

# Theoretische Analyse und praktische Durchführung einer Dozentenbewertung

Manuel Koller

17. Oktober 2006

Bachelor-Arbeit in Betreuung bei Herrn Prof. Dr. Künsch  
Eidgenössische Technische Hochschule Zürich



# Inhaltsverzeichnis

<b>1</b>	<b>Praktischer Teil</b>	<b>7</b>
1.1	Planung und Programmierung der Stimmsoftware . . . . .	7
1.1.1	Datenschutz . . . . .	8
1.1.2	Ablauf einer Abstimmung . . . . .	8
1.1.3	Testlauf . . . . .	8
1.2	Die Abstimmung . . . . .	9
<b>2</b>	<b>Analyse</b>	<b>11</b>
2.1	Gesammelte Daten . . . . .	11
2.2	Das Modell . . . . .	11
2.2.1	Fixe und zufällige Effekte . . . . .	15
2.2.2	Ausgabe . . . . .	16
2.3	Kontrolle des Modells . . . . .	16
2.3.1	Bootstrap . . . . .	17
2.4	Resultate . . . . .	18
2.4.1	Anpassung der Daten . . . . .	18
2.4.2	Genauigkeit der Schätzwerte . . . . .	20
2.4.3	Einfluss der Vergangenheit . . . . .	22
2.4.4	Grosse und kleine Veranstaltungen . . . . .	22
<b>3</b>	<b>Bestimmen der Gewinner</b>	<b>25</b>
<b>A</b>	<b>Gewinner</b>	<b>27</b>
<b>B</b>	<b>Screenshots der Website</b>	<b>29</b>
<b>C</b>	<b>Output von PROC MIXED</b>	<b>33</b>
<b>D</b>	<b>Plots für alle Departemente</b>	<b>35</b>



# Einleitung

Anlässlich des Tages der Lehre wurde im November 2005 vom VSETH erstmals die goldene Eule verliehen. Die goldene Eule ist eine Auszeichnung für exzellente Lehre an der ETH; sie wird in jedem Departement vergeben.

Zur Bestimmung der Preisträger wurde unter den Studierenden eine Abstimmung durchgeführt. Die Studierenden konnten drei Dozenten<sup>1</sup> vorschlagen, die ihrer Meinung nach die goldene Eule verdient hatten. Der Dozent mit den meisten Stimmen erhielt den Preis.

Theoretische Überlegungen zeigen aber, dass die Verleihung nach Anzahl Nennungen gewichtige Nachteile hat. Dozierende, die nur wenige Studierende unterrichten, werden benachteiligt gegenüber Lehrpersonen mit sehr vielen Studierenden. Bei ihnen kann schon eine kleine Anzahl ihrer Studierenden genügen, um auf eine relevante Zahl von Vorschlägen zu kommen. Mittelmässige oder negative Bewertungen werden dagegen überhaupt nicht berücksichtigt. Dozierende andererseits, die nur Serviceveranstaltungen geben, haben einen potenziellen Nachteil gegenüber Dozierenden, die studiumseigene Fächer unterrichten.

Es gingen viele Kommentare von Abstimmenden ein, in denen angemerkt wurde, wie schwierig es gewesen sei, sich für drei Dozierende zu entscheiden, und dass es oft grosse Qualitätsunterschiede unter den Auserwählten gegeben habe.

Im Rahmen dieser Arbeit wird ein geeignetes Wahlverfahren entwickelt und umgesetzt. Der Einfluss der oben beschriebenen Probleme und das Identifizieren von noch unentdeckten Einflüssen wird untersucht, und es werden Lösungsstrategien aufgezeigt.

Das im VSETH mit den ersten beiden Verleihungen der goldenen Eule gewonnene Wissen soll in dieser Arbeit festgehalten und so künftig als Grundlage für Verleihungen des Preises dienen.

Bedanken möchte ich mich hier bei Herrn Professor Dr. Künsch, dessen Unterstützung die Untersuchung und die Bachelor-Arbeit erst möglich machte, und bei Herrn Dr. Roth vom Seminar für Statistik, der mir mit der Implementation in SAS sehr geholfen hat und für meine Fragen immer ein offenes Ohr hatte.

---

<sup>1</sup>Der Einfachheit halber werden hier und in der Folge die männlichen Begriffe gebraucht. Die weiblichen Bezeichnungen sind immer mitgemeint.



# Kapitel 1

## Praktischer Teil

### 1.1 Planung und Programmierung der Stimm-Software

Aufgrund der Unterschiede zur Abstimmung vom letzten Jahr konnte die damals verwendete Software nicht mehr benützt und musste komplett ersetzt werden. Die neue Auswertung baut auf einer anderen und wesentlich umfangreicheren Datenbasis auf. Damit stieg auch der Aufwand für die Abstimmenden. Damit diese nicht durch ein langwieriges und kompliziertes Verfahren abgeschreckt wurden, mussten die Daten zuerst aufwändig aufbereitet werden, bevor sie dem Benutzer präsentiert werden konnten.

Das Ziel war, eine Website zu programmieren, die die Abstimmung für den Abstimmenden möglichst einfach und kurz macht. Es sollte eine Liste der verfügbaren Dozenten übersichtlich präsentiert und mit einer möglichst klaren Navigation versehen werden. Es durfte keinen Datenverlust durch Navigieren geben. Auch sollte es eine Gedächtnisstütze geben, damit sich der Abstimmende nicht nur an die Extremfälle unter der Dozenten erinnert.

Damit ergab sich folgende Struktur der Website: Zuerst gibt es eine sehr kurz gehaltene Anleitung, in der der Ablauf und das Vorgehen erklärt und ein paar Tipps zur Handhabung gegeben werden. Dann wird eine Liste mit allen Dozenten präsentiert, von welchen eine beliebige Auswahl bewertet werden kann. Abschliessend gibt es noch die Möglichkeit, einen Kommentar abzugeben.

Da an der ETH aber seit 2001 mehr als 2200 verschiedene Dozierende unterrichtet haben, mussten diese zuerst gefiltert werden. Natürlicherweise sollten nur Dozierende angezeigt werden, bei welchen der Studierende schon mindestens eine Lerneinheit belegt hatte. Diese Daten mussten jedoch von der ETH zur Verfügung gestellt werden. Vom technischen Standpunkt aus gesehen würde sich eine Lösung mit SOAP-Schnittstellen anbieten. Das System mit den SOAP-Schnittstellen war zum Zeitpunkt der Programmierung aber noch in Entwicklung und nicht bereit für den Produktions-Einsatz. Ausserdem hätte eine Schnittstelle, die genau die für die Abstimmung erforderlichen Daten bereitstellt, extra programmiert werden müssen. Der erhebliche Mehraufwand wäre nicht in der zur Verfügung stehenden Zeit realisierbar gewesen. Darum hat das Rektorat die benötigten Daten als Snapshot der Datenbank zur Verfügung gestellt. Dieser konnte dann als Quelle der Datenbank für die Website benützt

werden.

Die Website für die Abstimmung wurde mit PHP-Skripten generiert und die Daten von einer MySQL-Datenbank serviert. Obwohl die Tabelle mit den Verlesungsdaten knapp 500'000 Einträge hatte, vermochte der Server die Last ohne Probleme zu bewältigen. Auch beim grossen Ansturm zu Beginn der Abstimmung. Die Authentifizierung der Studierenden wurde per ETH-LDAP durchgeführt. Das hatte den Vorteil, dass damit auch gleich noch das Departement und der Studiengang des Abstimmenden festgestellt werden konnten.

### 1.1.1 Datenschutz

Viel Wert wurde auf die richtige Anonymisierung der Daten gelegt. Dafür wurde für jeden Dozenten eine zufällige Kennzahl generiert, die für die weitere Auswertung benützt wurde. Erst ganz am Schluss, beim Bestimmen der Gewinner, wurden einzelne wieder umgewandelt. Auch sollten die Bewertungen nicht zum Stimmgebenden zurückverfolgt werden können. Die Stimmen einzelner Studenten mussten jedoch untereinander erkennbar bleiben.<sup>1</sup> Daher wurde für Identifikation der Bewertungen ein *Universal Unique Identifier* (global einzigartig in Raum und Zeit) verwendet. Mit diesem konnten die Stimmen ohne die Identität des Stimmenden gespeichert und dennoch sicher auseinander gehalten werden. Die Identität der Stimmenden wurde jeglich in einer separaten Tabelle gespeichert, um mehrfach Abstimmungen zu verhindern.

### 1.1.2 Ablauf einer Abstimmung

Die Anleitung zur Abstimmung wurde so kurz wie möglich gehalten, da sonst das Risiko bestand, dass viele Abstimmende sie nicht lesen würden.<sup>2</sup> Sie bekamen Tipps zur Abstimmung und wurden aufgefordert, nur für Dozierende abzustimmen, deren Lehreinheiten sie wirklich besucht und noch in guter Erinnerung hatten. Um dann die Abstimmung zu beginnen musste sich jeder mit seinem n.ethz-Login authentifizieren. Dann wurde eine Liste mit allen Dozierenden ausgegeben, bei welchen der Abstimmende schon je eine Lerneinheit belegt hatte. Diese wurden nach Semester und Fach gruppiert. Als zusätzliche Erinnerungshilfe konnte man sich die Lehreinheiten von den Dozierenden anzeigen lassen.

Es musste die Frage *Wie gut ist der Unterricht dieses Dozenten?* mit einer Bewertung von 1 bis 5 beantwortet werden. Die Skala lautete: *mies mühsam neutral angenehm genial*. Wurde für einen Dozenten keine Bewertung abgegeben, so wurde das wie *keine Äusserung möglich* interpretiert, was auch noch angekreuzt werden konnte.

### 1.1.3 Testlauf

In einem Testlauf durch Vertreter der Fachvereine konnten noch einige kleine Fehler ausgebügelt werden. Auch konnte die Navigation und die Darstellung verbessert werden. Damit wurde ein intuitiver und fehlerfreier Ablauf sichergestellt.

---

<sup>1</sup>Dies war für den Studenteneffekt nötig, siehe 2.2.

<sup>2</sup>Screenshots finden sich in Anhang B.

## 1.2 Die Abstimmung

An der Abstimmung durften nur im Sommersemester 2006 an der ETH eingeschriebene Studierende teilnehmen. Studierende des didaktischen Ausweises, HörerInnen und Doktorierende waren nicht zugelassen. Für die Erhebung der Daten wurden alle Dozierenden seit dem Wintersemester 2001/2002 berücksichtigt. Ausgeschlossen wurden jeglich Dozierende, die nur Praktikas oder Kolloquien gegeben hatten. Bei der Auswertung wurden aus Gründen der Berechenbarkeit und des Bekanntheitsgrades nur Dozierende zugelassen, die 10 oder mehr Bewertungen erhalten hatten.

Für die Gewinner in den einzelnen Departementen galt zusätzlich noch das Kriterium, dass sie in den letzten zwei Semestern, also im Wintersemester 05/06 und im Sommersemester 06, eine Lehrinheit an der ETH gehalten haben mussten.

Neben einem Polykum-Artikel und Plakaten in allen Vorlesungsgebäuden der ETH wurde vor allem mit einem Rundmail für die Abstimmung geworben. Studierende, die zwei Wochen nach Abstimmungsbeginn noch nicht teilgenommen hatten, erhielten noch ein Erinnerungsmail.

Der Rücklauf kurz nach dem Versand der ersten E-Mail war überwältigend: Mehr als 10% stimmten schon am ersten Tag ab. Im ganzen nahmen 2972 Studierende an der Abstimmung teil, das sind mehr als 33% der zur Abstimmung zugelassenen Studierenden.

8442 Studierende mit durchschnittlich 52 und maximal 250 verschiedenen Dozierenden konnten für 2166 Dozierende abstimmen. Von 94'580 Stimmen konnten 55'521 Stimmen für 987 Dozierende mit mehr als 10 Stimmen gezählt werden. Diese erhielten durchschnittlich 65.09 (Standardabweichung 79.19) und maximal 792 Stimmen. Viele Stimmen wurden nicht für die definitive Auswertung benützt, weil sie Veranstaltungen bewerteten, die mehr als vier Semester zurücklagen. Diese wurden aber bewusst auch erhoben, um den Einfluss dieser Stimmen beurteilen zu können.

Es gab sehr viel Lob für die Abstimmung und auch viele Kommentare zur Lage der Didaktik an der ETH. Oft wurden auch einzelne Dozierende noch speziell kommentiert. Häufig waren auch Kommentare zur Skala und zur Wortwahl dafür. Einige wünschten sich mehr Kriterien und dass man nur für Dozenten der letzten ein, zwei Jahre sollte abstimmen können. Nur wenige Kommentare waren darauf zurückzuführen, dass die Anleitung nicht gelesen worden war.



