

Nr. 1/2018

Arbeitszeiten von Professorinnen und Professoren in Deutschland 2016

Claus Weihs, Tanja Hernández Rodríguez,
Maximilian Doeckel, Christoph Marty, Holger
Wormer

Technical Report

Arbeitszeiten von Professorinnen und Professoren in Deutschland 2016

Claus Weihs* Tanja Hernández Rodríguez† Maximilian Doeckel‡
Christoph Marty§ Holger Wormer¶

26. Januar 2018

Zusammenfassung

In dieser Studie werden belastbare Prognoseintervalle der wöchentlichen Gesamtarbeitszeit von Universitätsprofessorinnen und -professoren aus Daten einer Umfrage aus dem Jahre 2016 und a-priori Informationen aus früheren Studien bestimmt. Neben der Gesamtarbeitszeit werden auch Teilarbeitszeiten zum Beispiel für Lehre und Forschung ermittelt. Die Ergebnisse von frequentistischer und Bayesianischer Analyse werden verglichen.

Aus den gültigen Fragebögen von aktiven Vollzeit arbeitenden Universitätsprofessorinnen und -professoren ergeben sich bei der direkten Schätzung $56 h$ für die durchschnittliche wöchentliche Gesamtarbeitszeit und 95 %-Prognoseintervalle von $35 h$ bis $80 h$. Frequentistische und Bayesianische Analyse führen zu ähnlichen Ergebnissen, Fächergruppen und Geschlechter unterscheiden sich wenig.

Wird die Gesamtarbeitszeit als Summe der Arbeitszeiten für Teilaufgaben geschätzt, führt dies zu einem wesentlich größeren Mittelwert von $63 h$ und deutlich unterschiedlichen 95 %-Prognoseintervallen im Bayesianischen Fall mit $[42, 85] h$ und im frequentistischen Fall mit $[28, 113] h$. Messungen für die Gesamtarbeitszeit aus unabhängig voneinander ermittelten Teilarbeitszeiten erscheinen deshalb nur verlässlich, wenn eine Bayesianische Analyse mit Vorinformationen über die Gesamtarbeitszeit durchgeführt wird, denn offenbar sind Summen von Teilarbeitszeiten tendenziell größer als eine Gesamtarbeitszeitschätzung, sowohl im Mittel als auch in der Variation. Ein möglicher Grund ist die fehlende Übersicht über die insgesamt angegebene Arbeitszeit, wenn kein Summenzähler während des Ausfüllens des Fragebogens mitgeführt wird.

Der Anteil forschungsnaher Tätigkeiten an der Arbeitszeit erscheint mit etwa 60 % deutlich höher als der Anteil von Lehre und Betreuung und Prüfung von Studierenden mit 23 % und der Anteil administrativer Tätigkeiten mit 17 %.

Die größten signifikanten Differenzen in den Erwartungswerten der Fächergruppen treten immer zwischen den *Geistes-/Sozialwissenschaften* und einer der anderen Fächergruppen auf, sowohl bei der Gesamtarbeitszeit als auch bei Teilarbeitszeiten. Der Unterschied zwischen dem erwarteten Gesamtarbeitsaufwand von Professorinnen und Professoren ist eher klein.

Stichworte: Gesamt- und Teilarbeitszeiten, Universitätsprofessoren und -professorinnen, frequentistisch und Bayesianisch, Prognoseintervall

* (gleichberechtigter Erstautor), Fakultät Statistik, TU Dortmund, email: claus.weihs@tu-dortmund.de

† (gleichberechtigte Erstautorin), Fachbereich Life Science Technologies, Hochschule Ostwestfalen-Lippe, email: tanja.hernandez@hs-owl.de

‡ Institut für Journalistik, TU Dortmund, email: maximilian.doeckel@tu-dortmund.de

§ Institut für Journalistik, TU Dortmund, email: christoph.marty@tu-dortmund.de

¶ Institut für Journalistik, TU Dortmund, email: holger.wormer@tu-dortmund.de

Workload of German Professors in 2016

Abstract In this study, we determine reliable prediction intervals for the weekly total workload of active German full-time university professors from a 2016 survey and prior information from earlier studies. Additionally, also workloads for subtasks are determined. The results of frequentist and Bayesian analysis are compared.

From the valid questionnaires, a mean of 56 h and 95 %-prediction intervals from 35 h to 80 h arise as direct estimates of the weekly total workload. Frequentist and Bayesian analysis lead to similar results, subject groups and sexes differ only slightly.

Total workload estimated as the sum of workloads of subtasks reaches a significantly higher mean of 63 h and distinctly different 95 %-prediction intervals in the Bayesian case with [42, 85] h and in the frequentist case with [28, 113] h. Therefore, measurements of the total workload from independently determined workloads of subtasks only appear to be reliable if a Bayesian analysis with prior information on the total workload is carried out since sums of independent parts of the workload appear to be bigger than direct estimates of the total workload, in the mean as well as in variation. A possible reason is the missing overview over the total workload already specified when no counter is shown during completion of the questionnaire.

The share of research related activities of the total workload is with approximately 60 % distinctly higher than the share of teaching and student examination with 23 % and the share of administration with 17 %.

The greatest significant differences of subject groups appear between *Humanities and Social Sciences* and one of the other subject groups, for the total workload as well as for workloads for subtasks. The difference between the mean total workload of female and male professors appears to be small.

Keywords: total workload, workload for subtasks, university professors, frequentist and Bayesian, prediction interval

1 Einleitung

Der Forschungs- und Innovationsstandort Deutschland ist leistungsfähiger denn je. So heißt es auf einer Internetseite der Deutschen Bundesregierung (vgl. Bundesregierung (2014)). Forschung und Lehre haben in Deutschland einen hohen Stellenwert und die Hauptbeauftragten, um exzellente Lehre zu vermitteln, Forschung zu betreiben und zu koordinieren und wissenschaftlichen Nachwuchs zu fördern sind Deutschlands Professorinnen und Professoren. Die Anforderungen an den Professorenberuf sind hoch und die Aufgabenbereiche vielfältig. Einerseits garantiert der Gesetzgeber im Grundgesetz, Artikel 5, Absatz 3 (vgl. Deutscher Bundestag (2014)): *Kunst und Wissenschaft, Forschung und Lehre sind frei*, andererseits ist der Katalog mit den Tätigkeitsfeldern des Professorenberufs umfangreich.

Ein wesentlicher Teil der Aufgaben besteht in der Lehre. Dabei stieg die Zahl der Studierenden an deutschen Universitäten zwischen 2000 und 2015 um fast die Hälfte auf 2,76 Millionen (Statistisches Bundesamt (2017a)). Gründe für diese Zunahme sind die demografische Entwicklung, die steigende Bildungsbeteiligung sowie die doppelten Abiturjahrgänge. Um weiterhin allen Studierwilligen die Chance auf ein qualitativ hochwertiges Hochschulstudium zu eröffnen, schufen Bund und Länder mit ihrem bereits zweimal erneuerten Hochschulpakt zusätzliche Ausbildungskapazitäten (Hochschulpakt (2017)). Die auf zeitlich befristeten Bund-Länder-Vereinbarungen basierenden Hochschulpakete zielen u.a. auf die Einstellung zusätzlichen Personals sowie die Erhöhung des Frauenanteils bei der Besetzung von Professuren. Insgesamt aber hielt die Zahl neu geschaffener Professuren nicht Schritt mit der gestiegenen Zahl der Studierenden, sondern nahm von 2000 bis 2015 nur um ein knappes Viertel auf 46.344 Professorinnen

und Professoren zu (Statistisches Bundesamt (2017b); statista (2017)).

Allerdings stellt die Lehre nur eine der wichtigen Anforderungen an Professorinnen und Professoren dar. Neben der Forschung zählen dazu u.a. auch die (akademische Selbst-) Verwaltung und das Einwerben von Drittmitteln. Kann es gelingen, all diesen Anforderungen innerhalb der bei EU-Projekten angenommenen wöchentlichen Arbeitszeit von 41 Stunden (*h*) nachzukommen? Wie viele Stunden arbeiten Deutschlands Professorinnen und Professoren wirklich? Und wie lässt sich ihr Arbeitspensum zuverlässig messen?

Die vorliegende Studie stellt ein Messmodell zur Diskussion, das auf den Selbsteinschätzungen von Professorinnen und Professoren basiert. Grundlage ist eine E-Mail-Befragung, für die in einem Lehrforschungsprojekt an der Technischen Universität Dortmund gemeinsam mit ZEIT ONLINE ein Fragebogen entwickelt worden ist. Um diesen Fragebogen zu testen und erste Tendenzen zu erkennen, wurde im Rahmen eines Fallstudienprojekts mit Statistik- und Datenjournalismus-Studierenden der TU Dortmund eine Pilotbefragung vorgenommen. Diese ergab, dass Professorinnen und Professoren im Durchschnitt über das Jahr gesehen weit mehr als 41 *h* wöchentlich arbeiten.

Um diese auf einer sehr kleinen Stichprobe basierenden Tendenzen zu überprüfen, wurde eine Vollerhebung unter den deutschen Professorinnen und Professoren vorgenommen. Grundlage für die Adressliste der E-Mail-Umfrage war *Kürschners Deutscher Gelehrten-Kalender*, der den Anspruch hat, die Kontaktdaten aller deutschen Professorinnen und Professoren zu enthalten. Zunächst wurde die Datenbank am 24.1.2016 komplett ausgelesen, was zu 88.356 Adressen führte. Diese wurden dann nach verschiedenen Kriterien gefiltert. So wurden nur Personen in die Liste aufgenommen, deren Titel ein *Prof* enthielt, *HonPr*, *Akad* oder *TitPr* jedoch nicht, um Honorarprofessoren, Titularprofessoren und wissenschaftliche Mitarbeiter auszuschließen. Weiterhin wurden nur Kontakte in die Liste aufgenommen, deren *Land*-Eintrag *Deutschland* enthielt, die eine E-Mail-Adresse angegeben hatten (bei mehreren E-Mail-Adressen wurde die erste benutzt) und deren E-Mail-Adressen auf *DE*, *de*, *info*, *net*, *org*, *eu*, *edu* oder *com* endeten. Die bereinigte Kontaktliste enthielt 35.726 E-Mail-Adressen. Die erste Aussendung der E-Mails erfolgte am 10.10.2016, Erinnerungsmails am 27.10.2016 und am 16.11.2016. Rund 3500 Adressen stellten sich dabei als ungültig heraus.

Ziel dieser Arbeit ist es, statistische Methoden und ein geeignetes Messmodell zu präsentieren, die es ermöglichen, aktuelle und belastbare Erkenntnisse zur zeitlichen Arbeitsbelastung des Professorenberufs aus den Ergebnissen dieser Umfrage zu gewinnen. Die gewonnenen Erkenntnisse sollen dabei nicht nur die Daten der Neuerhebung berücksichtigen, sondern auch den bis dahin verfügbaren Informationsstand angemessen einbeziehen. Dabei beschränken wir uns auf die Analyse von vier Fächergruppen, denen sich die Teilnehmer und Teilnehmerinnen der Umfrage zugeordnet haben, und von sieben Tätigkeitsfeldern, die sich aus dem Fragebogen ergeben. Außerdem werden die Unterschiede zwischen Professorinnen und Professoren charakterisiert. Weitergehende Analysen, z.B. einzelner Fächer und einzelner Fragen, werden in einer späteren Analyse diskutiert.

