpretation einfacher Varianzanalysen zu umgehen.

Eine Lösung für die genannten Hindernisse bei der Analyse der Unterschiede zwischen den vier Behördentypen bietet der *Kruskal-Wallis Test*. Diese einseitige, nicht-parametrische Ranggruppenanalyse vergleicht die Mediane (Rangpositionen) der Gruppen.<sup>150</sup> Sie erfordert keine Normalverteilung der Daten und gestattet gewisse Unterschiede in den Gruppengrößen. Wie eine ANOVA deckt der *Kruskal-Wallis-Test* lediglich auf, ob signifikante Unterschiede zwischen den Medianen vorliegen. Welche der Untergruppen sich konkret wie stark unterscheiden, wird nicht berichtet. Hierfür sind *post-hoc Analysen* notwendig. Diese sind jedoch in konventioneller Statistiksoftware nicht für *Kruskal-Wallis-Tests* implementiert, weshalb doch auf eine ANOVA zurückgegriffen werden musste. Um für die oben berichtete Heterogenität der Varianzen zu kontrollieren, wurde hierfür statt klassischer ANOVAs mit ihren F-Werten der nicht von einer Homogenität der Varianzen ausgehende *Welch Test* interpretiert. Der Vergleich zwischen *Kruskal-Wallis-* und *Welch-Test* lieferte bei der überwältigenden Mehrheit der Variablen identische Ergebnisse.<sup>151</sup> Es erscheint vertretbar, den ANOVA-basierten, aber für varianzheterogene Stichproben geeigneten *post-hoc Test* nach *Games-Howell*<sup>152</sup> vorsichtig zu interpretieren.

Die Operationalisierung der Mehrzahl der in Kapitel 5 entwickelten theoretischen Konstrukte erfolgt mittels additiver Indizes aus mehreren Items (vgl. zur Begründung Gliem/Gliem 2003). Zur Schätzung der internen Konsistenz und damit der Verwendbarkeit der Indizes wird der Trennschärfekoeffizient *Cronbach's Alpha* eingesetzt (vgl. Cronbach 1951). Diese Maßzahl kann Werte zwischen 0 und 1 annehmen, wobei ein höherer Wert eine größere interne Konsistenz signalisiert. Aufgrund der relativen (aber schwachen) Abhängigkeit der Er-

149 Ist das Ergebnis des Levene-Tests *nicht*  $\geq .20$ , so wird hier von heterogenen Varianzen ausgegangen.

150 Test der Nullhypothese identischer Mediane der verglichenen Gruppen.

151 Ein parallel durchgeföhrter *Brown Forsythe Test* lieferte ebenso durchgängig beinahe identische Resultate, seine Ergebnisse werden deshalb nicht berichtet.

152 Der *Games-Howell Test* beruht nicht auf der Annahme homogener Varianzen und identischer Größe der Gruppen. Ein parallel durchgeföhrter *Tukey HSD-Test* lieferte nur marginal abweichende Ergebnisse.

gebnisse von der Zahl der einbezogenen Items (vgl. Peterson 1994; Schendera 2007: 157) wird teilweise eine gestufte Bewertung vorgeschlagen: Werte ab  $\alpha \geq .5$  bei zwei Items, ab  $\alpha \geq .6$  bei drei Items und  $\alpha \geq .7$  ab vier Items werden noch als akzeptabel angesehen (Bagozzi 1980). Diese Werte werden bei allen hier gebildeten Items erreicht.

### 6.3 Performanz – Deskriptive Ergebnisse und Varianzanalyse

Die deskriptiven Ergebnisse und Ranggruppenanalyse werden hier kurorisch für die Einzelitems und die Konstruktebene dargestellt. Wo es sinnvoll erscheint werden die Verteilungen in den vier realltypischen Organisationsmodellen auch graphisch in Boxplots dargestellt.<sup>153</sup>

#### 6.3.1. Rechtsstaatlichkeit

Wie in Kapitel 5.4.1.1.1 erläutert, wird das Konstrukt *Rechtsstaatlichkeit* über zwei getrennte Ansätze erfasst – erstens *unbotmäßige Politisierung* des Verwaltungsvollzugs sowie zweitens *unangemessene Prioritätensetzung* im Verwaltungsvollzug.

##### 6.3.1.1 Unbotmäßige Politisierung

Die *unbotmäßige Politisierung des Verwaltungsvollzugs* wird über die *Zahl normbrechender Weisungen* gemessen, welche über die folgende Frage erfasst wurde:<sup>154</sup> „Wenn Sie an das vergangene Kalenderjahr denken – wie oft erhielten Sie in 2010 formale oder informale "Anweisungen" von Vorgesetzten aus Ihrer Behörde, die außerhalb des Normbereichs lagen?“ Die Antwortenden konnten zwischen Häufigkeitskategorien von 0 bis >50 wählen, deren Mittelwert als absolute Zahl in die statistische Auswertung einging. Die Auswertung zeigt, dass keines der Organisationsmodelle derartige zweifelhafte hierarchische Eingriffe verhindert. Allerdings zeigen *Kruskal-Wallis-* und *Welch Test* eine große, nicht-zufällige Varianz in ihrer Häufigkeit. So berichten in Mittelinstanzen (Regierungspräsidien, Bezirksregierungen und Landesverwaltungsämter) tätige Mitar-

153 Auf eine alle Ergebnisse der oben genannten Analysen beinhaltende Darstellung des Ranggruppenvergleichs sowie der Post-hoc Tests wird hier aus Gründen der Lesbarkeit verzichtet. Diese finden sich im elektronischen Anhang 2 auf der Homepage des Autors.

154 Für eine Druckversion des internetbasierten Fragebogens s. den elektronischen Anhang 1.