# Kapitel 2

## Analyse

### 2.1 Gesammelte Daten

Folgende Daten wurden zur Verwendung in SAS aus der Datenbank ausgelesen:

<i>y</i>	Bewertung
<i>s</i>	Nummer des Studenten (anonymisiert)
<i>d</i>	Nummer des Dozenten (codiert)
<i>studage</i>	Semester des Studenten (aufgerundet auf eine gerade Zahl)
<i>lectage</i>	Semester der Vorlesung
<i>dept</i>	Stammdepartement des Dozenten (codiert)
<i>service</i>	Indikatorvariable für Servicevorlesungen

### 2.2 Das Modell

Für die Auswertung wurde ein *general linear mixed model* gewählt. Grundsätzlich wird angenommen, dass die Daten normalverteilt sind und in einem linearen Zusammenhang mit einer gewissen Menge unbekannter Parametern stehen. Gegenüber einem einfachen linearen Modell sind in einem general linear mixed model auch kompliziertere Kovarianz-Matrizen möglich. Dies erlaubt es zusätzlich einen zufälligen Effekt pro abstimmender Person zu berücksichtigen. Das Modell wird mit

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\epsilon}$$

beschrieben.  $\boldsymbol{\beta}$  ist der Vektor der unbekanntten fixen Effekte mit der bekannten Designmatrix  $\mathbf{X}$ .  $\boldsymbol{\gamma}$  beschreibt den Vektor der unbekanntten zufälligen Effekte mit der bekannten Designmatrix  $\mathbf{Z}$ . Die Residuen  $\boldsymbol{\epsilon}$  werden als unabhängig und identisch normalverteilt angenommen. Es wird auch noch angenommen, dass  $\boldsymbol{\gamma}$  und  $\boldsymbol{\epsilon}$  unkorreliert sind und Erwartungswert  $\mathbf{0}$  und Varianzen  $\mathbf{G}$  beziehungsweise  $\mathbf{R}$  haben. Die Varianz von  $\mathbf{y}$  ist damit

$$\mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R}$$

Folgendes Modell wurde gewählt:

$$y = \mu + \text{studage} + \text{dept} + \text{dept} * \text{service} + z + \epsilon$$

$y$  ist die Bewertung.  $studage$ ,  $dept$ ,  $dept*service$  sind fixe Effekte.  $z$ , der Studenteneffekt, ist ein zufälliger Effekt.  $\mu$ , der gesuchte Dozenteneffekt, wird als fixer Effekt behandelt.  $\epsilon$  ist der zufällige Fehler.

Um dies in die obige Form zu bringen, werden noch einige Zuordnungsfunktionen in Abhängigkeit der Stimme  $k = 1, \dots, N$  benötigt:

$d(k) \in \{1, \dots, m\}$	Nummer des Dozierenden
$e(d(k)) \in \{1, \dots, 15\}$	Nummer des Departements des Dozierenden $d(k)$
$s(k) \in \{1, \dots, n\}$	Nummer des Studierenden
$t(s(k)) \in \{2, 4, 6, 8, 10\}$	Anzahl Semester des Studierenden $s(k)$
$u(k) \in \{0, 1\}$	Service-Indikator-Variablen

Damit werden die Effekte und Fehler zu:

$y_k$	Bewertung Nummer $k$
$\mu_{d(k)}$	Dozenteneffekt des Dozierenden $d(k)$
$studage_{t(s(k))}$	Effekt der Anzahl Semester
$dept_{e(d(k))}$	Effekt des Departements $e(d(k))$
$dept_{e(d(k))} * service_{u(k)}$	Effekt der Servicevorlesungen im Dept. $e(d(k))$
$z_{s(k)}$	zufälliger Effekt des Studenten $s(k)$
$\epsilon_k$	zufälliger Fehler der Stimme $k$ .

Sei  $\Gamma_{i,j}^{n \times m}$  die  $n \times m$  Matrix mit allen Zellen gleich 0 und einer 1 in der Zelle in Zeile  $i$  und Spalte  $j$ . Für  $N$  die Anzahl Bewertungen,  $n$  die Anzahl abstimmender Studierenden,  $m$  die Anzahl bewerteter Dozenten und  $M := m + 50$  können die Modellmatrizen geschrieben werden als

$$\mathbf{X} = \sum_{k=1}^N \left[ \Gamma_{k,d(k)}^{N \times M} + \Gamma_{k,m+t(s(k))/2}^{N \times M} + \Gamma_{k,m+5+e(d(k))}^{N \times M} + \Gamma_{k,m+20+2e(d(k))-u(k)}^{N \times M} \right]$$

$$\mathbf{Z} = \sum_{k=1}^N \Gamma_{k,s(k)}^{N \times n}$$

$$\boldsymbol{\beta} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_m \\ studage_2 \\ studage_4 \\ studage_6 \\ studage_8 \\ studage_{10} \\ dept_1 \\ \vdots \\ dept_{15} \\ dept_1 * service_1 \\ dept_1 * service_0 \\ \vdots \\ dept_{15} * service_1 \\ dept_{15} * service_0 \end{pmatrix} \quad \mathbf{b} = \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix} \quad \boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_N \end{pmatrix}$$

Es wird angenommen, dass die zufälligen Effekte alle unabhängig und normalverteilt sind:

$$z_{s(k)} \sim \mathcal{N}(0, \sigma_z) \quad \forall k$$

$$\epsilon_k \sim \mathcal{N}(0, \sigma_\epsilon) \quad \forall k$$

Am Anfang der Entwicklung stand ein einfaches Modell, bei dem nur Dozierenden-, Studenten- und Serviceeffekte berücksichtigt wurden. Dann wurden sukzessive mehr Effekte eingefügt, um den vorliegenden Bedingungen Rechnung zu tragen. Die Grenzen setzten dabei die vorliegenden Daten. So war es zum Beispiel nicht möglich, einen Effekt für das Departement oder für den Studiengang des Studierenden einzufügen. Auf diese Daten wurde zu Gunsten der Anonymisierung der Bewertungen verzichtet. Diese und weitere eventuell fehlenden Parameter werden vom zufälligen Fehler der Stimme aufgenommen.

Mit dem *studage*-Effekt wird erreicht, dass etwaige Unterschiede im Bewertungsverhalten von verschiedenen Gruppen Studierender aneinander angeglichen werden. Der *dept*-Effekt wurde eingeführt, um die Departemente untereinander anzugleichen und den *service*-Effekt von diesen Unterschieden zu befreien.

Zuerst wurde im Modell nur ein *service*-Effekt für alle Departemente berücksichtigt. Ein Durchgang mit separaten *service*-Effekten (dargestellt in Abbildung 2.1) hat aber ziemliche Unterschiede zwischen den Departementen ergeben. Im Departement 3 ist die Differenz zwischen *service* und normalen Veranstaltungen am grössten, wobei die ersteren besser bewertet wurden. In zehn Departementen wurden die normalen Lerneinheiten höher bewertet.

Das ergibt für PROC MIXED in SAS folgenden Aufruf:

```
proc mixed data=Eule.koller_red CL NOCLPRINT;
  class dept s d studage lectage service;
  model y = dept d(dept) studage dept*service
    / S OUTPREDM=Eule.Resultate_predm
      OUTPRED=Eule.Resultate_pred;
  random intercept / subject=s(studage);
  parms 0.07402 0.9145;
  lsmeans dept studage;
  lsmeans dept*service;
  lsmeans d(dept) / CL;
  ods output LSMeans=Eule.Resultate
    CovParms=Eule.Resultate_covparams
    SolutionF=Eule.Resultate_solutionf;
run;
```

Skripte werden in SAS immer sequentiell ausgeführt. Ein Aufruf einer Prozedur beginnt mit `proc NAME`, gefolgt von Argumenten, die für die ganze Prozedur gelten. Danach werden die verschiedenen Funktionen der Prozedur aufgerufen und weitere benötigte Informationen übergeben. Dies erfolgt immer mit dem folgenden Schema:

```
Anweisung Argumente / Optionen;
```

Mit `run;` wird die Definition der Prozedur abgeschlossen und der eingelesene Code wird ausgeführt.

Die `class` Anweisung gibt an, welche Spalten der zu bearbeitenden Datentabelle für die Erstellung der Indikatorvariablen in den Designmatrizen des Modells benutzt werden sollen.

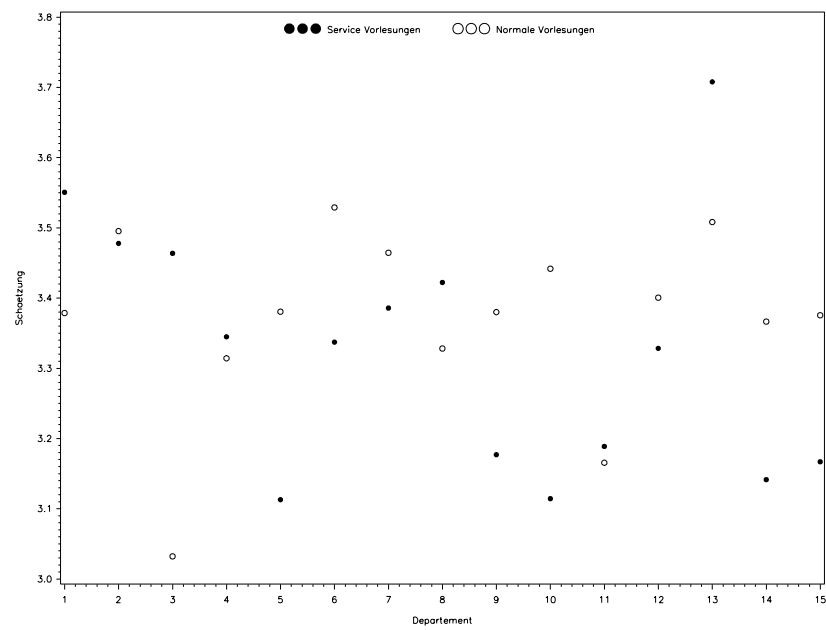


Abbildung 2.1: Effekt der Servicevorlesungen in den verschiedenen Departementen. Ein Schlüssel für die Zuordnung der Departemente findet sich in Anhang A.

In `model` wird die abhängige Variable ( $y$ ) auf der linken Seite des = und auf der rechten Seite die fixen Effekte definiert. `d(dept)` beschreibt, dass die Variablen `d` und `dept` nur in einer Kombination vorkommen. `dept*service` definiert alle vorkommenden Kombinationen beider Variablen als einzelne Effekte. Die Option `S` verlangt, dass die Lösungen für die fixen Effekte ausgerechnet werden. Die weiteren Optionen definieren die Tabellen, in die die addierten Effekte für jede Stimme ausgegeben werden sollen. Dies wird für den nachfolgenden Bootstrap-Prozess benötigt. Ein Abschnittseffekt (`intercept`) wird automatisch miteinbezogen.

Mit `random` werden die zufälligen Effekte definiert. Hier muss der Abschnittseffekt explizit gefordert werden. Die `subject`-Option definiert den zufälligen Effekt für die Testpersonen.

Die Kovarianzmatrizen  $\mathbf{G}$  und  $\mathbf{R}$  werden mit der *restricted maximum likelihood (REML)* Methode geschätzt. Dabei benützt PROC MIXED einen *ridge-stabilized Newton-Raphson* Algorithmus um die *log-likelihood* Funktion zu optimieren. Die maximale Anzahl Iterationen ist jedoch beschränkt und reicht häufig nicht aus, um das Optimum zu finden. Darum werden mit `parms` für diese Daten sinnvolle Startwerte eingesetzt. So wird vermieden, dass die Prozedur mit der Fehlermeldung *infinite likelihood* abgebrochen werden muss.

Die Kovarianzmatrizen tauchen in den *mixed model equations*

$$\begin{bmatrix} \mathbf{X}'\widehat{\mathbf{R}}^{-1}\mathbf{X} & \mathbf{X}'\widehat{\mathbf{R}}^{-1}\mathbf{Z} \\ \mathbf{Z}'\widehat{\mathbf{R}}^{-1}\mathbf{X} & \mathbf{Z}'\widehat{\mathbf{R}}^{-1}\mathbf{Z} + \widehat{\mathbf{G}}^{-1} \end{bmatrix} \begin{bmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{\gamma}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\widehat{\mathbf{R}}^{-1}\mathbf{y} \\ \mathbf{Z}'\widehat{\mathbf{R}}^{-1}\mathbf{y} \end{bmatrix}$$

als geschätzte  $\widehat{\mathbf{R}}$  und  $\widehat{\mathbf{G}}$  wieder auf. Diese Gleichungen werden durch

$$\begin{aligned} \widehat{\boldsymbol{\beta}} &= (\mathbf{X}'\widehat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\widehat{\mathbf{V}}^{-1}\mathbf{y} \\ \widehat{\boldsymbol{\gamma}} &= \widehat{\mathbf{G}}\mathbf{Z}'\widehat{\mathbf{V}}^{-1}(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}) \end{aligned}$$

gelöst, die geschätzten fixen und zufälligen Effekte.

In dieser Form können die Effekte aber nur schlecht miteinander verglichen werden, da sie oft ganz unterschiedliches Gewicht im Modell haben. Darum berechnet man die *Least-squares means*<sup>1</sup>, dabei werden mittlere Effekte von den anderen erklärenden Grössen addiert. So werden die Differenzen zwischen den Effekten vergleichbar gemacht.

Die `lsmeans` Anweisung definiert, für welche fixen Effekte die *Least-squares means* berechnet werden sollen. Dies benötigt jedoch sehr viel Arbeitsspeicher und es werden darum Daten auf die Festplatte ausgelagert. Dies verlängert die Berechnung enorm. Für den Bootstrap-Prozess wurden darum die *Least-squares means* nur für den Dozenteneffekt berechnet. Die `CL` Option bewirkt, dass *t-type* 95%-Vertrauensintervalle für jeden angegebenen Effekt ausgegeben werden.

### 2.2.1 Fixe und zufällige Effekte

Der Unterschied zwischen fixen und zufälligen Effekten liegt in den Annahmen, die man darüber macht. Nimmt man an, dass alle möglichen Werte, die ein fixer Effekt annehmen kann, in der Studie vorhanden sind, so wählt man einen fixen Effekt. Handelt sich es aber eher um eine zufällige Auswahl aller möglichen Werte, so berechnet man es als zufälligen Effekt.