Die Arbeit ist so gegliedert, dass nach einer Einführung in die theoretischen Grundlagen und ihrer Umsetzung in Abschnitt 2 eine Beschreibung der Ausgangsdatenlage in Abschnitt 3 erfolgt. Hier werden diejenigen Studien aus der Literatur vorgestellt, die relevante Informationen zur Arbeitsbelastung von Professoren liefern und in dieser Arbeit verwendet werden. Abschnitt 4 nimmt Bezug auf die Fragen unseres Fragebogens und stellt ein Messmodell für den zeitlichen Gesamtaufwand auf der Basis von Teilaufwänden vor. In Abschnitt 5 wird erläutert, wie aus den vorliegenden Ausgangsdaten a-priori Information in Form von a-priori Verteilungen ermittelt wird. In Abschnitt 6 werden die Ergebnisse der Auswertung des Fragebogens vorgestellt und in Abschnitt 7 werden diese Ergebnisse diskutiert und Schlussfolgerungen präsentiert.

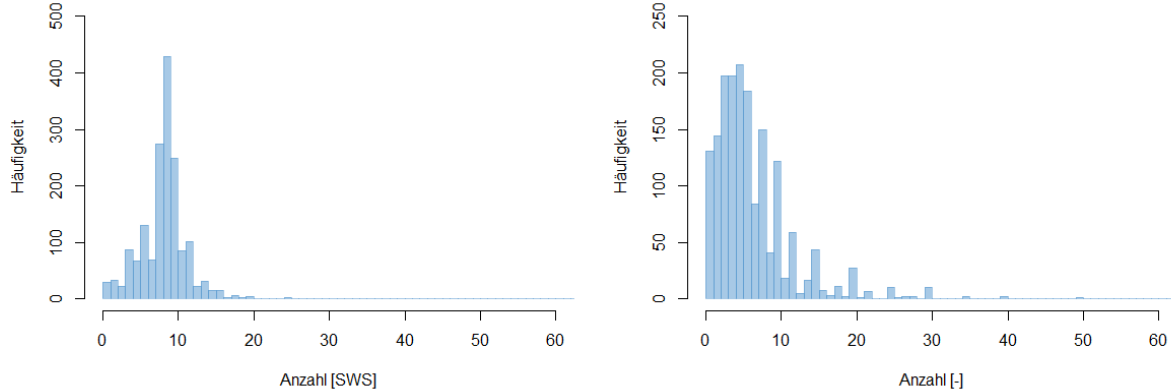


Abb. 1: Histogramme typischer Verteilungen: X_2 links, X_{11} rechts (Bezeichnungen s. Abschnitt 4).

2 Statistische Methoden

In dieser Arbeit werden frequentistische und Bayes-Modelle zur Bestimmung von Prognoseintervallen und zum Vergleich von Erwartungswerten gegenübergestellt. Sei dazu Y die betrachtete Zufallsvariable (z.B. die wöchentliche Gesamtarbeitszeit) und $y := (y_1 \dots y_n)$ ein Vektor von Realisierungen dieser Zufallsvariable.

2.1 Prognoseintervalle frequentistisch

Zur Schätzung der Verteilungsmitte und eines frequentistischen 95 %-Prognoseintervalls dieser Zufallsvariable werden für die Stichprobe der Mittelwert \bar{y} sowie das 2,5 %-Quantil $y_{0,025}$ und das 97,5 %-Quantil $y_{0,975}$ berechnet. Als Prognoseintervall wird dann $[y_{0,025}, y_{0,975}]$ verwendet. Damit werden auch nicht-symmetrische Verteilungen zugelassen, passend zu den Verteilungen der hier beteiligten Variablen (vgl. Abb. 1).

2.2 Prognoseintervalle Bayesianisch

2.2.1 Theorie

Im Bayes-Modell basieren die Prognoseintervalle nicht ausschließlich auf den Angaben der Professorinnen und Professoren, die an der E-Mail-Umfrage teilgenommen haben, sondern auch auf vorher schon gewonnenen Erkenntnissen, wie sie in Abschnitt 3 vorgestellt werden. Wir interessieren uns für die Verteilung von Prognosen für nicht beobachtete Werte \tilde{y} der Zielvariable, gegeben die beobachteten Daten y . Dazu verwenden wir die so genannte prädiktive a-posteriori Dichte $f(\tilde{y}|y)$ (vgl. Spiegelhalter et al. (2004)). Für deren Berechnung benötigt man die Festlegung eines statistischen Modells für die Abhängigkeit der Zielvariable Y von Modellparametern θ in Form einer Datendichte (Likelihood) $f(y|\theta)$, die Festlegung des a-priori Wissens über die Modellparameter θ in Form einer a-priori Dichte $f(\theta)$ und die Berechnung der a-posteriori Dichte $f(\theta|y)$ der Modellparameter mit Hilfe des Bayes-Theorems. Sowohl y als auch θ können mehrdimensional sein. Aus der prädiktiven a-posteriori Verteilung kann dann ein $100(1 - \beta)$ %-Prognoseintervall $S_{1-\beta}$ für die Zielvariable ermittelt werden (vgl. Bernardo und Smith

(2000), S. 259), so dass

$$\int_{S_{1-\beta}} f(\tilde{y}|y) d\tilde{y} = 1 - \beta. \quad (1)$$

Es gibt verschiedene Arten solcher so genannten Kreditabilitätsintervalle S . Wir verwenden die *highest posteriori density (hpd)*-Intervalle, die alle Werte des Parameters \tilde{y} enthalten, für die die a-posteriori Dichte $f(\tilde{y}|y)$ größer als eine Konstante c_β ist, wobei c_β so gewählt wird, dass die Überdeckungswahrscheinlichkeit $(1 - \beta)$ beträgt. *hpd*-Intervalle eignen sich insbesondere für nicht-symmetrische a-posteriori Dichten.

2.2.2 Umsetzung

Für den Zeitaufwand Y für eine bestimmte Tätigkeit gehen wir in dieser Studie davon aus, dass die Daten einer Gamma-Verteilung mit den unbekanntenen Parametern α und λ folgen, d.h.

$$Y|\alpha, \lambda \sim \Gamma(\alpha, \lambda).$$

Diese Annahme stützt sich zum einen auf die Tatsache, dass die betrachteten Zufallsvariablen nur positive Werte annehmen können. Zum anderen hat die Pilotbefragung gezeigt, dass sich die Gamma-Verteilung in vielen Fällen zur Beschreibung der Verteilungen der betrachteten Zufallsvariablen eignet. Die Gammaverteilung ist sehr flexibel und wird durch die Parameter α (shape) und λ (scale) festgelegt. Die Dichtefunktion lautet:

$$f(y|\alpha, \lambda) = \frac{\lambda^\alpha}{\Gamma(\alpha)} y^{\alpha-1} e^{-\lambda y} \quad \text{für } y > 0 \quad \text{und} \quad = 0 \quad \text{sonst.} \quad (2)$$

Erwartungswert und Varianz berechnen sich als $E(Y) = \frac{\alpha}{\lambda}$, $\text{Var}(Y) = \frac{\alpha}{\lambda^2}$.

Da die Parameter der Verteilung unbekannt sind, nehmen wir a-priori je eine Verteilung $f(\alpha)$ und $f(\lambda)$ für beide Parameter an. Daraus berechnet sich die prädiktive a-posteriori Dichte einer neuen Beobachtung \tilde{y} gegeben die Daten y mittels

$$f(\tilde{y}|y) = \int_0^\infty \int_0^\infty f(\tilde{y}|\alpha, \lambda) f(\alpha, \lambda|y) d\alpha d\lambda. \quad (3)$$

Da sich dieses Integral nicht analytisch lösen lässt, wird die prädiktive a-posteriori Verteilung neuer Beobachtungen per Simulation wie folgt bestimmt:

1. Ziehe α_1 aus $f(\alpha_1|y)$
2. Ziehe λ_1 aus $f(\lambda_1|\alpha_1, y)$
3. Ziehe \tilde{y} aus der Likelihood $f(\tilde{y}|\alpha_1, \lambda_1)$ (s. Gleichung (2))
4. Wiederhole Schritte 1 bis 3 m Mal. Die m Werte $\tilde{y}_1, \dots, \tilde{y}_m$ repräsentieren Züge aus der prädiktiven a-posteriori Verteilung $f(\tilde{y}|y)$.
5. Zur Bestimmung des *hpd*-Kreditabilitätsintervalls für \tilde{y} ordne $\tilde{y}_1, \dots, \tilde{y}_m$ der Größe nach, so dass $\tilde{y}_{(1)} \leq \dots \leq \tilde{y}_{(m)}$. Konstruiere dann alle $100(1 - \beta)\%$ -Kreditabilitätsintervalle von \tilde{y} , also

$$(\tilde{y}_{(1)}, \tilde{y}_{\lfloor m(1-\beta) \rfloor}), \dots, (\tilde{y}_{\lfloor m\beta \rfloor}, \tilde{y}_{(m)}).$$

Das *hpd*-Kreditabilitätsintervall für \tilde{y} ist das Kreditabilitätsintervall mit der kürzesten Länge.

Alle Berechnungen wurden in der Programmiersprache R (vgl. R (2016)) implementiert, wobei die Umsetzung der Schritte 1 bis 4 mittels Markov Chain Monte Carlo (MCMC)-Simulation unter Verwendung von R und BRugs (vgl. Thomas et al. (2006)), einem R-Interface für OpenBUGS (vgl. Kruschke (2010)), erfolgte. Dabei wird der Gibbs-Sampler zur Simulation der a-posteriori Verteilung sowie der prädiktiven a-posteriori Verteilung angewendet. Es wurden jeweils 3 Markov-Ketten simuliert. Dabei werden nur Realisierungen akzeptiert, deren Markov-Ketten konvergiert haben und eine geringe Autokorrelation aufweisen. Um zu entscheiden, wie viele Iterationen im Anschluss an die Burn-in Phase durchgeführt werden müssen, so dass Signifikanzaussagen auf dem 5 %-Niveau zu den Ergebnissen möglich sind, wird der Monte Carlo-Fehler (*MC error*) herangezogen, ein Schätzer des Monte Carlo-Standardfehlers $\frac{s}{\sqrt{N}}$ des betrachteten Mittelwerts. Um s zu schätzen, wird die *batch means Methode* von Roberts (1996) verwendet. Die Simulationen wurden solange durchgeführt, bis der MC error $< 5\%$ war.

2.3 Beurteilung von Erwartungswertsdifferenzen

Um zu untersuchen, ob unterschiedliche Fächergruppen oder Geschlechter im Mittel unterschiedliche Wochenarbeitszeiten aufweisen, werden die Mittelwerte dieser Fächergruppen verglichen. Mittelwerte großer Stichproben sind nach dem zentralen Grenzwertsatz nahezu normalverteilt. Differenzen von Mittelwerten zweier unabhängiger Stichproben der Größen n, m der Zufallsvariablen X, Y sind approximativ $N(\mu_X - \mu_Y, \sigma_X^2/n + \sigma_Y^2/m)$ verteilt. Wegen der Symmetrie der Normalverteilung verwenden wir im frequentistischen Modell das Konfidenzintervall

$$\left[(\bar{x} - \bar{y}) - 1,96 \cdot \sqrt{\sigma_X^2/n + \sigma_Y^2/m}, (\bar{x} - \bar{y}) + 1,96 \cdot \sqrt{\sigma_X^2/n + \sigma_Y^2/m} \right]$$

als Unsicherheitsintervall für die Erwartungswertdifferenz. Falls dieses Intervall die Null nicht enthält, schließen wir auf einen (auf dem 5 %-Niveau) signifikanten Unterschied der Erwartungswerte.

Im Bayesianischen Modell wird das analoge Intervall der prädiktiven a-posteriori Verteilung verwendet.

Zum Vergleich von k Fächergruppen unter Berücksichtigung des multiplen Testens wird das schrittweise Verfahren von Ryan verwendet (cp. Lüpsen (2016)). Dabei werden die geschätzten Mittelwerte der Größe nach geordnet und zunächst nur die größte Differenz auf Signifikanz getestet. Wenn diese signifikant ist, wird der größte bzw. der kleinste Mittelwert weggelassen und in den daraus entstehenden beiden Gruppen von Fächergruppen die größten Differenzen getestet. Usw. bis keine signifikanten Unterschiede mehr gefunden werden. Ab der Betrachtung von weniger als $k-1$ Mittelwerten wird das Signifikanzniveau wie folgt angepasst: $\alpha_{k-i} = 1 - (1 - \alpha)^{\frac{k-i}{k}}$, $i = 2, 3, \dots$. Falls eine Differenz nicht signifikant von Null verschieden ist, wird die entsprechende Gruppe von Fächergruppen als homogen angesehen und die Erwartungswerte der enthaltenen Fächergruppen als nicht unterschiedlich.

3 Ausgangsdatenlage

In diesem Abschnitt werden alle die von uns verwendeten Studien vorgestellt, die a-priori Informationen zu Professorenarbeitszeiten angeben. Wie diese Studien konkret bei der Bestimmung von a-priori Verteilungen verwendet werden, zeigt Abschnitt 5.

3.1 A-priori Wissen

3.1.1 Statistisches Bundesamt (DESTATIS)

Auf der Internetseite des Statistischen Bundesamtes sind unter der Rubrik *Bildung, Forschung, Kultur* zahlreiche Veröffentlichungen zum Thema Hochschule zu finden (vgl. Statistisches Bundesamt (2017)). Die Publikation *Promovierende in Deutschland* (vgl. Statistisches Bundesamt (2016a)) liefert Daten zur Betreuungsrelation (Promovierende) im Wintersemester 2014/2015, die in unserer Studie verwendet wurden: Demnach gab es über alle Fächergruppen gemittelt 6 Promovierende je Professur, wobei etwa 50 % der Professorinnen und Professoren zwischen einem und fünf Promovierenden betreute. Dabei muss berücksichtigt werden, dass in manchen Bereichen nicht alle Professorinnen und Professoren ein Promotionsrecht besitzen, um Promovierende betreuen zu können. Daher wird die Betreuungsrelation bei Professorinnen und Professoren mit Promotionsrecht etwas über dem errechneten Durchschnitt liegen.

3.1.2 LESSI-Studie (INCHER)

In der Studie "Wandel von Lehre und Studium an deutschen Hochschulen - Erfahrungen und Sichtweisen der Lehrenden (LESSI)", die das International Centre for Higher Education Research der Universität Kassel (INCHER-Kassel) im Auftrag des Projekts nexus der Hochschulrektorenkonferenz durchgeführt hat (vgl. Schomburg et al. (2012)), wurden die teilnehmenden Professorinnen und Professoren nach der wöchentlichen Arbeitszeit gefragt, wobei nach Vorlesungszeit und vorlesungsfreier Zeit unterschieden wurde. In der Vorlesungszeit lag die angegebene Wochenarbeitszeit bei Universitätsprofessorinnen und Professoren im Mittel bei 56,5 h und bei FH-Professorinnen- und Professoren bei 50,6 h ($n = 5.622$). In der vorlesungsfreien Zeit lag die angegebene Wochenarbeitszeit bei Universitätsprofessorinnen und Professoren im Mittel bei 51,9 h und bei FH-Professorinnen und Professoren bei 40,3 h ($n = 5.491$). Des Weiteren werden in der Studie die vertraglich festgelegten und die tatsächlichen (von den Professoren angegebenen) Lehrverpflichtungen in Semesterwochenstunden (*SWs*) angegeben. Diese unterscheiden sich stark nach Statusgruppe und Qualifikationsstufe.

Neben den vertraglich festgelegten Lehrverpflichtungen wurde auch der tatsächliche Umfang der Lehre erhoben. Für die Universitätsprofessorinnen und Professoren mit Lehrtätigkeit ergab sich für die vertragliche Lehrverpflichtung im Mittel 8 *SWs* und für den tatsächlichen Zeitumfang 9 *SWs*.

3.1.3 iFQ-Wissenschaftlerbefragung (2010)

In einer Wissenschaftlerbefragung vom Institut für Forschungsinformation und Qualitätssicherung (iFQ) wurden im April und Mai 2010 Professorinnen und Professoren deutscher Universitäten zum Thema Forschungsbedingungen befragt (vgl. Böhmer et al. (2011)). Die Auswertungen basieren auf den Angaben von 3131 Professorinnen und Professoren (nach Bereinigung). Untersucht wurden unter anderem Antragsaktivität und Antragserfolg sowie das Tätigkeitsspektrum des Professorenberufs. Neben anderen Aspekten wurden die Teilnehmer danach gefragt, welchen Anteil ihrer Arbeit sie im Jahresdurchschnitt für die verschiedenen Tätigkeiten verwenden. Es zeigte sich, dass der größte Anteil auf den Bereich *Lehr- und Prüfungsverpflichtung* (26,1 %) fiel, gefolgt von *Forschung* (21,4 %), *Betreuung von Studierenden und Promovierenden* (13,9 %), *Gremien/Selbstverwaltung* (10,2 %), *Drittmittelakquise* (8,9 %), *Begutachtungen* (8,8 %), *Management* (8,6 %) und *Sonstiges* (1,8 %).

3.1.4 Deutscher Hochschulverband (DHV)

In einer Online-Umfrage vom 15.10. bis 15.12.2007 hat der deutsche Hochschulverband (DHV) nach der geschätzten Arbeitszeit pro Woche gefragt (vgl. Hartmer (2008)). Die Frage wurde von 3000 Hoch-

schullehrern beantwortet. Das Ergebnis der Selbsteinschätzung wurde in der Zeitschrift *Forschung & Lehre 02/2008* veröffentlicht. Demnach gaben 0,38% der Befragten einen Schätzwert von bis zu 40 h, 11,1% einen Schätzwert zwischen 41 und 50 h, 42% zwischen 51 und 60 h, 31,7% zwischen 61 und 70 h und 14,82% über 70 h an.

4 Die E-Mail-Umfrage

Der Fragebogen zu unserer Befragung von Professorinnen- und Professoren¹ beinhaltet 31 Fragen, die Aufschluss darüber geben sollen, wie viele Stunden Professorinnen und Professoren in Deutschland aktuell in die verschiedenen Tätigkeitsbereiche investieren und welche tatsächliche zeitliche Arbeitsbelastung daraus resultiert. In die Entwicklung des Fragebogens flossen die Erfahrungen aus der Pilotbefragung sowie Résumés diverser fachlicher Diskussionen ein.

4.1 Schätzung der Gesamtarbeitszeit

Zunächst wird in dem Fragebogen nach der Einschätzung der Höhe der wöchentlichen Gesamtarbeitszeit gefragt. Diese Gesamtschätzung Y_{Gesamt} wird auf zwei Arten verwendet, zum einen zum Vergleich mit der aus den Teilarbeitszeiten in Abschnitt 4.2.8 berechneten Gesamtarbeitszeit Y_{Summe} , zum anderen als a-priori Information für die Gesamtarbeitszeit Y_{Summe} (s. Abschnitt 5.3).

4.2 Summe der Einzeltätigkeiten: Das Messmodell

Jeder Frage dieses Fragebogens, die eine quantitative Antwort erfordert, wurde eine Variable zugeordnet. Diese Variablen dienen der Aufstellung der Modellgleichungen zur Ermittlung der wöchentlichen Arbeitszeit. Die Modellgleichungen beinhalten Variablen, welche als Zufallsgrößen in die Schätzung einfließen und feste Faktoren, welche zu Umrechnungszwecken verwendet werden. Die Gesamtarbeitszeit ist in Tätigkeitsbereiche unterteilt, so dass sie aus den Verteilungen einzelner Aufgabenfelder ermittelt werden kann. Die Modellgleichungen sowie die Zuordnung der Fragen zu den Variablen werden nachfolgend erläutert. Wir verwenden folgende Bezeichnungen:

- f_{Index} : Umrechnungsfaktoren auf Wochenstunden. Der Index symbolisiert die Zuordnung.
- X_{Index} : Zufallsvariable zur Beschreibung einer abgefragten Anzahl. Der Index bezieht sich auf die Nummerierung der Fragen im Fragebogen.
- Y_{Index} : Zufallsvariable zur Beschreibung eines abgefragten zeitlichen Aufwands für eine Tätigkeit. Der Index bezieht sich wieder auf die Nummerierung im Fragebogen.

Bei der Bestimmung der Umrechnungsfaktoren gehen wir von 48 Arbeitswochen pro Jahr aus. Für manche Tätigkeitsbereiche ergeben sich während der Vorlesungszeit andere Arbeitszeitbelastungen als in der vorlesungsfreien Zeit. Dabei gehen wir immer von 13 Vorlesungswochen pro Semester aus, d.h. wir nehmen an, dass die Professoren selbst 13 Wochen pro Semester Vorlesungen halten. Bei der Umrechnung ist weiter zu berücksichtigen, dass sich die Fragen auf unterschiedliche Zeiträume beziehen (Semester, 2 Semester, Studienjahr, etc.), um die Fragen einfacher und realistischer beantworten zu können.

¹s. https://www.statistik.tu-dortmund.de/fileadmin/user_upload/PDF-Dokumente_Allgemein/Professorenbefragung-1.pdf

Es wird folgende Unterteilung der Tätigkeitsfelder wie im Fragebogen vorgenommen: i) Lehre & Betreuung von Abschlussarbeiten: Fragen 2-10, ii) Wissenschaftlicher Nachwuchs: Fragen 11-14, iii) Administration: Fragen 16-17, iv) Drittmittel, Gutachten & Herausgeberschaften: Fragen 18-19, 23-25, v) Forschung: Fragen 20-22, vi) Tagungen & Vorträge: Fragen 26-30, vii) Sonstiges: Frage 31.

Im Folgenden entwickeln wir das Modell für die durchschnittliche wöchentliche Arbeitszeit aus den Teilangaben des Fragebogens.