<sup>1</sup>Die *Least-squares means* werden manchmal auch als *adjusted means* bezeichnet.

Fixe Effekte haben den Vorteil, dass man weniger Annahmen über sie treffen muss. Die Berechnung von fixen Effekten benötigt aber sehr viel Arbeitsspeicher. Bei etwa 76'000 Stimmen und 2500 fixen Effekten wächst die Auslagerungsdatei zwischenzeitlich auf 46 Gigabyte an. Wegen beschränkter Ressourcen ist darum zum Beispiel kein fixer *d\*lectage*-Effekt möglich. Wählt man stattdessen einen zufälligen Effekt, so gelingt die Berechnung. Wie man in Abbildung 2.2 sieht, sind die Dozenteneffekte annähernd normalverteilt. Ein Vergleichslauf mit dem Dozenteneffekt als zufälligen Effekt zeigte darum auch kaum Unterschiede.

### 2.2.2 Ausgabe

In der vorangegangenen Beschreibung sind schon einige Optionen betreffend der Ausgabe der Prozedur beschrieben worden. In Anhang C findet sich die Ausgabe für den Durchlauf mit den vorliegenden Daten. Zuerst werden einige Informationen über das Modell, die Dimensionen und die Anzahl der Beobachtungen ausgegeben. Es folgt die *Iteration History* des *Newton-Raphson*-Algorithmus, eine Zeile über die Konvergenz dieses Verfahrens und danach die geschätzten Kovarianz-Parameter  $\sigma_z$  (*Intercept*) und  $\sigma_\epsilon$  (*Residual*). Die *Fit Statistics* liefern einige Statistiken über das geschätzte Modell. Ausserdem werden die Resultate der durchgeführten *Type 3*-Tests der fixen Effekte aufgelistet.

## 2.3 Kontrolle des Modells

Ein *Type 3*-Test eines fixen Effekts<sup>2</sup> testet die Signifikanz des Effektes für den Fall, dass alle anderen Effekte miteinbezogen werden. Zum Beispiel werden für den *studage*-Effekt die Hypothesen

$H_0 : studage = 0$  gegeben, dass  $d(dept)$ ,  $dept$ ,  $dept*service$  miteinbezogen werden im Modell

gegen  $H_A : studage \neq 0$  gegeben, dass  $d(dept)$ ,  $dept$ ,  $dept*service$  miteinbezogen werden im Modell

getestet. Typischerweise wird aus einem  $p$ -Wert grösser als 0.05 geschlossen, dass *dept* kein signifikanter Effekt im vorliegenden Modell ist.

Die Resultate dieses Tests sind in Tabelle 2.1 dargestellt. Ausser beim *studage*-Effekt sind alle Tests auf Signifikanz erfolgreich. Der Einfluss dieses Effektes ist denn auch minim. Zur Überprüfung dieser Feststellung in zukünftigen Auswertungen wurde der Effekt jedoch im Modell beibehalten.

Effekt	$Pr > F$
<i>dept</i>	< .0001
$d(dept)$	< .0001
<i>studage</i>	0.1122
$dept*service$	< .0001

Tabelle 2.1: *Type 3*-Test der fixen Effekte

<sup>2</sup>*Type 3*-Tests werden im Abschnitt *Default output* im Kapitel *The MIXED procedure* in [3] und in [7] als *Tests for Individual Coefficients* genauer beschrieben.

### 2.3.1 Bootstrap

PROC MIXED stellt zur Überprüfung der Genauigkeit der geschätzten Werte Vertrauensintervalle zur Verfügung. Diese sind hier aber nur beschränkt interessant, da am Ende eine Rangliste der Dozierenden gebildet wird. Die Vertrauensintervalle der Schätzwerte lassen sich nicht einfach auf Vertrauensintervalle der Ränge übertragen. Am einfachsten ist es, solche Intervalle mit einem Bootstrap zu simulieren. Eine solche Simulation kann ohne eine zusätzliche Datenerhebung durchgeführt werden. Mit den bestehenden Daten und den geschätzten fixen Effekten werden neue Realisierungen mit zufälligen Unterschieden simuliert. Dann wird das Modell von neuem berechnet und die neuen Resultate mit den alten verglichen.

Die einfachste Variante eines Bootstraps benutzt die empirische Verteilung, um Realisierungen zu simulieren, also im Wesentlichen durch ein Ziehen mit Zurücklegen. Damit wird aber die Student-Dozent-Studiengang-Struktur der Daten zerstört und viele Effekte könnten nicht mehr berücksichtigt werden. Ein Test hat ergeben, dass sich die Realisierungen damit sehr stark unterscheiden und keine brauchbaren Resultate liefern.

Darum wurde entschieden, gemäss dem Modell zu simulieren, das heisst, für jede Bewertung die fixen Effekte als richtig angenommen und je mit einem simulierten Studenteneffekt und zufälligen Fehler versehen. Die erhaltenen Werte wurden wieder auf ganze Zahlen zwischen 1 und 5 gerundet (dabei wurde die Verkleinerung der Varianz ignoriert) und damit das Modell neu berechnet. Das ganze wurde 21 Mal wiederholt.

Die Kovarianzen wurden von PROC MIXED im Bootstrap-Prozess durchgehend höher eingeschätzt, obwohl durch die Rundung die Varianz eher verkleinert werden sollte.

	$\sigma_z$	$\sigma_\epsilon$	Verhältnis
ursprüngliche Daten	0.07323	0.9021	12.32
Bootstrap	0.09650	1.3776	14.28

Tabelle 2.2: Geschätzte Varianzen

Der Grund dieses Unterschiedes konnte nicht eruiert werden. Da mit dem Bootstrap-Prozess die Lage der Vertrauensintervalle viel besser bestimmt werden kann, wurde der Bootstrap aber trotzdem weitergeführt.

Der Bootstrap wurde in SAS mit Hilfe von Makros programmiert, die nacheinander mehrmals aufgerufen wurden. Darum ist es nicht möglich, die Zufallszahlen einfach mit einem Seed zu generieren. Die eigentlich dafür vorgesehene Methode, CALL RANNOR, funktioniert im Zusammenspiel mit Makros nicht. Es wurde daher der Empfehlung aus [2] gefolgt, wo vorgeschlagen wird die Seeds in einem separaten Schritt zu generieren. Von einem Startseed ausgehend werden Zufallszahlen generiert und jede 100'000ste als ein Seed für eine Realisierung genommen.

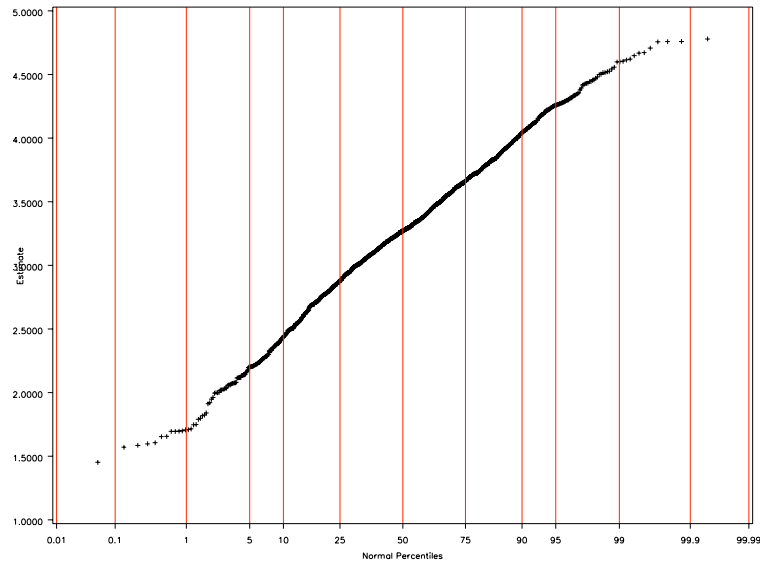


Abbildung 2.2: Quantilplot aller geschätzten Dozenteneffekte.

## 2.4 Resultate

Im folgenden sind alle Abbildungen aus demselben Departement, ausser bei der Abbildung 2.2, die über alle Departemente gerechnet wurde. Mit *Estimate* oder *Schaetzung* ist immer der Dozierendeneffekt gemeint.

### 2.4.1 Anpassung der Daten

Die Abbildung 2.3 zeigt einen Vergleich zwischen den unkorrigierten und den korrigierten Mittelwerten. Man sieht, dass Werte nach unten wie auch nach oben korrigiert wurden. Abbildung 2.4 zeigt, dass die Werte über das ganze Spektrum im gleichen Masse angepasst wurden. Es kam also nicht zu dem zu erwarteten Effekt, dass die oberen Werte generell nach unten und die unteren nach oben korrigiert wurden. Dass viele Werte unter der Gerade mit Steigung 1 liegen, kommt daher, dass die Werte gegenüber den anderen Departementen korrigiert wurden. Es gibt andere Departemente, die eher gehoben wurden. Dies ist eine Folge der Wahl des Modells und hat für die Auswertung keinen Einfluss, da nur die Dozenten innerhalb eines Departements miteinander verglichen werden. Das Modell benötigt den interdepartementalen Vergleich jedoch, um die Effekte der Service-Vorlesungen und Einflüsse von Studenten ausserhalb von ihrem Stammdepartement auszugleichen.

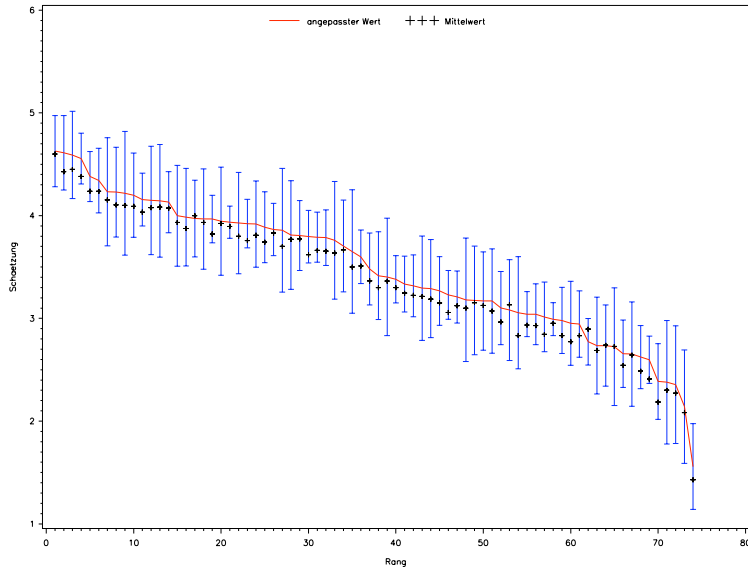


Abbildung 2.3: Unveränderte und korrigierte Mittelwerte mit 95%-Vertrauensintervallen

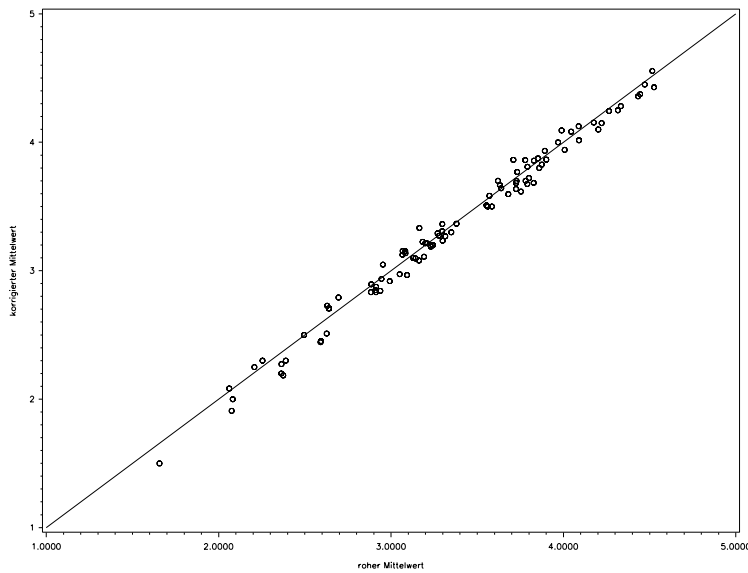


Abbildung 2.4: Anpassung der Mittelwerte

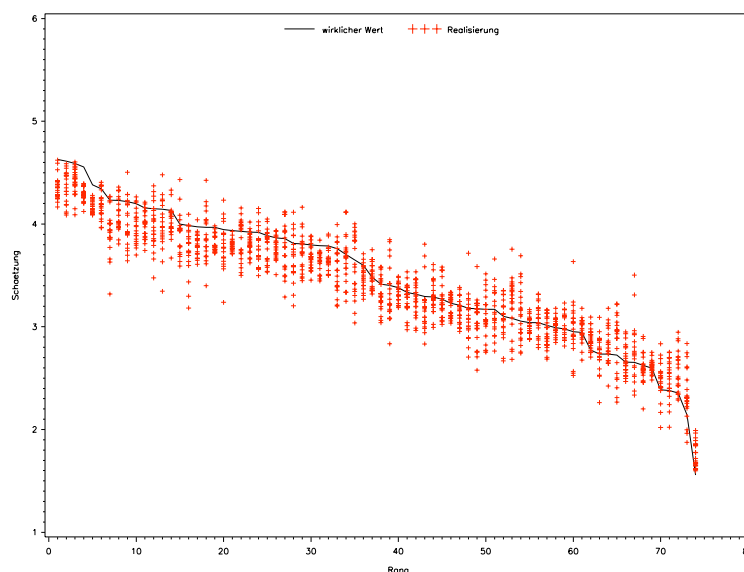


Abbildung 2.5: Resultate des Bootstraps zur Überprüfung der Vertrauensintervalle

## 2.4.2 Genauigkeit der Schätzwerte

Die von PROC MIXED angegebenen Vertrauensintervalle in Abbildung 2.3 mit den Werten der verschiedenen Realisierungen im Bootstrap in Abbildung 2.5 haben in der Länge eine gute Übereinstimmung. Die Intervalle in der oberen Hälfte sind vielfach zu hoch, im Gegensatz zu denen in der unteren Hälfte, die eher etwas zu weit nach unten gehen. Das jeweilige andere Ende stimmt aber sehr gut mit den simulierten Werten überein. Das liegt daran, dass PROC MIXED nicht weiss, dass nur Werte zwischen 1 und 5 möglich sind.