4.2.1 Lehre

Der Bereich Lehre wird durch Fragen nach den gelehrteten *SWs*, den Vor- und Nachbereitungszeiten, den Prüfungen und den betreuten Studien- und Abschlussarbeiten abgedeckt. Für den Bereich Lehre ergibt sich aus dem Fragebogen das folgende Berechnungsmodell:

$$\begin{aligned}
 Y_{Lehre} = & f_{SWs} \cdot X_2 + f_J \cdot X_2 \cdot (Y_{3a}/3 + 2 \cdot Y_{3b}/3) + f_J \cdot Y_4 + f_J \cdot \frac{1}{60} \cdot X_5 \cdot Y_6 \\
 & + f_J \cdot (X_{7a} \cdot Y_{8a} + X_{7b} \cdot Y_{8b} + X_{7c} \cdot Y_{8c} + X_{7d} \cdot Y_{8d} + X_{7e} \cdot Y_{8e}) \\
 & + f_J \cdot (X_{9a} \cdot Y_{10a} + X_{9b} \cdot Y_{10b} + X_{9c} \cdot Y_{10c} + X_{9d} \cdot Y_{10d} + X_{9e} \cdot Y_{10e} \\
 & + X_{9f} \cdot Y_{10f} + X_{9g} \cdot Y_{10g} + X_{9h} \cdot Y_{10h} + X_{9i} \cdot Y_{10i} + X_{9j} \cdot Y_{10j}), \text{ wobei}
 \end{aligned}$$

- X_2 : Anzahl der gelehrteten *SWs* im vergangenen Semester (inkl. Blockveranstaltungen),
- $Y_{3a,3b}$: Zeit in *h* für Konzeption, Vor- und Nachbereitung einer Lehrveranstaltung im Umfang von 2 *SWs* insgesamt pro Semester, für neue (a), bereits gehaltene (b) Lehrveranstaltungen,
- Y_4 : Persönlich investierte Zeit in *h* in Vorbereitung, Durchführung und Korrektur von Klausuren insgesamt im vergangenen gesamten Studienjahr,
- X_5 : Anzahl der Prüflinge (mündlich oder praktisch) im vergangenen Studienjahr (ohne Promotion),
- Y_6 : Persönlich investierter Zeitaufwand in Minuten für die Vorbereitung, Durchführung und Bewertung einer typischen mündlichen oder praktischen Prüfung im vergangenen Studienjahr insgesamt pro Prüfling,
- $X_{7a,7b,7c,7d,7e}$: Anzahl der im vergangenen Studienjahr korrigierten und besprochenen Studienarbeiten (Hausarbeiten (a), Laborberichte (b), Projektarbeiten (c), Referate (d), Sonstiges (e)),
- $Y_{8a,8b,8c,8d,8e}$: Durchschnittlich benötigter Zeitaufwand in *h* für die Korrektur und Besprechung pro Arbeit (Aufteilung wie bei $X_{7a,7b,7c,7d,7e}$),
- $X_{9a,9b,9c,9d,9e}$: Anzahl der in den vergangenen 2 Semestern persönlich betreut und/oder bewerteten Abschlussarbeiten (Bachelor (a), Master (b), Diplom (c), Magister (d), Sonstiges (e)) als Erstgutachter,
- $X_{9f,9g,9h,9i,9j}$: Anzahl der in den vergangenen 2 Semestern persönlich betreut und/oder bewerteten Abschlussarbeiten (Bachelor (f), Master (g), Diplom (h), Magister (i), Sonstiges (j)) als Zweitgutachter,
- $Y_{10a,10b,10c,10d,10e}$: Im Mittel selbst investierte Zeit in *h* in EINE Arbeit (Aufteilung wie bei $X_{9a,9b,9c,9d,9e}$) als Erstgutachter,
- $Y_{10f,10g,10h,10i,10j}$: Im Mittel selbst investierte Zeit in *h* in EINE Arbeit (Aufteilung wie bei $X_{9f,9g,9h,9i,9j}$) als Zweitgutachter,
- $f_{SWs} = 0,75 \cdot 2 \cdot 13 \cdot \frac{1}{48} = 0,75 \cdot \frac{13}{24}$ und $f_J = \frac{1}{48}$.

4.2.2 Wissenschaftlicher Nachwuchs

Für den Bereich Wissenschaftlicher Nachwuchs werden nur die Promovenden berücksichtigt. Habilitanden und Postdoktoranden bleiben unberücksichtigt, da der zeitliche Aufwand für deren Betreuung auf Grund der im hohen Maße selbstständigen und selbstverantwortlichen Arbeitsweise als vernachlässigbar klein angenommen wird. Für den Bereich Wissenschaftlicher Nachwuchs ergibt sich aus dem Fragebogen das folgende Berechnungsmodell:

$$Y_{Wiss.Nachwuchs} = f_{Diss} \cdot (X_{11} \cdot Y_{12a} + X_{11} \cdot Y_{12b}) + \frac{1}{3} \cdot f_J \cdot (X_{13a} \cdot Y_{14a} + X_{13b} \cdot Y_{14b}), \text{ wobei}$$

- X_{11} : Anzahl der im vergangenen Studienjahr betreuten Promotionsvorhaben,
- $Y_{12a,12b}$: Selbst investierte monatliche Zeit in h in die Betreuung EINES Promotionsvorhabens als Erstgutachter (a), Zweitgutachter (b),
- $X_{13a,13b}$: Anzahl der in den vergangenen 3 Studienjahren (2013-2015) bewerteten Promotionsvorhaben (Dissertationen) als Erstgutachter (a), Zweitgutachter (b),
- $Y_{14a,14b}$: Durchschnittliche selbst investierte Zeit in h in die Bewertung EINER Dissertation inkl. Gutachten als Erstgutachter (a), Zweitgutachter (b),
- $f_{Diss} = \frac{1}{4} \cdot \frac{1}{2} = \frac{1}{8}$ (angenommen werden 4 Wochen im Monat sowie gleich viele Erst- wie Zweitbetreuungen).

4.2.3 Administration

Der wöchentliche über das Jahr gemittelte Arbeitsaufwand für Administration wird aus dem entsprechenden wöchentlichen Zeitaufwand in der Vorlesungszeit und in der vorlesungsfreien Zeit durch gewichtete Mittelung berechnet. Für den Bereich Administration ergibt sich aus dem Fragebogen das folgende Berechnungsmodell:

$$Y_{Administration} = f_{Vorl.-Zeit} \cdot Y_{16a} + f_{Vorl.-frei} \cdot Y_{16b} + f_J \cdot Y_{17}, \text{ wobei}$$

- $Y_{16a,16b}$: Durchschnittlicher wöchentlicher Zeitaufwand in h für Verwaltungstätigkeiten in der Vorlesungszeit (a), in der vorlesungsfreien Zeit (b),
- Y_{17} : Durchschnittlich investierte Zeit in h in Beiratstätigkeiten pro Studienjahr,
- $f_{Vorl.-Zeit} = 2 \cdot 13 \cdot \frac{1}{48} = \frac{13}{24}$ und $f_{Vorl.-frei} = 2 \cdot (24 - 13) \cdot \frac{1}{48} = \frac{11}{24}$.

4.2.4 Drittmittelbeschaffung, Gutachten & Herausgeberschaften

Die Anzahl der Drittmittelanträge wurde für die vergangenen 3 Jahre erfragt zusammen mit dem Zeitaufwand pro Antragstyp. Es wurde eine Unterteilung in die verschiedenen Arten von Anträgen vorgenommen, da der Arbeitsaufwand für die unterschiedlichen Antragsarten im Allgemeinen stark variiert. Für den Bereich Drittmittelbeschaffung, Gutachten & Herausgeberschaften ergibt sich aus dem Fragebogen das folgende Berechnungsmodell:

$$\begin{aligned} Y_{Drittmittel} = & f_{Drittm.} \cdot (X_{18a} \cdot Y_{19a} + X_{18b} \cdot Y_{19b} + X_{18c} \cdot Y_{19c} + X_{18d} \cdot Y_{19d} \\ & + X_{18e} \cdot Y_{19e} + X_{18f} \cdot Y_{19f} + X_{18g} \cdot Y_{19g} + X_{18h} \cdot Y_{19h}) \\ & + f_J \cdot (X_{23a} \cdot Y_{24a} + X_{23b} \cdot Y_{24b} + X_{23c} \cdot Y_{24c}) + f_J \cdot Y_{25}, \text{ wobei} \end{aligned}$$

- $X_{18a,18b,18c,18d,18e,18f,18g,18h}$: Anzahl der selbst verfassten Drittmittelanträge (DFG-Einzelantrag (a), Forschergruppenantrag DFG, SFB oder vergleichbar (b), BMBF (c), EU (d), öffentliche Stiftungen (z.B. VolkswagenStiftung) (e), Drittmittel von privaten Geldgebern (Industrie, private Stiftungen etc.) (f), Kleinere Anträge (g), Sonstige (h)) in den vergangenen 3 Jahren (2013-2015),
- $Y_{19a,19b,19c,19d,19e,19f,19g,19h}$: Durchschnittlicher selbst investierter Zeitaufwand in h pro Antrag (Aufteilung wie bei $X_{18a,18b,18c,18d,18e,18f,18g,18h}$),
- $X_{23a,23b,23c}$: Anzahl der selbst verfassten Gutachten (Peer Reviews) im vergangenen Studienjahr (Manuskripte (a), Drittmittelanträge (b), Gruppenbegutachtung (c)),
- $Y_{24a,24b,24c}$: Selbst investierter Zeitaufwand in h für ein Gutachten (Peer Review) (Aufteilung wie bei $X_{23a,23b,23c}$),
- Y_{25} : Im vergangenen Studienjahr selbst investierte Zeit in h für Herausgeberschaften,
- $f_{Drittm.} = \frac{1}{3,48}$.

4.2.5 Forschung

Der Zeitaufwand für die Forschung wird aufgeteilt in das eigentliche Forschen und den Zeitaufwand für Publikationen. Für den Bereich Forschung ergibt sich aus dem Fragebogen das folgende Berechnungsmodell:

$$\begin{aligned}
 Y_{Forschung} = & f_{Vorl-Zeit} \cdot Y_{20a} + f_{Vorl.-frei} \cdot Y_{20b} \\
 & + f_{Publ.} \cdot (X_{21a1} \cdot Y_{22a1} + X_{21b1} \cdot Y_{22b1} + X_{21c1} \cdot Y_{22c1} + X_{21d1} \cdot Y_{22d1} \\
 & + X_{21e1} \cdot Y_{22e1} + X_{21f1} \cdot Y_{22f1} + X_{21g1} \cdot Y_{22g1} + X_{21h1} \cdot Y_{22h1} \\
 & + X_{21a2} \cdot Y_{22a2} + X_{21b2} \cdot Y_{22b2} + X_{21c2} \cdot Y_{22c2} + X_{21d2} \cdot Y_{22d2} \\
 & + X_{21e2} \cdot Y_{22e2} + X_{21f2} \cdot Y_{22f2} + X_{21g2} \cdot Y_{22g2} + X_{21h2} \cdot Y_{22h2}), \text{ wobei}
 \end{aligned}$$

- $Y_{20a,20b}$: Wöchentliche Arbeitszeit in h zum Forschen in der Vorlesungszeit (a), in der vorlesungsfreien Zeit (b),
- $X_{21a1,\dots,21h1}$: Anzahl an Veröffentlichungen in den letzten 2 Jahren (2014-2015) als alleinige(r) Autor(in) (Publikationen mit Peer Review (a1), Publikationen ohne Peer Review (b1), Wissenschaftliche Monografien (c1), Populärwissenschaftliches Sachbuch (d1), Patentanmeldungen (e1), Künstlerische Veröffentlichungen (f1), Populäre Medien (g1), Sonstiges (h1)),
- $Y_{22a1,\dots,22h1}$: Selbst investierter Zeitaufwand in h in die Erstellung der jeweils letzten Veröffentlichung als alleinige(r) Autor(in) (Aufteilung wie bei $X_{21a1,\dots,21h1}$),
- $X_{21a2,\dots,21h2}$: Anzahl an Veröffentlichungen in den letzten 2 Jahren (2014-2015) als Co-Autor(in) (Publikationen mit Peer Review (a2), Publikationen ohne Peer Review (b2), Wissenschaftliche Monografien (c2), Populärwissenschaftliches Sachbuch (d2), Patentanmeldungen (e2), Künstlerische Veröffentlichungen (f2), Populäre Medien (g2), Sonstiges (h2)),
- $Y_{22a2,\dots,22h2}$: Selbst investierter Zeitaufwand in h in die Erstellung der jeweils letzten Veröffentlichung als Co-Autor(in) (Aufteilung wie bei $X_{21a2,\dots,21h2}$),
- $f_{Publ.} = \frac{1}{2,48} = \frac{1}{96}$.

4.2.6 Tagungen und Vorträge

Der Bereich Tagungen und Vorträge beinhaltet die Ausrichtung und den Besuch von Tagungen sowie den Zeitaufwand für das Vorbereiten und Halten von Vorträgen. Für den Bereich Tagungen und Vorträge ergibt sich aus dem Fragebogen das folgende Berechnungsmodell:

$$\begin{aligned} Y_{Tagungen} &= f_{Tagung} \cdot (X_{26a} \cdot Y_{27a} + X_{26b} \cdot Y_{27b} + X_{26c} \cdot Y_{27c} + X_{26d} \cdot Y_{27d}) \\ &+ f_{Besuche} \cdot (Y_{28a} + Y_{28b} + Y_{28c}) + f_J \cdot X_{29} \cdot Y_{30} \\ &= \tilde{Y}_{Tagungen} + Y_{Summe} \cdot \frac{1}{7 \cdot 48} \cdot (Y_{28a} + Y_{28b} + Y_{28c}), \text{ wobei} \end{aligned}$$

- $X_{26a,26b,26c,26d}$: Anzahl der selbst ausgerichteten Tagungen in den letzten 2 Jahren (2014-2015) (Wissenschaftliche Tagungen in zentraler Funktion (a), Nicht-wissenschaftliche Tagungen in zentraler Funktion (b), Wissenschaftliche Tagungen in unterstützender Funktion (c), Nicht-wissenschaftliche Tagungen in unterstützender Funktion (d)),
- $Y_{27a,27b,27c,27d}$: Durchschnittlich selbst investierter Zeitaufwand in h pro Tagung (Aufteilung wie bei $X_{26a,26b,26c,26d}$),
- $Y_{28a,28b,28c}$: Wie viele Tage haben Sie im vergangenen Jahr auf Tagungen oder bei Vorträgen an Universitäten verbracht? (Wissenschaftliche Tagungen (a), Nicht-wissenschaftliche Tagungen (b), Sonstiges (z.B. Universitätsvorträge) (c)),
- X_{29} : Anzahl der selbst gehaltenen Vorträge im vergangenen Jahr,
- Y_{30} : Typischer selbst investierter Zeitaufwand in h für die Vorbereitung eines Vortrags,
- $f_{Tagung} = \frac{1}{2 \cdot 48} = \frac{1}{96}$ und $f_{Besuche} = \frac{Y_{Summe}}{7 \cdot 48}$ (Eine Näherung für die Arbeitsstunden für einen Tagungstag ermitteln wir mit Hilfe der durchschnittlichen Gesamtarbeitszeit pro Tag. Dabei gehen wir davon aus, dass die Wochenarbeitszeit irgendwann in den 7 Tagen einer Woche erbracht wird.).

4.2.7 Sonstiges

Um die Möglichkeit zu bieten, auch solche Tätigkeiten zu erwähnen, die mit den Fragen 1 bis 30 nicht abgedeckt wurden, konnten die Teilnehmer unter Frage 31 weitere Tätigkeiten angeben. Diese Antworten wurden händisch ausgewertet und daraus die Stundenzahl für $Y_{Sonstiges}$ bestimmt. Die Antworten auf diese Frage wurden auch verwendet, um Fragebögen auszuschließen, etwa wenn ein Abbruch angedeutet wird (vgl. Abschnitt 6.1).

4.2.8 Gesamtarbeitszeit

Das Messmodell für die wöchentliche Gesamtarbeitszeit ergibt sich als Summe der oben genannten Terme für die einzelnen Tätigkeitsfelder wie folgt:

$$\begin{aligned} Y_{Summe} &= Y_{Lehre} + Y_{Wiss.Nachwuchs} + Y_{Administration} + Y_{Drittmittel} + Y_{Forschung} + \tilde{Y}_{Tagungen} \\ &+ Y_{Summe} \cdot \frac{1}{7 \cdot 48} \cdot (Y_{28a} + Y_{28b} + Y_{28c}) + Y_{Sonstiges}, \text{ d.h.} \end{aligned}$$

$$\begin{aligned} Y_{Summe} &= (Y_{Lehre} + Y_{Wiss.Nachwuchs} + Y_{Administration} + Y_{Drittmittel} + Y_{Forschung} + \tilde{Y}_{Tagungen} \\ &+ Y_{Sonstiges}) / (1 - \frac{1}{7 \cdot 48} \cdot (Y_{28a} + Y_{28b} + Y_{28c})). \end{aligned}$$

5 A-priori Verteilungen

Die Angaben in den in Abschnitt 3 genannten Studien werden über a-priori Verteilungen in die Auswertung der neuen Befragung eingebunden. Dabei ist allerdings zu beachten, dass aus den Studien in erster Linie Mittelwerte ohne Vertrauensintervalle oder Standardabweichungen hervorgehen. Die Unsicherheit der Vorinformation wird jedoch bei der Wahl der a-priori Verteilung benötigt.

In diesem Abschnitt soll die Bestimmung der Parameter von a-priori Verteilungen, der so genannten Hyperparameter, aus Mittelwerten erläutert werden und zu welchen Variablen welche a-priori Information genutzt wird. Außerdem werden Sensitivitätsanalysen zur Untersuchung des Einflusses der für die Festsetzung der a-priori Verteilungen notwendigen Hyperparameter auf die prädiktiven a-posteriori Verteilungen vorgestellt.

5.1 Hyperparameter informativer a-priori-Verteilungen

Falls a-priori Wissen zu Mittelwert \bar{x} und Standardabweichung s_x aus vorangegangenen Studien vorliegt, können die Hyperparameter aus diesen Informationen geschätzt werden. Diese so genannte Empirische Bayes-Methode wird nachfolgend für die hier verwendeten Gamma-Verteilungen $\text{Gamma}(\alpha, \lambda)$ mit Hyperparametern α und λ erläutert, die hier durchgängig für alle Variablen unterstellt wird.

Falls nur der Mittelwert aus einer Studie bekannt ist, muss auch die Standardabweichung der Daten aus der Vorstudie geschätzt werden. Dies kann mit Hilfe einer Schätzung der *fractional standard variation* (fsd) geschehen, so dass $s_x = \text{fsd} \cdot \bar{x}$. Stattdessen kann natürlich die empirische Standardabweichung als Schätzer verwendet werden. Bei der Gammaverteilung können daraus Schätzer für α und λ mit Hilfe der Momentenmethode bestimmt werden als $\hat{\alpha} = \frac{\bar{x}^2}{s_x^2}$ und $\hat{\lambda} = \frac{\bar{x}}{s_x^2}$.

Wie nehmen nun an, dass die Parameter α und λ jeweils wiederum einer Gamma-Verteilung folgen, also $\alpha \sim \text{Gamma}(c, d)$, $\lambda \sim \text{Gamma}(a, b)$. Dann liefert ein nochmaliges Anwenden der Momentenmethode folgende Schätzer:

$$\hat{c} = \frac{\hat{\alpha}^2}{\text{Var}(\alpha)} = \left(\frac{\bar{x}^2}{s_x^2}\right)^2 \cdot \frac{1}{\text{Var}(\alpha)}, \quad \hat{d} = \frac{\hat{\alpha}}{\text{Var}(\alpha)} = \frac{\bar{x}^2}{s_x^2} \cdot \frac{1}{\text{Var}(\alpha)} \quad (4)$$

$$\hat{a} = \frac{\hat{\lambda}^2}{\text{Var}(\lambda)} = \left(\frac{\bar{x}}{s_x^2}\right)^2 \cdot \frac{1}{\text{Var}(\lambda)}, \quad \hat{b} = \frac{\hat{\lambda}}{\text{Var}(\lambda)} = \frac{\bar{x}}{s_x^2} \cdot \frac{1}{\text{Var}(\lambda)} \quad (5)$$

Die Einstellungsparameter fsd, $\text{Var}(\alpha)$ und $\text{Var}(\lambda)$ werden dabei so gewählt, dass sie die Verlässlichkeit bzw. den Informationsgehalt der Information widerspiegeln.

Beispiel: Wöchentliche Arbeitszeit für Forschung: In der in Abschnitt 3 genannten LESSI-Studie (vgl. Schomburg et al. (2012)) wurde der Zeitaufwand der Vollzeitbeschäftigten für verschiedene Arbeitsbereiche, gegliedert nach Vorlesungszeit und vorlesungsfreier Zeit, erhoben. Diesen Fragen wurden die Variablen Y_{20a} und Y_{20b} zugeordnet. Die Befragung der Universitätsprofessorinnen und -professoren ergab für den Arbeitsbereich *Forschung* einen arithmetischen Mittelwert von 14,5 h in der Vorlesungszeit und 24,9 h in der vorlesungsfreien Zeit.

Diese arithmetischen Mittelwerte gehen, wie im vorangegangenen Abschnitt erläutert, in die Schätzung der Hyperparameter für die Gammaverteilung der Variablen Y_{20a} und Y_{20b} ein. Für die Variable Y_{20a} müssen neben $\bar{x} = 14,5$ noch s_x^2 , $\text{Var}(\alpha)$ und $\text{Var}(\lambda)$ gewählt werden. Die Stichprobenvarianz wurde mit Hilfe von fsd (*fractional standard deviation*) bestimmt. Wird z.B. $\text{fsd} = 0,3$ angenommen, so ergibt sich $s_x^2 = (\text{fsd} \cdot \bar{x})^2 = (0,3 \cdot 14,5)^2 = 18,92$.

Die Unsicherheit bezüglich der Parameter α und λ wird über die entsprechenden Varianzen ausgedrückt. In diesem Beispiel werden auf Grund der vorliegenden Informationen kleine Varianzen gewählt und zwar

$\text{Var}(\alpha) = \text{Var}(\lambda) = 0,02$. Daraus resultieren folgende a-priori Verteilungen: $\alpha \sim \text{Gamma}(6174, 556)$, $\lambda \sim \text{Gamma}(19, 81)$.

Die Einstellungen der Parameter fsd , $\text{Var}(\alpha)$ und $\text{Var}(\lambda)$ wurden aus der folgenden Sensitivitätsanalyse abgeleitet, die zunächst für einige derjenigen Variablen durchgeführt wurde, zu denen keine belastbaren Angaben aus Vorstudien gefunden werden konnten.

5.2 Sensitivitätsanalyse für die Einstellungsparameter schwach-informativer a-priori Verteilungen

Für diejenigen Fragen, zu denen keine belastbaren Angaben aus Vorstudien gefunden werden konnten, sollten die entsprechenden a-priori Verteilungen so gewählt werden, dass letztere nur einen schwachen Einfluss auf die Auswertungsergebnisse haben. Wie in diesem Fall die Einstellungsparameter zu wählen sind, wurde im Rahmen der folgenden Untersuchung für vier Beispielfragen bzw. Variablen untersucht. Für die vier Variablen Y_{14a} , Y_{16a} , X_{21a1} und X_{26a} wurde jeweils ein Plackett-Burman-Versuchsplan (vgl. Weihs & Jessenberger (1999, S. 239)) zur Untersuchung des Einflusses der einzelnen Einstellungsparameter aufgestellt (als Beispiel vgl. Tabelle 1). Variiert wurden die Faktoren *Mittelwert* auf zwei individuellen Faktorstufen, *fsd* (fractional standard deviation) auf den Faktorstufen 0,1 und 0,5 sowie $\text{Var}(\alpha)$ und $\text{Var}(\lambda)$ jeweils auf den Faktorstufen 0,02 und 0,1.