Um einen Eindruck der Auswirkungen der Unsicherheiten auf die Rangliste zu erhalten, kann man die erwartete Rangänderung für die einzelnen Dozenten mit dem Bootstrap, wie in [4] beschrieben, approximieren. Dies ist in Abbildung 2.6 dargestellt. + markiert die erwartete Änderung des Ranges, würde eine reine Permutation durchgeführt. Die Permutation erreicht die maximalen Erwartungswerte an den Rändern, während die Resultate des Bootstraps zeigen, dass die Ränge nur in der Umgebung schwanken. In der Mitte sind die Schwankungen grösser, da dort die Unterschiede zwischen den einzelnen Dozenten geringer sind als am Rand (siehe Abbildung 2.2). Das Maximum der Erwartungswerte bleibt in allen Departementen unter einem Fünftel der totalen Anzahl Rängen.

Die Unsicherheiten in der Rangierung kommen durch den Bootstrap gut zum Vorschein. Am besten kann man die Verteilung der Ränge eines Dozenten durch einen Boxplot, wie in Abbildung 2.7 dargestellt, visualisieren. Im vorliegenden Beispiel sieht man, dass die ersten vier Ränge (reale Daten) allesamt mindestens einmal den ersten Rang erreicht haben. Die 75%-Quantile des vierten Ranges sind aber ziemlich viel höher als die der ersten drei.

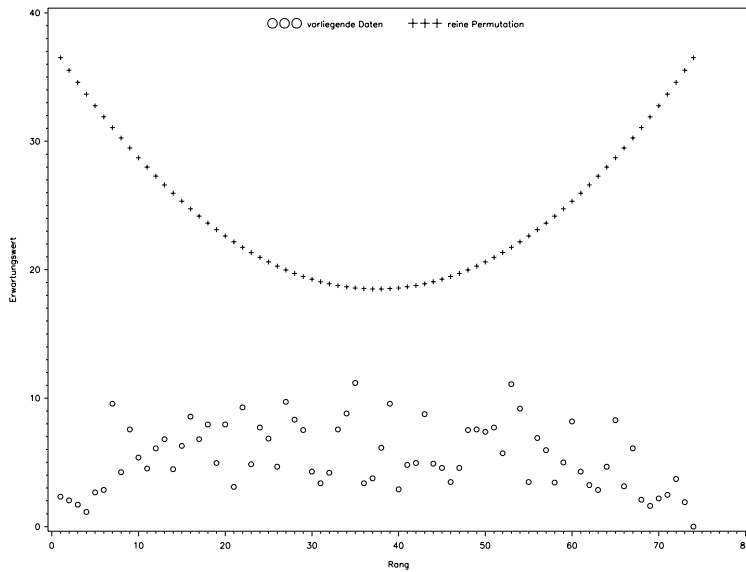


Abbildung 2.6: Erwartete Rangänderung

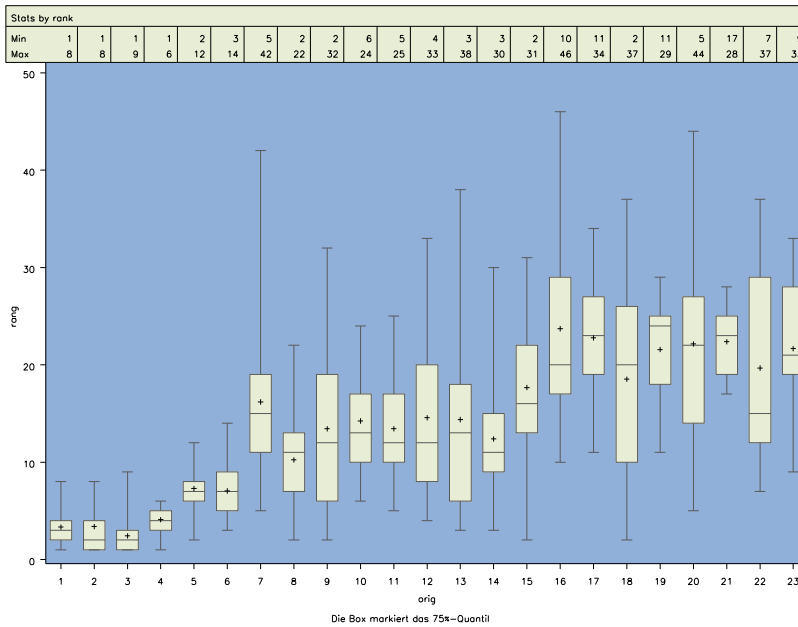


Abbildung 2.7: Boxplot der erreichten Ränge (*rang*) im Bootstrap. *orig* bezeichnet den Rang im Modell mit den richtigen Daten

Eine andere Art dies zu visualisieren wäre die Anzahl Gewinne - gemittelt über alle Realisierungen - für jeden Rang zu plotten. Da die Ränge stark abhängig sind, gäbe ein solcher Plot ein falsches Bild wieder. So verändert sich das Bild oft enorm, wenn man den ersten Rang in der Berechnung ignoriert, und es kommt vor, dass dann andere Ränge in den Vordergrund treten, die vorher nie einen ersten Rang im Bootstrap belegt hatten.

### 2.4.3 Einfluss der Vergangenheit

Die Studierenden konnten in der Abstimmung für alle Dozierenden abstimmen, bei welchen sie seit dem Wintersemester 2001 eine Veranstaltung besucht hatten. Ein Durchlauf des Modells mit den kompletten Daten und zusätzlichen *studage*- und *lectage*-Effekten ergibt folgende *Least-squares means*:

studage	2	4	6	8	10	
Schätzwert	3.1863	3.2416	3.2657	3.3148	3.3517	
lectage	1	2	3	4	5	6
Schätzwert	3.4054	3.3313	3.2902	3.2205	3.2410	3.1437

Werte von *lectage* höher als 6 wurden auf 6 reduziert, damit alle Effekte geschätzt werden konnten.

Die Resultate zeigen einen klaren Trend: Je länger der Abstimmende schon an der ETH studiert, umso besser werden seine Bewertungen. Mit dem *Reverse Arrangement Test*, aus [5], ergibt sich dafür eine 99% Wahrscheinlichkeit eines Trendes nach oben im Vergleich zu keinem Trend.

Führt man den gleichen Test für einen Abwärtstrend des *lectage*-Effekts bei grösserem zeitlichen Abstand zur Erfahrung durch, so kommt man auf eine 99% Wahrscheinlichkeit eines Trendes im Vergleich zu keinem Trend.

Die Resultate zeigen eine Diskrepanz: Obwohl Veranstaltungen, die länger zurück liegen, schlechter abschneiden, stimmen die gleichen Studierenden im Durchschnitt besser ab. Offenbar ändert sich das Abstimmungsverhalten, wenn die Erinnerung nicht mehr ganz so klar ist. Darum wurde entschieden, nur die Bewertungen für Lerneinheiten der letzten vier Semester zu berücksichtigen.

### 2.4.4 Grosse und kleine Veranstaltungen

Abbildung 2.8 zeigt ein Streudiagramm mit der Anzahl Studierenden und dem Schätzwert des Dozierendeneffekts. Es sind keine klaren Effekte erkennbar, die in Zusammenhang mit der Grösse der Veranstaltungen stehen könnten. Um die Frage zu klären, ob grosse Vorlesungen besser als kleine Vorlesungen bewertet werden, wurden die Dozierenden in zwei Gruppen am Median der Stimmenanzahl geteilt. Getestet wurde

$H_0$ : Die Mittelwerte der Bewertungen in den beiden Gruppen sind gleich  
gegen  $H_A$ : Die Mittelwerte unterscheiden sich.

Ein *Wilcoxon Zwei-Stichproben* Test wurde mit einem 95%-Signifikanzniveau verworfen. Der Test gibt uns ein Indiz dafür, dass die Veranstaltungen mit weniger Studierenden unterschiedlich bewertet wurden. Das Modell wurde um einen zusätzlichen Effekt erweitert, der die beiden Gruppen aneinander anpassen sollte. PROC MIXED hat diesen Effekt aber als nicht schätzbar eingestuft. Die im selben Plot dargestellte Kurve der mit *lowess* geglätteten Schätzwerte erklärt

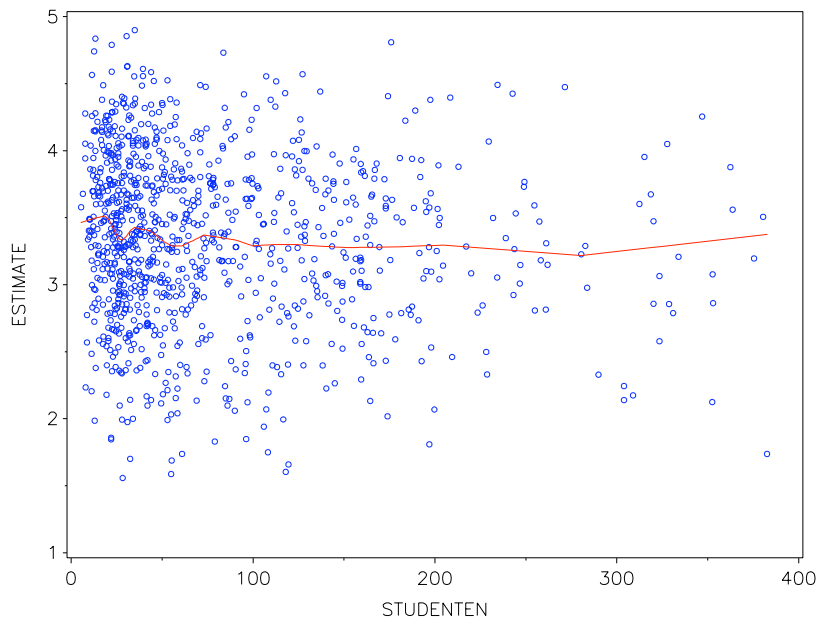


Abbildung 2.8: Schätzwerte der Dozierendeneffekte in Abhängigkeit der Anzahl Studierende. Die rote Linie bildet die mit *lowess* geglätteten Schätzwerte ab. Die Fenstergröße beinhaltet 20% der Beobachtungen. Aus Datenschutzgründen wurde der Plot bei 400 Studierenden abgeschnitten.

das Verhalten von SAS: Es sind keine klaren Gruppen erkennbar und mögliche Unterteilungen entsprechend willkürlich. Im Vergleich zu der Spannweite der Resultate, sind die Schwankungen aber klein.

Dozierende, die sowohl grosse als auch kleine Veranstaltungen gehalten haben, werden im Modell nicht speziell berücksichtigt. Die Stimmen, die ein Dozierender für seine kleine Veranstaltungen erhielt, würden im Modell gegenüber den Bewertungen für seine grossen Lehreinheiten benachteiligt. Wie der obige Vergleich gezeigt hat, ist die Beziehung mit der Anzahl Studenten und damit der Grösse der Veranstaltung vernachlässigbar, der Nachteil also höchstens ein theoretischer.



## Kapitel 3

# Bestimmen der Gewinner

Die Analyse mit dem Bootstrap hat gezeigt, dass man nicht einfach den vom Modell vorgeschlagenen Sieger übernehmen, sondern vielmehr eine Auswahl aus den ersten paar Rängen nehmen sollte. Als Kriterium wurde verwendet, dass das 75%-Quantil der Bootstrap-Schätzwerte eines möglichen Kandidaten mindestens das 25%-Quantil des ersten Ranges erreichen musste um in die engere Auswahl zu kommen.

Von dieser Auswahl wurde mit folgenden Kriterien der Sieger bestimmt:

1. Verteilung der Stimmen, insbesondere Anzahl bzw. Existenz von stark negativen Stimmen
2. Abstand des Schätzwerts zu denjenigen der folgenden Ränge
3. Diversität der für den Dozierenden stimmenden Studierenden (*studage*)
4. Anzahl der Stimmen
5. Diversität der Art der Vorlesungen
6. Beurteilung durch eine Person aus dem dem Departement zugeordneten Fachverein.

Die Meinung des Fachvereines wurde eingeholt, um Fehlentscheidungen zu vermeiden. Zum Beispiel könnten aktuelle Entwicklungen (etwa ein Todesfall) in die Entscheidung miteinbezogen werden, die durch die vorhanden Daten noch nicht reflektiert wurden. Ausserdem ist das Vorlesungsverzeichnis für Departementsexterne oft die einzig brauchbare Ressource, um etwas über die gewählten Dozierenden herauszufinden.

Für das Departement 1 wären, wie aus Abbildung 2.7 ersichtlich, die ersten vier Ränge zu berücksichtigen.

Die Sieger des letzten Jahres sind bei einem Drittel der Departemente in die engere Auswahl gekommen. Nur in zwei Departementen ist der Sieger vom letzten Jahr nicht unter den ersten zehn Rängen platziert.