Tab. 1: Plackett-Burman-Versuchsplan für die Variable Y_{14a} , inkl. resultierende Zielvariable d

| Versuch | Mittelwert (h) | fsd | $\text{Var}(\alpha)$ | $\text{Var}(\lambda)$ | d |
|---------------------------------------|--------------------|---------------------|----------------------|-----------------------|------|
| 1 | 30 | 0,5 | 0,1 | 0,02 | 0,12 |
| 2 | 10 | 0,1 | 0,1 | 0,1 | 0,84 |
| 3 | 10 | 0,5 | 0,1 | 0,1 | 0,11 |
| 4 | 10 | 0,5 | 0,02 | 0,02 | 0,21 |
| 5 | 30 | 0,1 | 0,1 | 0,02 | 0,84 |
| 6 | 10 | 0,1 | 0,02 | 0,02 | 0,88 |
| 7 | 30 | 0,5 | 0,02 | 0,1 | 0,21 |
| 8 | 30 | 0,1 | 0,02 | 0,1 | 0,87 |
| $\hat{\beta}_{1,\dots,\hat{\beta}_4}$ | 0,00 | -0,35 | -0,03 | 0,00 | |
| p -Wert | 0,98 | $2,5 \cdot 10^{-5}$ | 0,02 | 0,9 | |

Zur Beurteilung des Einflusses der a-priori Verteilung auf die a-posteriori Verteilung wurde für jeden Versuch die Stichprobenverteilung mit der prädiktiven a-posteriori Verteilung der betrachteten Variable verglichen und überprüft, wie gut die beiden Verteilungen übereinstimmen. Betrachtet wurden zum einen der Mittelwert \bar{x} , das 0,025 %-Quantil $q_{0,025\%}$ und das 0,975 %-Quantil $q_{0,975\%}$ der Datenverteilung. Zum anderen wurde der Mittelwert μ_X der prädiktiven a-posteriori Verteilung betrachtet sowie die rechte und linke Grenze hpd_r, hpd_l des dazugehörigen 95 %-hpd-Intervalls. Aus diesen Größen wurde in Analogie zu Wünschbarkeitsindizes (Weihs & Jessenberger (1999, S. 176)) folgende Maßzahl zur Beurteilung der Übereinstimmung der beiden Verteilungen ermittelt:

$$d = \left| 1 - \frac{\mu_X}{\bar{x}} \cdot \frac{hpd_r - hpd_l}{q_{0,975} - q_{0,025}} \right|. \quad (6)$$

Bei guter Übereinstimmung der Mittelwerte liegt der Term $\frac{\mu_X}{\bar{x}}$ nahe 1 und bei guter Übereinstimmung der Verteilungsintervalle liegt der Term $\frac{hpd_r - hpd_l}{q_{0,975} - q_{0,025}}$ nahe bei 1. Stimmen sowohl Mittelwerte als auch Verteilungsintervalle gut überein, ist d folglich nahe 0. Je kleiner also der Wert für d , desto geringer der Einfluss der a-priori Verteilung auf die prädiktive a-posteriori Verteilung.

Mit den Daten des Plackett-Burman-Plans wurde eine lineare Regression durchgeführt mit d als Zielvariable, also dem Modell:

$$d = \beta_0 + \beta_1 \cdot \bar{x} + \beta_2 \cdot \text{fsd} + \beta_3 \cdot \text{Var}(\alpha) + \beta_4 \cdot \text{Var}(\lambda). \quad (7)$$

Der Vorteil des Plackett-Burman-Versuchsplans ist die Orthogonalität der Einflussfaktoren und die daraus resultierende eindeutige Bestimmung der Faktoreffekte. Bei der Regression hat sich gezeigt, dass nur fsd für alle vier untersuchten Variablen zum 5 %-Niveau signifikant ist (vgl. beispielhaft die Ergebnisse für Y_{14a} in Tabelle 1). Eine Erhöhung von fsd bewirkt wegen des negativen Vorzeichens von $\hat{\beta}_2$ eine Verringerung von d , d.h. eine bessere Übereinstimmung zwischen der Datenverteilung und prädiktiver a-posteriori Verteilung. Der Einfluss von $\text{Var}(\alpha)$ hat sich nur für die Variablen Y_{14a} und Y_{16a} auf dem 5 %-Niveau als signifikant erwiesen und die Vorzeichen der zugehörigen Koeffizienten $\hat{\beta}_3$ sind ebenfalls negativ. Deswegen wird für schwach informative a-priori Verteilungen die jeweils höhere Faktorstufe $\text{fsd} = 0,5$ bzw. $\text{Var}(\alpha) = 0,1$ verwendet. Der Faktor $\text{Var}(\lambda)$ hat keinen signifikanten Einfluss zum 5 %-Niveau gezeigt. Wegen des negativen Vorzeichens des zugehörigen Koeffizienten $\hat{\beta}_4$ empfiehlt sich jedoch auch hier die jeweils größere Einstellung. Auch die Mittelwerte haben keinen zum 5 %-Niveau signifikanten Einfluss. Die Untersuchungsergebnisse lassen also darauf schließen, dass sich folgende Einstellungen für die Bestimmung nur schwach-informativer a-priori Verteilungen empfehlen: $\text{fsd} = 0,5$ und $\text{Var}(\alpha) = \text{Var}(\lambda) = 0,1$. Für den Mittelwert reicht es, diesen in einem jeweils realistischen Wertebereich zu wählen (vgl. Tabelle 2).

Tab. 2: A-priori Mittelwerte (\bar{x}) (für $X_2, X_{11}, Y_{20a}, Y_{20b}$ geschätzt, sonst gesetzt)

| | | | | | | | |
|-----------|---------------------|-----------------------------|-------------------|---------------------|---------------------|-------------------|----------|
| Variable | X_2 | Y_{3a} | Y_{3b} | Y_4 | X_5 | Y_6 | |
| \bar{x} | 9 | 30 | 20 | 60 | 20 | 60 | |
| Variable | $X_{7a,\dots,e}$ | $Y_{8a,\dots,e}$ | X_{9a} | $X_{9b,\dots,e}$ | X_{9f} | $X_{9g,\dots,j}$ | |
| \bar{x} | 20 | 10 | 10 | 8 | 10 | 8 | |
| Variable | $Y_{10a,f}$ | $Y_{10b,\dots,e,g,\dots,j}$ | X_{11} | $X_{12a,b}$ | $X_{13a,b}$ | $Y_{14a,b}$ | |
| \bar{x} | 20 | 30 | 6 | 3 | 4 | 8 | |
| Variable | $Y_{16a,b}$ | Y_{17} | $X_{18a,\dots,h}$ | $Y_{19a,\dots,h}$ | Y_{20a} | Y_{20b} | |
| \bar{x} | 10 | 20 | 10 | 20 | 14,5 | 24,9 | |
| Variable | $X_{21a1,\dots,h1}$ | $Y_{22a1,b1}$ | Y_{22c1} | $Y_{22d1,\dots,h1}$ | $X_{21a2,\dots,h2}$ | $Y_{22a2,b2}$ | |
| \bar{x} | 5 | 15 | 30 | 15 | 12 | 15 | |
| Variable | Y_{22c2} | $Y_{22d2,\dots,h2}$ | $X_{23a,b,c}$ | $Y_{24a,b,c}$ | Y_{25} | $X_{26a,\dots,d}$ | |
| \bar{x} | 20 | 15 | 6 | 20 | 20 | 6 | |
| Variable | Y_{27a} | Y_{27b} | $Y_{27c,d}$ | $Y_{28a,b,c}$ | X_{29} | Y_{30} | Y_{31} |
| \bar{x} | 45 | 30 | 15 | 20 | 15 | 25 | 1 |

5.3 Informative a-priori Verteilungen

Für die Variablen $X_2, X_{11}, Y_{20a}, Y_{20b}, Y_{Gesamt}, Y_{Summe}$ und $Y_{Lehre}, \dots, Y_{Sonstiges}$ liegen Vorinformationen aus umfangreichen, repräsentativen Erhebungen vor. Daher empfiehlt es sich, die Parameter-Einstellungen zur Bestimmung der Hyperparameter so zu wählen, dass die resultierenden a-priori Verteilungen einen hohen informativen Einfluss auf die prädiktiven a-posteriori Verteilungen besitzen. Basierend auf der Sensitivitätsanalyse in Abschnitt 5.2 verwenden wir im Gegensatz zu dem schwachinformativen Fall die Parameter-Einstellungen $\text{Var}(\alpha) = \text{Var}(\lambda) = 0,02$. fsd wird auf der Basis der Wertebereiche der Variablen individuell wie folgt festgelegt: X_2 : $\text{fsd} = 0,3$, X_{11} : $\text{fsd} = 0,7$, Y_{20a} : $\text{fsd} = 0,4$, Y_{20b} : $\text{fsd} = 0,3$, Y_{Gesamt} : $\text{fsd} = 0,2$. Ob diese Wahl einen signifikanten Einfluss auf die Schätzung der Gesamtarbeitszeit hat, untersuchen wir in der Sensitivitätsanalyse in Abschnitt 6.4.4. Die Varianz von Y_{Summe} wird aus den Gesamtschätzungen Y_{Gesamt} bestimmt. Mittelwerte und Varianzen von $Y_{Lehre}, \dots, Y_{Sonstiges}$ werden

als Prozentanteile von Mittelwert und Varianz von Y_{Summe} geschätzt.

Folgende Literaturquellen wurden zur Bestimmung der a-priori Parameter verwendet:

- X_2 : In der LESSI-Studie (vgl. Schomburg et al. (2012)) wurde ein arithmetischer Mittelwert von 9 *SWs* für den tatsächlichen Lehraufwand eines Universitätsprofessors (mit Lehrauftrag) pro Semester ermittelt. Insgesamt wurden Angaben zwischen 1 und 20 *SWs* gemacht.
- X_{11} : In der Studie *Promovierende in Deutschland* des Statistischen Bundesamtes wurde die Anzahl der Promovierenden je Professur geschätzt (vgl. Statistisches Bundesamt (2016a)). Der Mittelwert über alle Fächergruppen liegt bei 6 Promovierenden je Lehrkraft. Die Mittelwerte variieren jedoch sehr stark von Fächergruppe zu Fächergruppe (zwischen 5 und 11 im WS2014/2015).
- Y_{20a} & Y_{20b} : In der LESSI-Studie (vgl. Schomburg et al. (2012)) wurde ein arithmetischer Mittelwert von 14,5 *h*/Woche in der Vorlesungszeit und 24,9 *h*/Woche in der vorlesungsfreien Zeit für Universitätsprofessoren ermittelt.
- Y_{Gesamt} : In der LESSI-Studie (vgl. Schomburg et al. (2012)) wurden arithmetische Mittelwerte veröffentlicht. In der Vorlesungszeit lag die angegebene Wochenarbeitszeit bei Universitätsprofessorinnen und -professoren im Mittel bei 56,5 *h*, in der vorlesungsfreien Zeit bei 51,9 *h*. Diese beiden Werte wurden entsprechend der Länge der repräsentierten Zeiten miteinander verrechnet. Als Gewichte wurden damit verwendet: 26/48 für die Vorlesungszeit und 22/48 für die vorlesungsfreie Zeit. Insgesamt ergab sich daraus eine mittlere Wochenarbeitszeit von 54,4 *h*.
- Y_{Summe} : Als a-priori Information für die Verteilung von Y_{Summe} werden sowohl die mittlere Wochenarbeitszeit von 54,4 *h* nach der LESSI-Studie verwendet (s. Y_{Gesamt}) als auch Mittelwert und Varianz von Y_{Gesamt} . Einerseits sind diese Informationen aktueller als die der LESSI-Studie, andererseits umfasst die LESSI-Studie deutlich mehr Professorinnen und Professoren (> 5000). Deshalb werden die Mittelwert-Informationen gleich gewichtet. Da die LESSI-Studie keine Information zur Varianz liefert, wird hier der Wert aus der Gesamtschätzung verwendet.
- $Y_{Lehre}, \dots, Y_{Sonstiges}$: Als a-priori Information für die Prozentanteile von $Y_{Lehre}, \dots, Y_{Sonstiges}$ an Y_{Summe} werden die Prozentsätze der iFQ-Wissenschaftlerbefragung (Böhmer et al., 2011) verwendet. Allerdings benutzt diese Studie eine etwas andere Abgrenzung der Teilaufgaben. Deshalb gehen wir wie folgt vor:
 - $Lehre(\%) = (Lehre + 0,5 \cdot Betreuung)(\%iFQ) = (26,1 + 7,0)\% = 33,1\%$,
 - $Wissenschaftlicher\ Nachwuchs(\%) = 0,5 \cdot Betreuung(\%iFQ) = 7,0\%$,
 - $Administration(\%) = (Gremien/Selbstverw.+0,5 \cdot Management)(\%iFQ) = (10,2+4,3)\% = 14,5\%$,
 - $Drittmittel/Gutachten(\%) = (Drittmittel + Begutachtungen)(\%iFQ) = (8,9 + 8,8)\% = 17,7\%$,
 - $Forschung(\%) = 0,9 \cdot Forschung(\%iFQ) = (0,9 \cdot 21,5)\% = 19,4\%$,
 - $Tagungen\&\ Vorträge(\%) = (0,5 \cdot Management+0,1 \cdot Forschung)(\%iFQ) = (4,3+0,1 \cdot 21,5)\% = 6,5\%$,
 - $Sonstiges(\%) = Sonstiges(\%iFQ) = 1,8\%$.

Diese Prozentsätze werden zur Bestimmung der a-priori Mittelwerte und a-priori Varianzen auf Mittelwert und Varianz von Y_{Summe} angewendet. Bei der Bestimmung der Varianzen wird dabei implizit die Unabhängigkeit der Angaben zu den Teilaufgaben angenommen. Das erscheint deshalb gerechtfertigt, weil die Angaben zu den Fragen der Teilaufgaben niemals verknüpft wurden beim

Ausfüllen des Fragebogens. Sowohl die Variablen $Y_{Lehre}, \dots, Y_{Sonstiges}$ als auch Y_{Summe} werden ebenfalls als Gamma-verteilt angenommen.²

Die aus der Vorinformation geschätzten a-priori Mittelwerte der Variablen $X_2, X_{11}, Y_{20a}, Y_{20b}$ finden sich ebenfalls in Tabelle 2. Als a-priori Mittelwert von Y_{Gesamt} wird $54,4 h$ aus der LESSI-Studie verwendet. Die a-priori Parameter Mittelwert und Varianz von Y_{Summe} und $Y_{Lehre}, \dots, Y_{Sonstiges}$ finden sich in Tabelle 3, fsd wird bei diesen Variablen nicht benötigt.

Tab. 3: A-priori Mittelwerte und Varianzen

| Variable | Mittelwert (h) | Varianz (h^2) |
|-------------------------|--------------------|-------------------|
| Y_{Summe} | 55,4 | 88,2 |
| Y_{Lehre} | 18,3 | 29,2 |
| $Y_{Wiss.Nachwuchs}$ | 3,9 | 6,2 |
| $Y_{Administration}$ | 8,0 | 12,8 |
| $Y_{Dritt./Gutachten}$ | 9,8 | 15,6 |
| $Y_{Forschung}$ | 10,8 | 17,1 |
| $Y_{Tagungen/Vorträge}$ | 3,6 | 5,7 |
| $Y_{Sonstiges}$ | 1,0 | 1,6 |

Die in Abschnitt 3 aufgeführte DHV-Studie wird nicht als Vorinformation einbezogen, sondern unseren Ergebnissen in der Diskussion in Abschnitt 7 gegenübergestellt.

6 Ergebnisse

Die Auswertung der Daten nach dem frequentistischen und dem Bayesianische Ansatz wird in den folgenden Abschnitten gegenübergestellt. Doch zunächst folgen Überlegungen zu Ausschluss-Kriterien und Repräsentativität der Studie.

6.1 Ausschluss-Kriterien

An der Befragung haben insgesamt 2609 Professorinnen und Professoren teilgenommen. Für die oben beschriebenen Auswertungen wurden davon 100 Antwortbögen entfernt, die die Beantwortung abgebrochen haben sowie weitere 59 Antwortbögen mit negativen Einträgen. Aus den übrigen 2450 Antwortbögen wurden die Antworten der 155 bereits emeritierten Professorinnen und Professoren entfernt, da diese den jetzigen Universitätsbetrieb nicht mehr repräsentativ widerspiegeln. Da es sich bei 93 % der Teilnehmer um Universitätsprofessuren handelt und die übrigen Gruppen zu gering vertreten sind, wurden weitere 155 Antwortbögen der anderen Gruppen (Fachhochschulen, private Hochschulen, ...) ausgeschlossen. Damit verbleiben 2140 Antwortbögen. Diese wurden für die Untersuchungen der Gesamtschätzung der durchschnittlichen wöchentlichen Arbeitszeit und für die Untersuchung der Teilarbeitszeiten und deren Summe wie folgt weiter analysiert.

Für die Untersuchung der Gesamtschätzung der durchschnittlichen wöchentlichen Arbeitszeit wurden zusätzlich sehr hohe Werte größer als $120 h$ sowie Gesamtschätzungen mit weniger als $20 h$ pro Woche

²Man beachte, dass die Summe unabhängiger Gamma-verteilter Zufallsvariablen wieder Gamma-verteilt ist (Moschopoulos (1985)). Produkte $X \cdot Y$ von Gamma-verteiltern Zufallsvariablen X, Y sind zwar nicht exakt Gamma-verteilt (vgl. Marques (2012)), aber Gamma-Verteilungen sind nach unseren eigenen Untersuchungen gute Näherungen. Deshalb können $Y_{Lehre}, \dots, Y_{Sonstiges}$ und Y_{Summe} als Gamma-verteilt angenommen werden.

entfernt, da bei letzteren nicht von einer Vollzeitbeschäftigung auszugehen ist. Insgesamt flossen damit 1992 Angaben in die Analyse der Gesamtschätzung ein, darunter 811 Angaben von *Geistes- und Sozialwissenschaftlern*, 257 von *Lebenswissenschaftlern*, 624 von *Naturwissenschaftlern* sowie 300 von *Ingenieurwissenschaftlern*. Die Anzahl Professorinnen liegt bei 543 (27,3%).

Für die Untersuchung der Teilarbeitszeiten und deren Summe wurden die zunächst verbleibenden 2140 Antwortbögen wie folgt weiter analysiert. In Frage 6 waren die sehr unterschiedlichen Größenordnungen der Angaben auffällig. Gekoppelt an die Tatsache, dass nur diese Frage in Minuten anstatt in Stunden beantwortet werden sollte, haben wir daraus geschlossen, dass einige Teilnehmer diese Frage in Stunden beantwortet haben. Aus diesem Grund wurden die Antworten zu dieser Frage wie folgt bearbeitet: Angaben zwischen 0 und 3 wurden als Stundenangaben gewertet und entsprechend in Minuten umgerechnet. 30 Antwortbögen mit Angaben zwischen 3 und 5 wurden entfernt, da von einer Mindestprüfzeit von 5 Minuten ausgegangen wird. 10 Antwortbögen mit Angaben größer als 180 Minuten zu Frage 6 wurden entfernt, da sie als unrealistisch gewertet werden. Somit wurden weitere 40 Antwortbögen entfernt.

Des Weiteren fiel auf, dass zu manchen Fragen unrealistisch hohe Angaben gemacht wurden, was in der Summe zu sehr hohen Gesamtzeitbelastungen führte. Da diese Angaben zu deutlichen Verzerrungen führen würden, wurden alle 337 Antwortbögen, die zu einer summierten Gesamtarbeitsbelastung von wöchentlich mehr als 120 h führten, entfernt. Außerdem wurde ein weiterer Antwortbogen entfernt mit einem Tagungsbesuch von mehr als 365 Tagen im Jahr. Im nächsten Schritt wurden 6 der nun 1762 Antwortbögen mit extremen Ausreißern entfernt.

Für die nun 1756 Antwortbögen wurde die Frage 31 untersucht. Dadurch wurden Antwortbögen identifiziert, die aus verschiedenen Gründen nicht vollständig ausgefüllt wurden oder solche, die laut eigenen Angaben aus verschiedenen Gründen (wie "außerplanmäßige Professur", "Forschungssemester", etc.) nicht repräsentativ sind. Außerdem ging aus einigen Bögen hervor, dass die Teilnehmer nicht mehr in Vollzeit tätig sind. In diesem Schritt wurden insgesamt 28 Antwortbögen entfernt.

Aus Frage 31 gingen aber ebenfalls Zeitangaben für sonstige Tätigkeiten hervor, die zu den übrigen Tätigkeitsbereichen hinzugerechnet werden müssen und in die weiteren Analysen einfließen.

Abschließend wurden diejenigen Antwortbögen entfernt, die nach Hinzunehmen des Bereichs „Sonstige Tätigkeiten“ zu einer summierten wöchentlichen Gesamtarbeitsbelastung von mehr als 120 h oder immer noch zu weniger als 20 h (keine Vollzeittätigkeit) führten. Dies sind insgesamt weitere 38 Bögen. Schließlich blieben 1690 Antwortbögen von nicht emeritierten Universitätsprofessorinnen und -professoren, die für die Auswertung der Arbeitszeiten für Teilaufgaben und deren Summe ausgewertet werden konnten, darunter die von 425 Professorinnen. Die *Geistes- und Sozialwissenschaften* sind mit 632 Teilnehmern am meisten vertreten, gefolgt von den *Naturwissenschaften* mit 558 Teilnehmern, die beiden anderen Fächergruppen mit 234 Teilnehmern der *Lebenswissenschaften* und 266 Teilnehmern aus den *Ingenieurwissenschaften* etwas weniger.

6.2 Repräsentativität

Unsere Studie beruht auf einer E-Mail-Umfrage, die einer online-Umfrage sehr ähnlich ist. Daraus ergeben sich prinzipiell verschiedene Einschränkungen der Validität. Die angefragte Population der Teilnehmer war zwar weitgehend vollständig (auf der Basis von Kürschners Gelehrtenkalender 2016), aber die Teilnahme war freiwillig. Zur Verringerung von Erhebungsverzerrungen wurde zur Bestimmung der mittleren wöchentlichen Arbeitszeiten von Professorinnen und Professoren über alle Fächergruppen auch eine *Zelungewichtung* nach Geschlecht bzw. Fächergruppe durchgeführt (vgl. z.B. Milic (2015)).

Eine aktuelle Fächeraufteilung der Professorinnen und Professoren an Universitäten wurde extra für diese Studie vom Statistischen Bundesamt berechnet (20.9.2017). Für Tabelle 4 wurde diese Aufteilung wie folgt umgerechnet: *Geist/Sozial*: {*Geistes-, Kunst-, Rechts-, Wirtschafts- und Sozialwissenschaften*}, *Natur*: {*Mathematik, Naturwissenschaften*}, *Leben*: {*Humanmedizin/Gesundheitswissenschaften, Agrar-, Forst-, Ernährungswissenschaften, Veterinärmedizin, Sport*} und *Ingenieur*: {*Ingenieurwissen-*

schaften}. Zentrale Einrichtungen (, die sich nicht an Kliniken befinden) wurden nicht zugeordnet. Daraus ergeben sich Sollprozentzahlen für eine Zellgewichtung nach Geschlecht bzw. Fächergruppe. Die in der Umfrage realisierten Prozentzahlen bei den für die Summation gültigen Fragebogen (s. Abschnitte 4.2.8, 6.1) sind: *Geist/Sozial*: 37%, *Natur*: 33%, *Leben*: 14%, *Ingenieur*: 16%. Man sieht, dass die *Naturwissenschaften* in der Umfrage stark überrepräsentiert sind und die *Geistes-/Sozialwissenschaften* unterrepräsentiert. Der realisierte Frauenanteil war mit 25% nur leicht zu hoch. Bei den Anteilen an der Gesamtschätzung der wöchentlichen Arbeitszeit (s. Tabelle 5) sind ebenfalls die *Naturwissenschaften* stark überrepräsentiert.