# Anhang B

## Screenshots der Website

### Die goldene Eule 2006

---



#### **Willkommen**

Die goldene Eule 2006 ist ein Sympathiepreis von den Studierenden an die Dozierenden der ETH Zürich, der vom VSETH verliehen wird.

Alle im Moment an der ETH angestellten Personen mit einem Lehrauftrag können den Preis gewinnen. Es wird ein Preis pro Departement verliehen.

Zur Abstimmung zugelassen sind alle Studierenden der ETH Zürich. Die Abstimmung läuft bis am 18. Juni.

[Zur Abstimmung](#)

[Resultate der goldenen Eule 2005](#)

[Kontakt](#)

---

**Anleitung**

Du loggst dich mit deinem n.ethz-Usernamen und Passwort ein. Dann kommt eine Liste von allen Dozierenden, bei welchen du je an einer Lehrveranstaltung eingeschrieben warst, aufgeteilt nach Semestern und gruppiert pro Fach. Jede Dozentin / jeder Dozent wird nur einmal angezeigt.

Die Punkte "mies" bis "genial" werden als Noten von 1 bis 5 interpretiert.

Hat es Dozierende darunter, die dir nicht bekannt sind, musst du nichts ankreuzen. Dies wird gleich interpretiert wie "k.A.". Geht dir die Liste zu weit in die Vergangenheit, klicke dich einfach bis zum Schluss durch.

Benütze die Pfeile um vor oder zurück zu navigieren.

Wenn du einen Dozenten / eine Dozentin keiner Lehrinheit zuordnen kannst, dann fahre mit dem Mauszeiger über den Namen und es wird dir eine Liste mit seinen / ihren Lehrheiten angezeigt.

Zum Schluss kannst du noch einen Kommentar abgeben.



Anmelde-Seite

Melde dich hier an:

n.ethz-Username

n.ethz-Passwort

Login

Stimm-Seite (Semester 2006S)

Wie gut ist der Unterricht dieser Dozenten?

	k.A.	mies	mühsam	neutral	angenehm	genial
Zehnder Eduard	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Künsch Hans Rudolf	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Lanford III Oscar E.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>



Stimm-Seite (Semester 2006S)



Wie gut ist der Unterricht dieser Dozenten?

	k.A.	mies	mühsam	neutral	angenehm	genial
Zehnder Eduard	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Künsch Hans Rudolf	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Lanford III	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Lerninheiten bei Künsch Hans Rudolf:  
Stochastische Simulation  
Mathematische Grundlagen der Statistik  
Wahrscheinlichkeit und Statistik



Kommentar-Seite



Dein Kommentar zur Abstimmung:

Senden



Die goldene Eule 2006



Vielen Dank für deine Teilnahme!



# Anhang C

## Output von PROC MIXED

### The Mixed Procedure

#### Model Information

Data Set	EULE.KOLLER_RED
Dependent Variable	y
Covariance Structure	Variance Components
Subject Effect	s(studage)
Estimation Method	REML
Residual Variance Method	Profile
Fixed Effects SE Method	Model-Based
Degrees of Freedom Method	Containment

#### Dimensions

Covariance Parameters	2
Columns in X	1038
Columns in Z Per Subject	1
Subjects	2969
Max Obs Per Subject	66

#### Number of Observations

Number of Observations Read	55521
Number of Observations Used	55521
Number of Observations Not Used	0

#### Parameter Search

CovP1	CovP2	Variance	Res Log Like	-2 Res Log Like
0.07402	0.9145	0.9021	-77642.9870	155285.974

Iteration History

Iteration	Evaluations	-2 Res Log Like	Criterion
1	2	155285.96987933	0.00000000

Convergence criteria met.

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Alpha	Lower	Upper
Intercept	s(studage)	0.07323	0.05	0.06702	0.08036
Residual		0.9021	0.05	0.8912	0.9132

Fit Statistics

-2 Res Log Likelihood	155286.0
AIC (smaller is better)	155290.0
AICC (smaller is better)	155290.0
BIC (smaller is better)	155302.0

PARMS Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
1	0.00	0.9488

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
dept	14	52E3	8.68	<.0001
d(dept)	972	52E3	24.59	<.0001
studage	4	2964	1.87	0.1122
dept*service	15	52E3	12.53	<.0001

## Anhang D

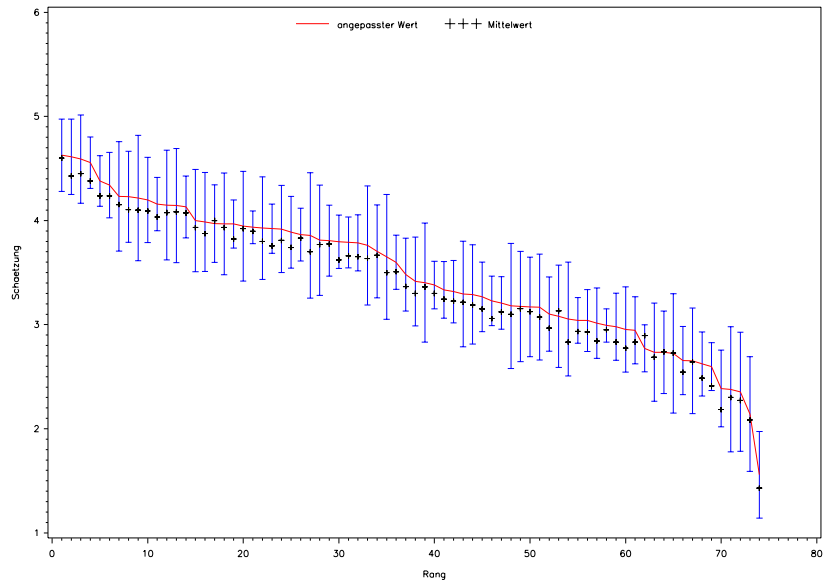
# Plots für alle Departemente

Auf den folgenden Seiten sind die Plots der Vertrauensintervalle und der Bootstrap-Quantilen für alle Departemente abgebildet.

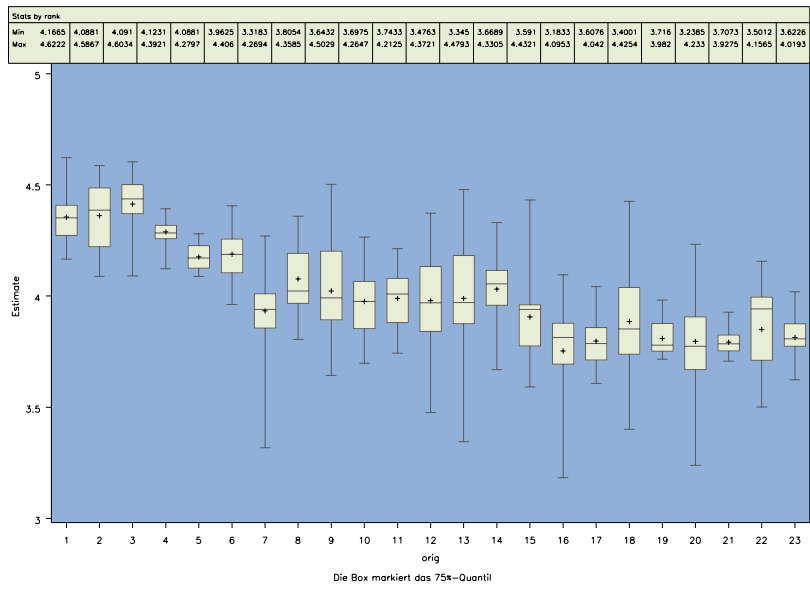
Der Erste zeigt jeweils die *t-type* 95%-Vertrauensintervalle, die von PROC MIXED berechnet wurde. Die Dozierenden sind in absteigender Reihenfolge in der x-Achse aufgelistet. In der y-Achse sind die Schätzwerte und die dazugehörigen Vertrauensintervalle abgebildet.

Der zweite Plot enthält die Resultate des Bootstraps für die ersten 23 Ränge. In der y-Achse sind die Schätzwerte abgebildet. Der untere Rand der Box markiert das 25%-Quantil, der Obere das 75%-Quantil. Das + markiert den Mittelwert und — den Median.

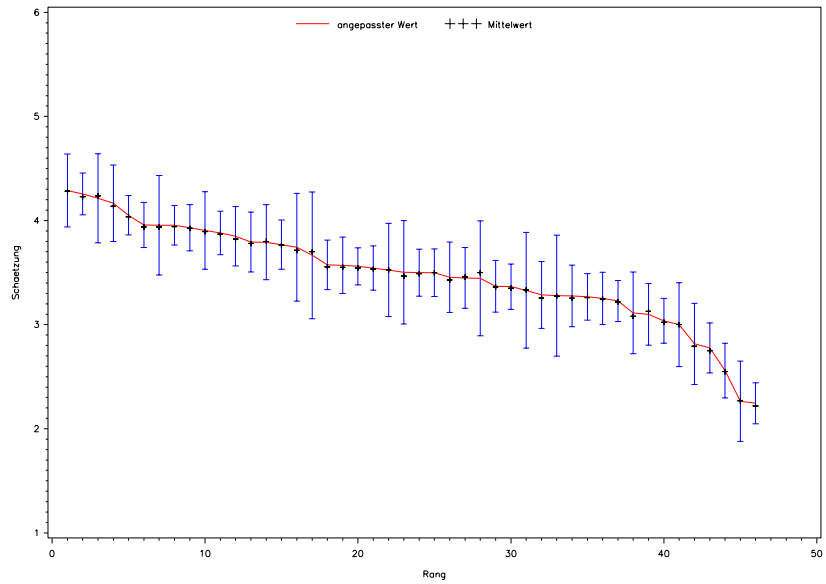
Departement 1, Least-Squares Means



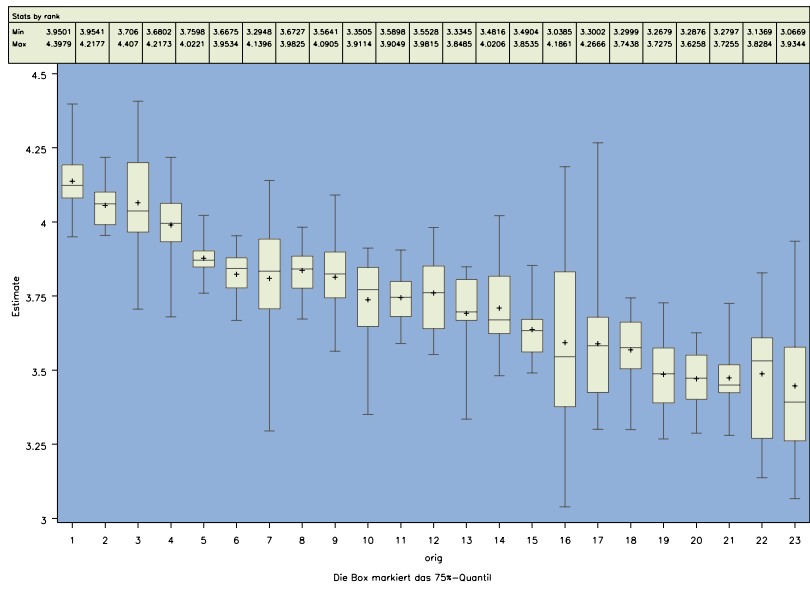
Departement 1, Least-Squares Means Boxplot



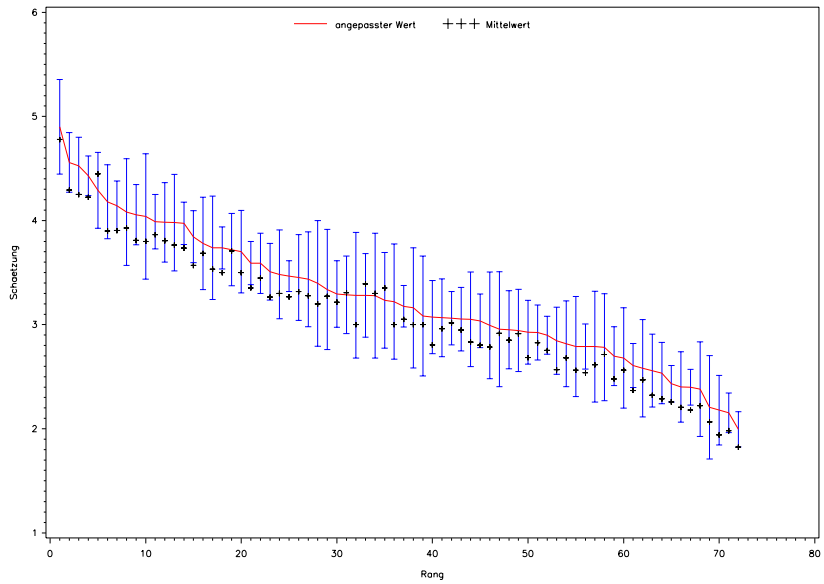
Departement 2, Least-Squares Means



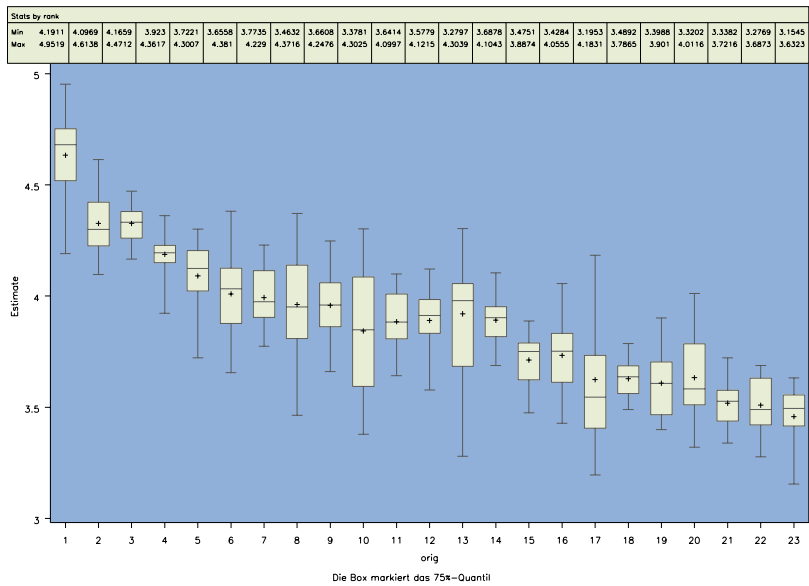
Departement 2, Least-Squares Means Boxplot



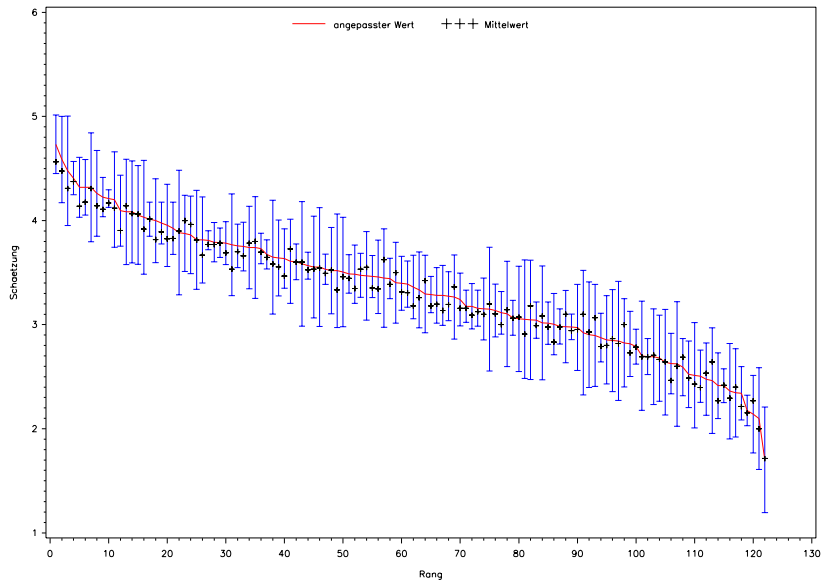
Departement 3, Least-Squares Means



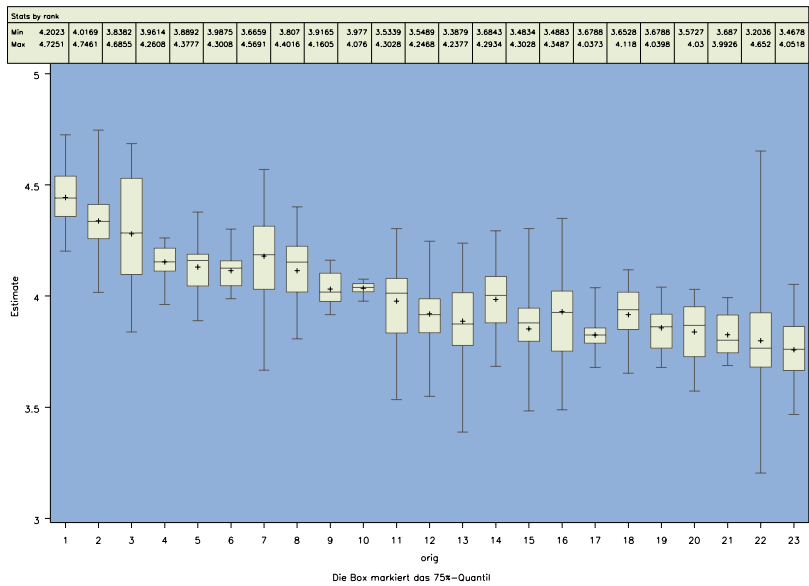
Departement 3, Least-Squares Means Boxplot



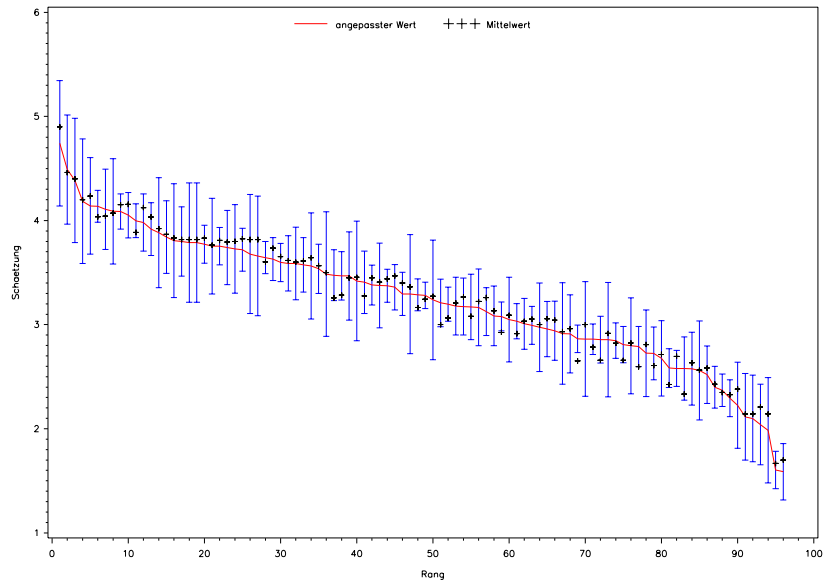
Departement 4, Least-Squares Means



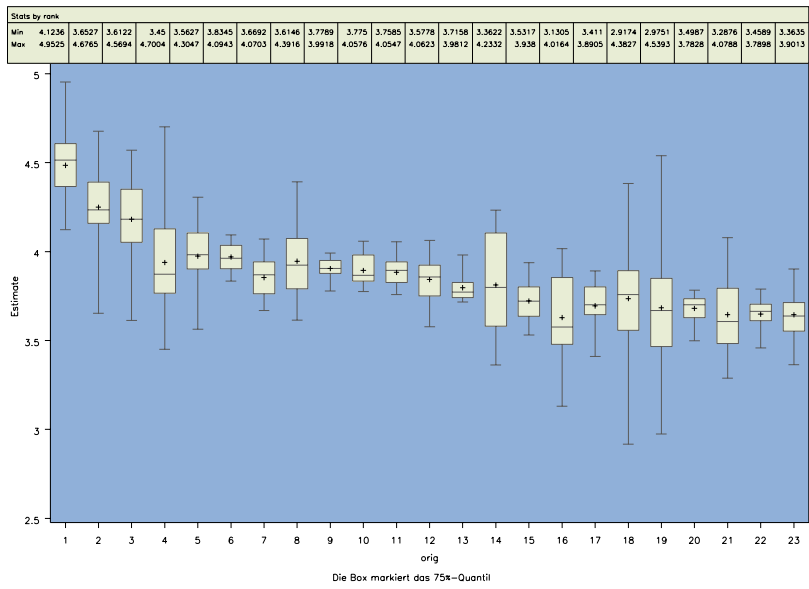
Departement 4, Least-Squares Means Boxplot



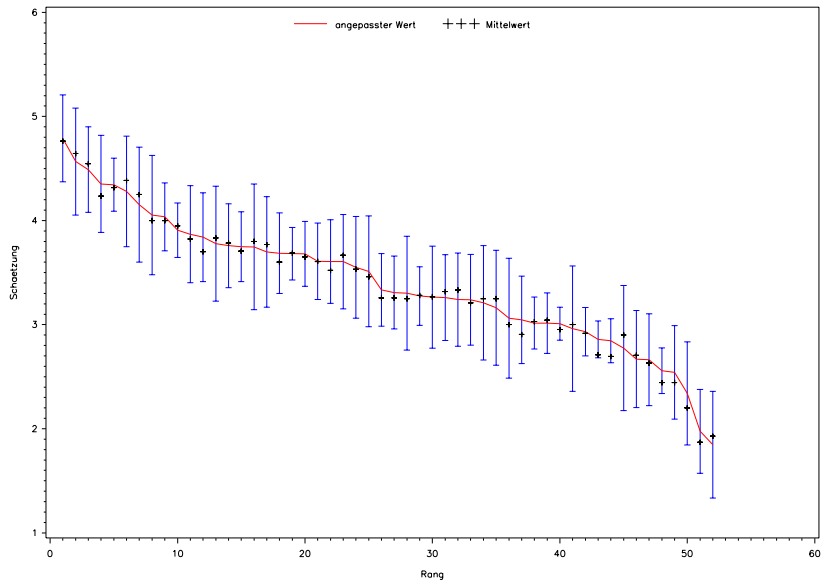
Departement 5, Least-Squares Means



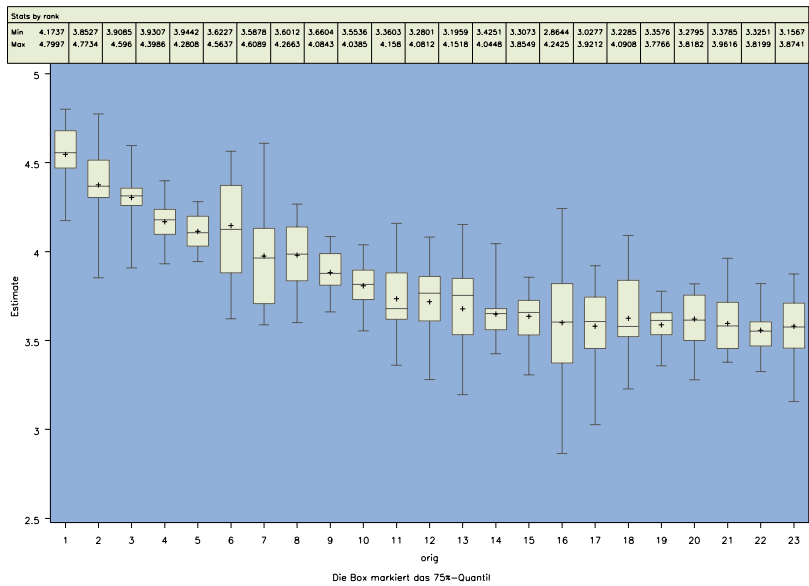
Departement 5, Least-Squares Means Boxplot



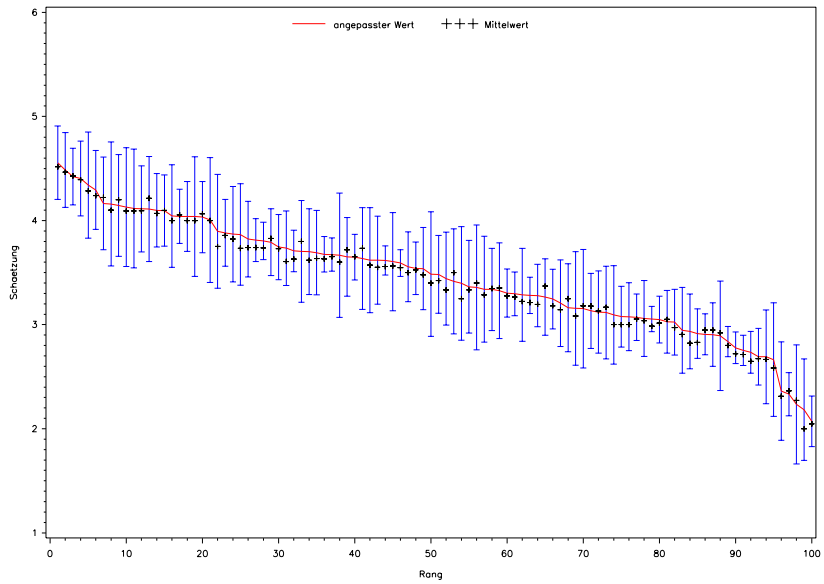
Departement 6, Least-Squares Means



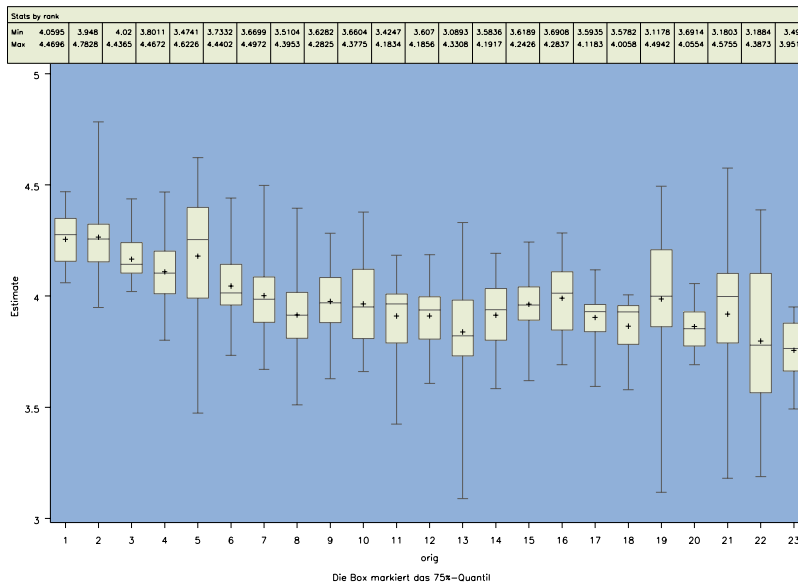
Departement 6, Least-Squares Means Boxplot



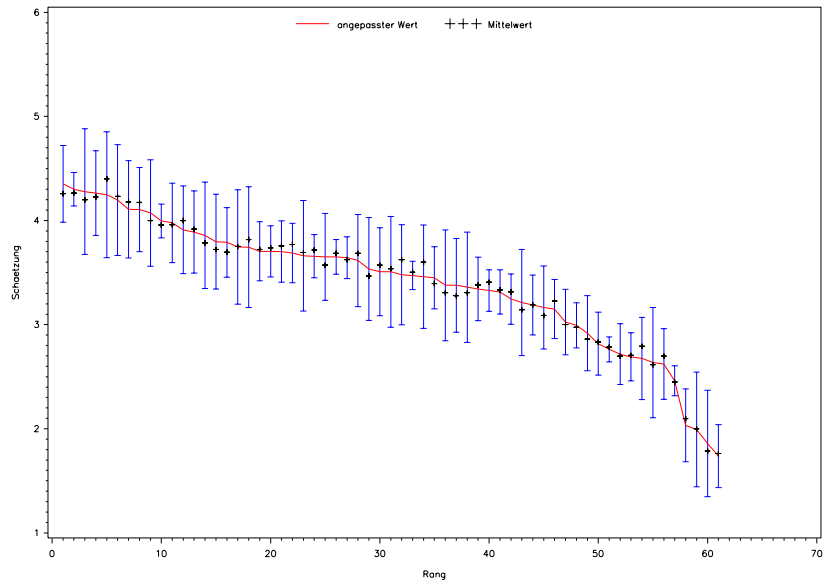
Departement 7, Least-Squares Means



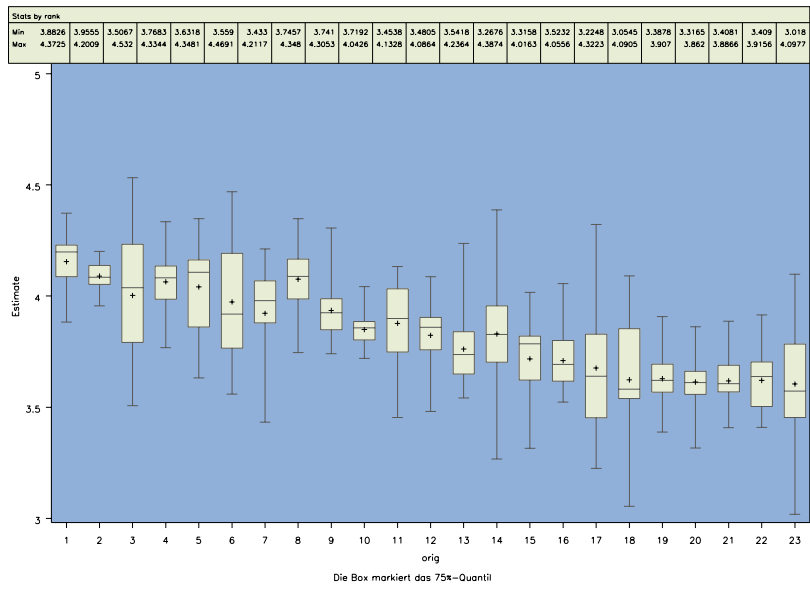
Departement 7, Least-Squares Means Boxplot



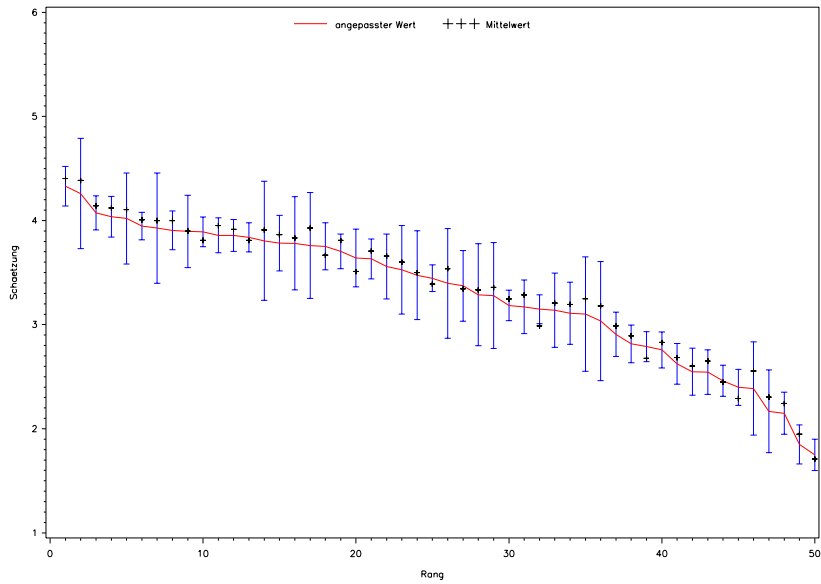
Departement 8, Least-Squares Means



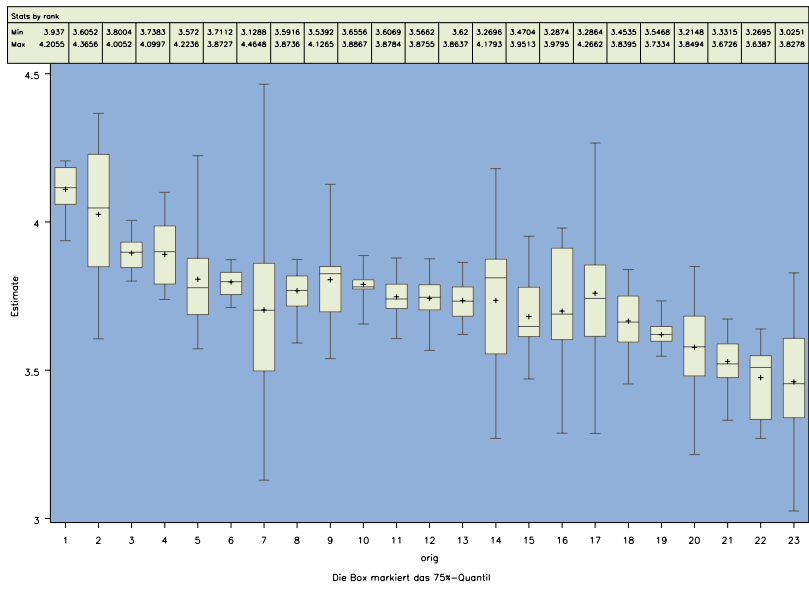
Departement 8, Least-Squares Means Boxplot



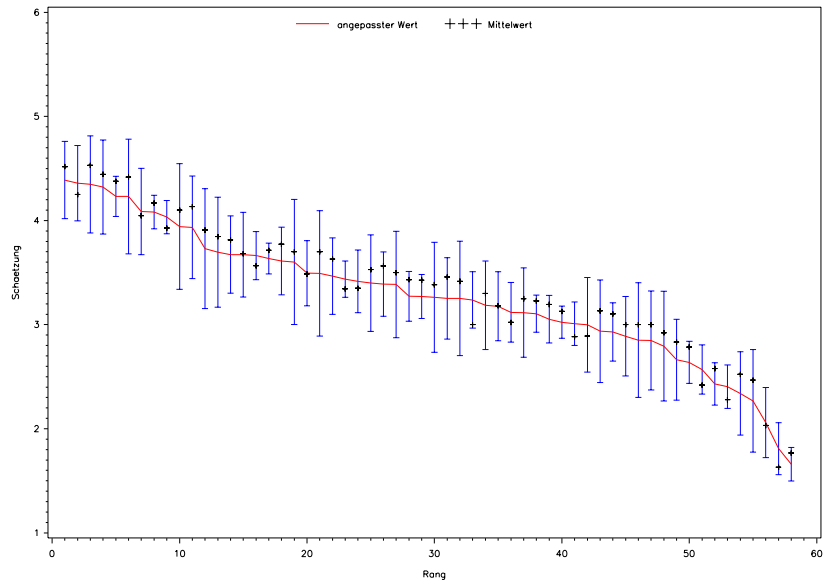
Departement 9, Least-Squares Means



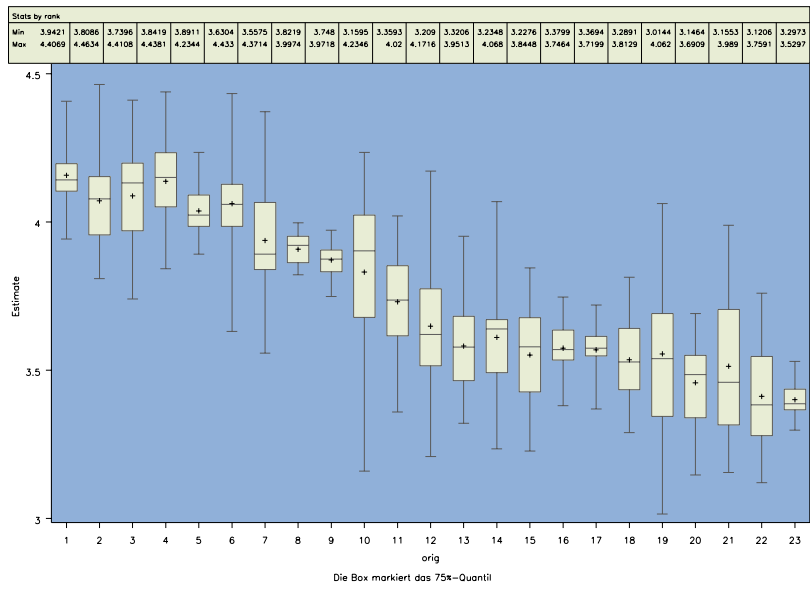
Departement 9, Least-Squares Means Boxplot



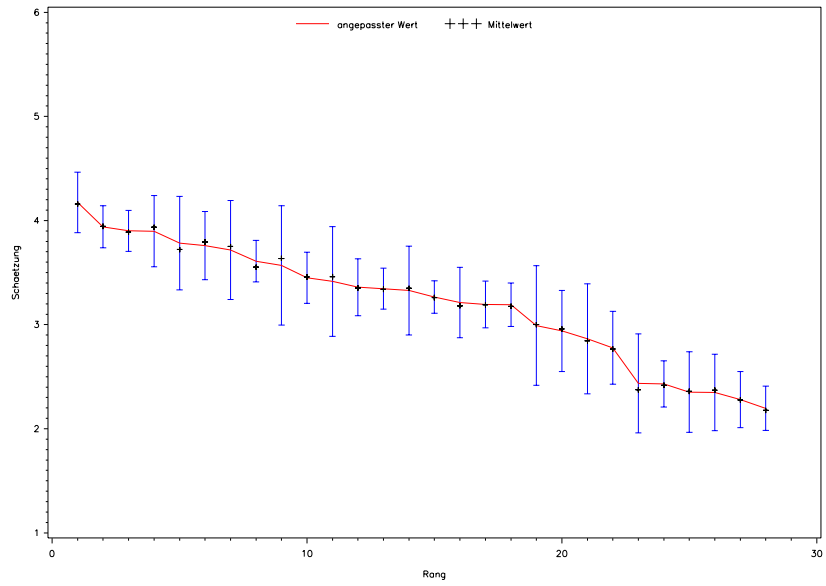
Departement 10, Least-Squares Means



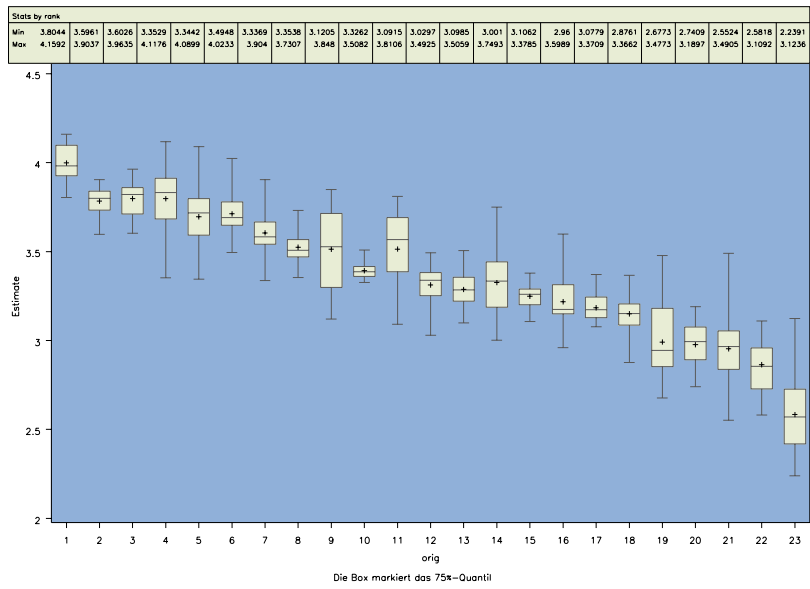
Departement 10, Least-Squares Means Boxplot



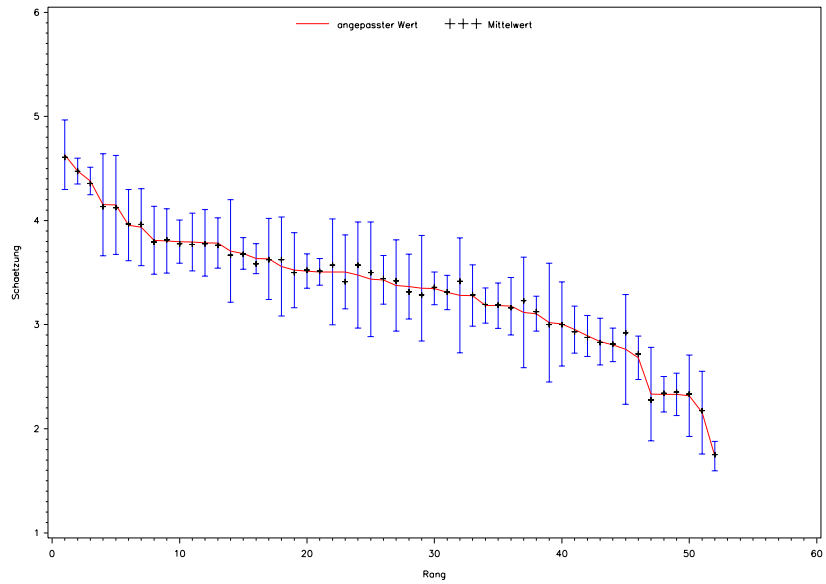
Departement 11, Least-Squares Means



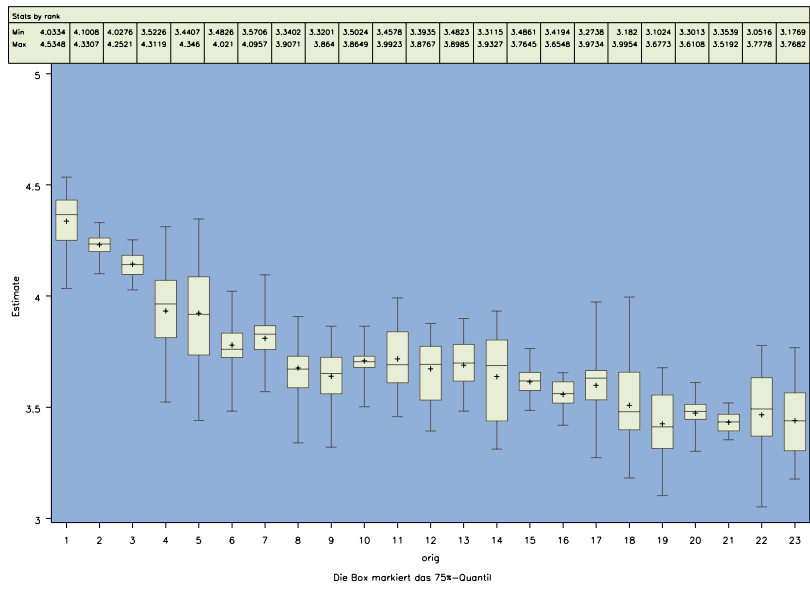
Departement 11, Least-Squares Means Boxplot



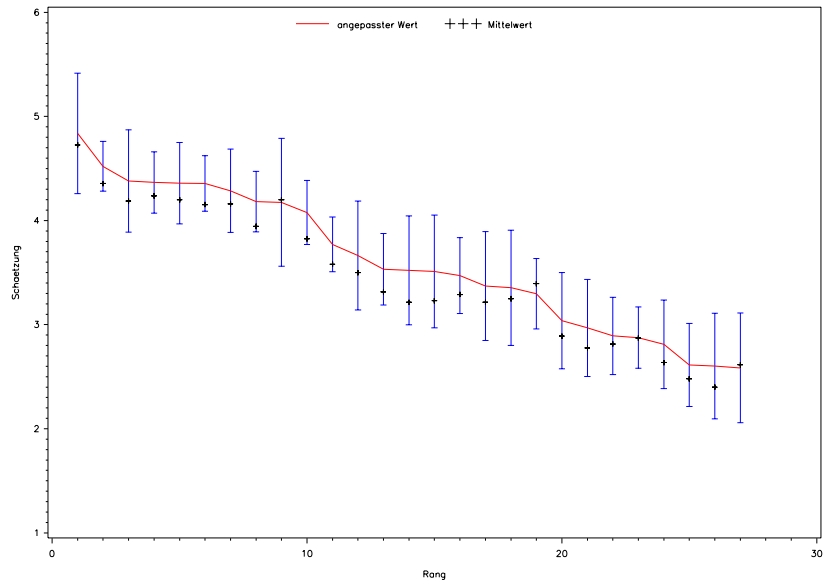
Departement 12, Least-Squares Means



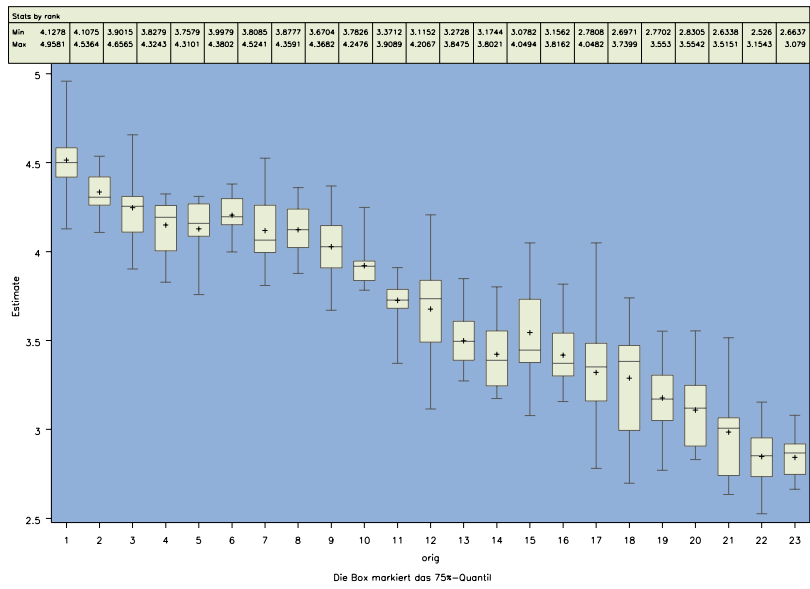
Departement 12, Least-Squares Means Boxplot



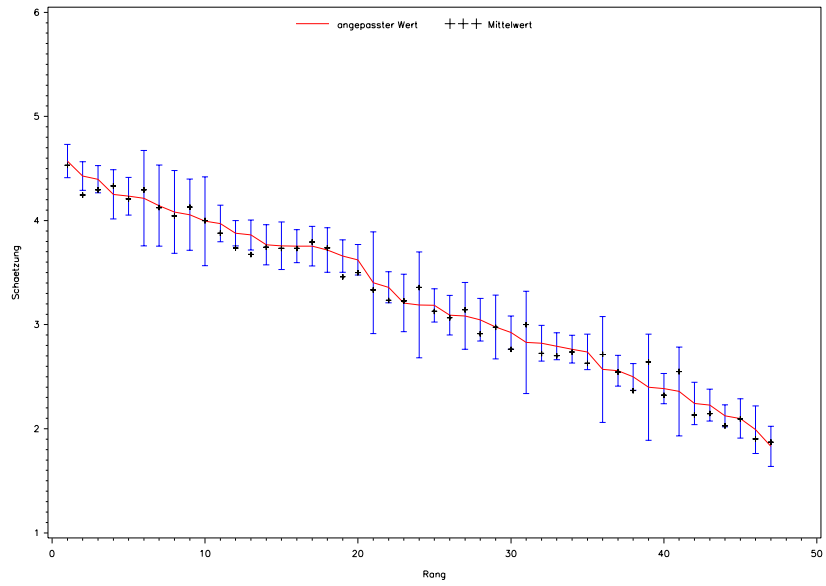
Departement 13, Least-Squares Means



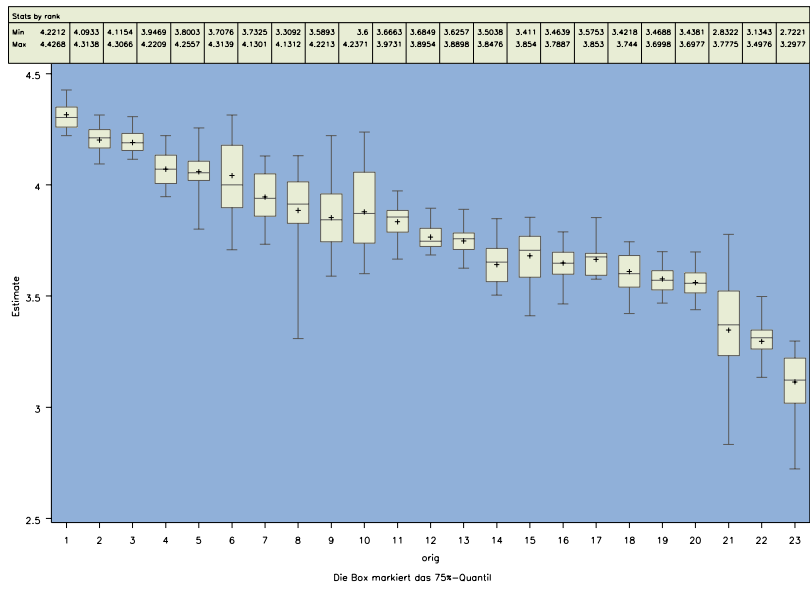
Departement 13, Least-Squares Means Boxplot



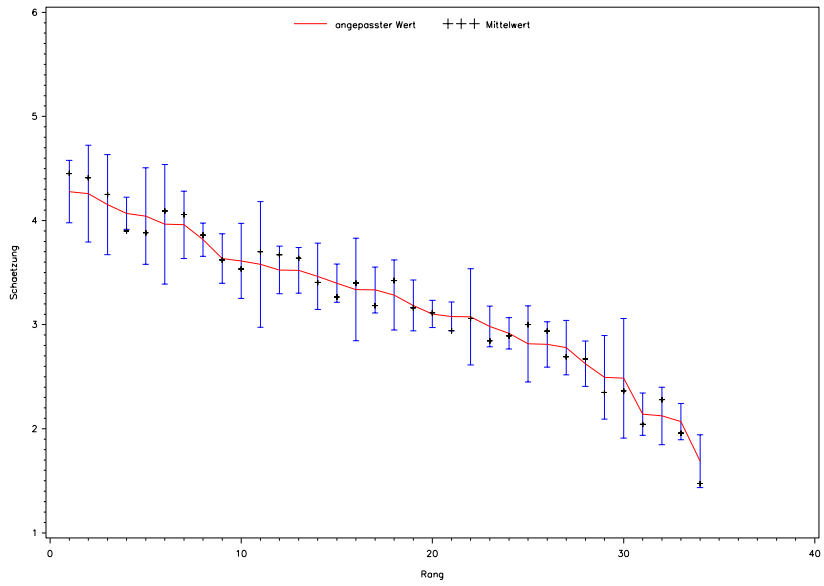
Departement 14, Least-Squares Means



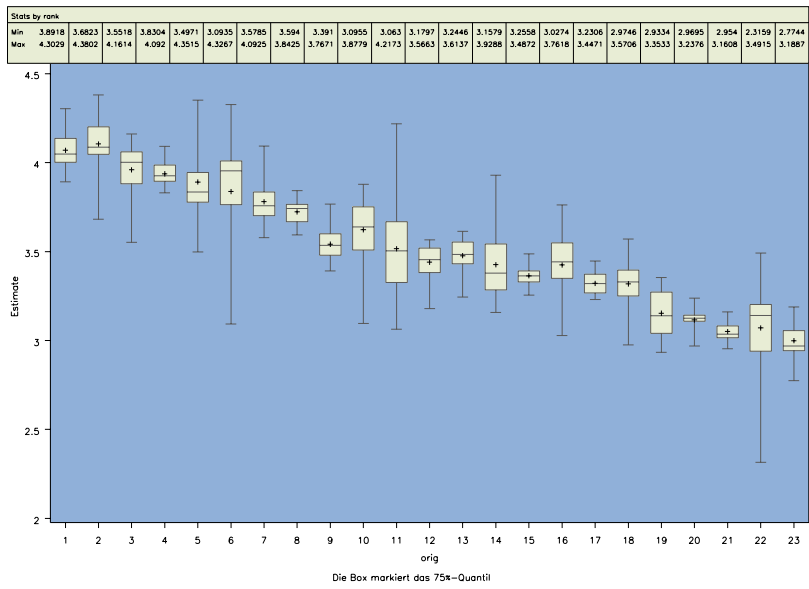
Departement 14, Least-Squares Means Boxplot



Departement 15, Least-Squares Means



Departement 15, Least-Squares Means Boxplot



# Literaturverzeichnis

- [1] Geert Verbeke, Geert Molenberghs (1997), *Linear Mixed Models in Practice*, New York: Springer.
- [2] Fan, Xitao, Ákos Felsővályi, Stephen A.Sivo, and Sean C. Keenan (2002), *SAS<sup>®</sup> for Monte Carlo Studies: A guide for Quantitative Researchers*, Cary, NC: SAS Institute Inc.
- [3] SAS Institute (2004) *SAS/STAT 9.1 User's Guide, Volumes 1-7*, SAS Institute Inc.
- [4] Jonas Andersson, Kenneth Carling and Stefan Mattson (1998), "Random Ranking of Hospitals Is Unsound", *Chance*, Vol. 11, NO. 3, 34-39.
- [5] NIST/SEMATECH e-Handbook of Statistical Methods, <http://www.itl.nist.gov/div898/handbook/>, 11. September 2006.
- [6] Carlin, B.P. and Louis, T.A. (1996), *Bayes and Empirical Bayes Methods for Data Analysis*, London: Chapman and Hall.
- [7] Rudolf J. Freund, William J. Wilson (2002), *Statistical Methods, Second Edition*, Academic Press.