Tab. 4: Aufteilung der Professorinnen und Professoren 2015 an Hochschulen auf die Fächergruppen

| | Geist/Sozial | Natur | Leben | Ingenieur | insg. |
|------------------|--------------|-------|-------|-----------|-----------------|
| alle Prof. | 10389 | 5565 | 4178 | 3589 | 23721 von 24101 |
| ungef. % (insg.) | 44 | 23 | 18 | 15 | |
| weiblich | 3267 | 921 | 735 | 403 | 5326 von 5425 |
| in % (Fach) | 31,4 | 16,5 | 17,6 | 11,2 | 22,5 |

6.3 Gesamtschätzung der wöchentlichen Arbeitszeit

Die Teilnehmer wurden gebeten, selbst einzuschätzen, wie viele Stunden sie insgesamt wöchentlich arbeiten. Ausgewertet wurden die bzgl. der Gesamtschätzung 1992 gültigen Antwortbögen (s. Abschnitt 6.1). Abbildung 2 zeigt die Verteilung der Gesamtschätzungen für alle Fächergruppen zusammen. Diese Verteilung liegt im Wesentlichen zwischen etwas unter 40 h und 80 h.

Bei dem frequentistischen Ansatz werden Mittelwerte und 95%-Prognoseintervalle für alle gültigen Antwortbögen zusammen berechnet. Außerdem werden die entsprechenden wöchentlichen Gesamtarbeitszeiten der vier Fächergruppen *Geistes- und Sozial-*, *Lebens-*, *Natur-* und *Ingenieurwissenschaften* und von Professorinnen und Professoren getrennt bestimmt. Die Ergebnisse sind in Tabelle 5 aufgeführt. Der Mittelwert für die Gesamtschätzung beträgt rund 56,4 h. Lebenswissenschaftler scheinen im Mittel einen etwas höheren Zeitaufwand zu haben, Professorinnen einen etwas kleineren als Professoren. Die 95%-Prognoseintervalle liegen fast immer bei 40 – 80 h, bei den Lebenswissenschaftlern liegt die linke Grenze etwas höher. Durch die Gewichtung mit den Sollprozentanteilen (aus Tabelle 5) der Fächergruppen bzw. der Geschlechter ändert sich das frequentistische Mittel aller gültigen Antworten um höchstens 0,1%, obwohl zumindest die Fächergruppen deutlich nicht richtig repräsentiert sind (Vergleich von Anteil(%) mit Soll(%)).

Für den Bayesianischen Ansatz sind die Mittelwerte der prädiktiven a-posteriori Verteilungen und die entsprechenden 95%-*hpd*-Prognoseintervalle aufgeführt. Dabei haben wir den Mittelwert 54,4 h der LESSI-Studie als a-priori Information für sämtliche Mittelwerte und durchweg $\text{fsd} = 0,2$, $\text{Var}(\alpha) = \text{Var}(\lambda) = 0,02$ verwendet. Auffällig ist, dass die Mittelwerte nahezu exakt mit den frequentistischen übereinstimmen und die *hpd*-Intervalle etwas breiter sind als die frequentistischen Intervalle. Insgesamt zeigen diese Ergebnisse, dass die Modellierung der diskret erscheinenden Gesamtschätzung (s. Abbildung 2) durch eine stetige Gammaverteilung zu angemessenen Ergebnissen führt.

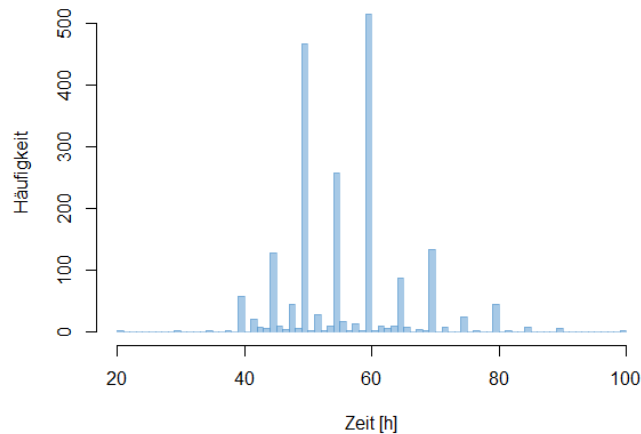


Abb. 2: Gesamtschätzung der wöchentlichen Arbeitszeit: alle Fächergruppen zusammen.

Tab. 5: Gesamtschätzung: Mittelwerte und 95%-Prognoseintervalle (frequentistisch und Bayesianisch) in h

| | frequentistisch | | Bayesianisch | | Anteil(%) | Soll(%) |
|----------------|-----------------|--------------------------|----------------|-----------------|-----------|---------|
| | Mittelwert | $[q_{0.025}, q_{0.975}]$ | a-post. Mittel | 95 %-hpd-Inter. | | |
| Alle | 56,4 | [40, 80] | 56,3 | [35, 79] | 100 | 100 |
| Geist-/Sozial | 55,9 | [40, 80] | 55,9 | [35, 79] | 41 | 44 |
| Leben | 58,0 | [44, 80] | 58,0 | [37, 81] | 13 | 18 |
| Natur | 56,2 | [40, 80] | 56,2 | [35, 78] | 31 | 23 |
| Ingenieur | 56,6 | [41, 80] | 56,6 | [35, 79] | 15 | 15 |
| Professorinnen | 55,5 | [40, 80] | 55,4 | [35, 78] | 27 | 23 |
| Professoren | 56,7 | [40, 80] | 56,7 | [35, 79] | 73 | 77 |

6.4 Summierter wöchentlicher Zeitaufwand

Die folgenden Ergebnisse beziehen sich auf den summierten wöchentlichen Zeitaufwand insgesamt und getrennt nach Tätigkeitsbereichen, Fächergruppen und Geschlechtern.

6.4.1 Frequentistische Analyse

Abbildung 3 zeigt die Verteilung der wöchentlichen Gesamtarbeitszeiten für alle Fächergruppen zusammen. Man beachte, dass Werte kleiner als $20 h$ und größer als $120 h$ nicht berücksichtigt wurden. Der Median der Summen liegt bei $59,3 h$. In Tabelle 6 finden sich frequentistische Mittelwerte und 95%-Prognoseintervalle. Danach ergibt sich insgesamt eine mittlere wöchentliche Arbeitszeit von $63,1 h$ mit einem 95%-Prognoseintervall von etwa $28 h$ bis $113 h$. Das mit den richtigen Zellgewichten der Fächergruppen (vgl. Abschnitt 6.2) berechnete Mittel beträgt $63,3 h$, ist also nur geringfügig höher als ohne Gewichtung.

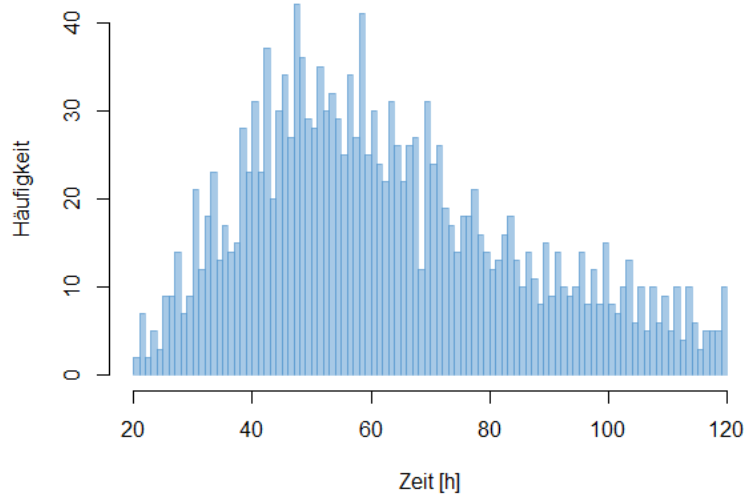


Abb. 3: Summe der wöchentlichen Teilarbeitszeiten

Tab. 6: Mittelwerte für den wöchentlichen Zeitaufwand in h , sowie 0,025 %- und 0,975 %-Quantile: frequentistisch

| | Gesamt | Lehre | Wiss.Nachwuchs | Administration |
|-----------------------|---------------------|-------------------|---------------------|-------------------|
| Alle Fächergruppen | 63,1 [27,5 , 113,1] | 14,4 [2,5 , 44,6] | 6,0 [0,2 , 22,7] | 10,5 [1,5 , 30,0] |
| Geistes-/Sozial-Wiss. | 65,4 [28,2 , 113,5] | 16,5 [4,0 , 50,5] | 4,2 [0,2 , 16,7] | 10,0 [1,9 , 26,3] |
| Lebens-Wiss. | 60,3 [25,4 , 108,7] | 10,6 [1,1 , 34,4] | 6,7 [0,2 , 27,2] | 10,6 [1,4 , 27,5] |
| Natur-Wiss. | 61,9 [30,3 , 111,8] | 13,8 [3,1 , 45,0] | 6,9 [0,5 , 23,4] | 10,2 [1,5 , 30,0] |
| Ingenieur-Wiss. | 62,9 [26,3 , 113,8] | 14,2 [3,6 , 40,0] | 7,5 [0,0 , 25,1] | 11,9 [2,3 , 31,7] |
| Professorinnen | 62,1 [26,9 , 113,0] | 14,9 [3,4 , 40,5] | 5,1 [0,1 , 20,6] | 10,1 [1,5 , 26,4] |
| Professoren | 63,5 [27,7 , 113,1] | 14,2 [2,2 , 45,2] | 6,3 [0,2 , 23,6] | 10,6 [1,6 , 30,3] |
| | Drittmittel | Forschung | Tagungen & Vorträge | Sonstiges |
| Alle Fächergruppen | 6,1 [0,2 , 20,6] | 18,6 [2,4 , 50,4] | 6,3 [0,4 , 21,0] | 1,3 [0,0 , 12,0] |
| Geistes-/Sozial-Wiss. | 5,1 [0,1 , 17,6] | 21,9 [1,9 , 56,0] | 6,8 [0,1 , 20,5] | 1,0 [0,0 , 10,0] |
| Lebens-Wiss. | 6,6 [0,9 , 22,2] | 17,6 [3,3 , 47,0] | 5,9 [0,5 , 20,8] | 2,5 [0,0 , 28,3] |
| Natur-Wiss. | 6,0 [0,6 , 18,6] | 17,9 [3,2 , 46,5] | 6,1 [0,6 , 20,8] | 1,1 [0,0 , 10,0] |
| Ingenieur-Wiss. | 8,3 [0,7 , 24,3] | 13,3 [1,2 , 34,2] | 6,0 [0,3 , 21,2] | 1,6 [0,0 , 10,0] |
| Professorinnen | 5,7 [0,2 , 18,2] | 18,6 [1,9 , 54,2] | 6,3 [0,4 , 21,1] | 1,7 [0,0 , 15,4] |
| Professoren | 6,3 [0,3 , 21,1] | 18,6 [2,6 , 47,8] | 6,3 [0,4 , 20,5] | 1,2 [0,0 , 10,0] |

Geistes- und Sozialwissenschaftler haben im Mittel den größten wöchentlichen Gesamtzeitaufwand, ebenso wie bei den Teilaufgaben *Lehre* und *Forschung*. Der Zeitaufwand für die Betreuung des *wissenschaftlichen Nachwuchses* und die Beantragung von *Drittmitteln* (u.ä.) ist hingegen im Mittel niedriger als bei den anderen Fächergruppen. *Lebenswissenschaftler* weisen die kleinste mittlere Wochenarbeitszeit auf und scheinen am wenigsten Zeit für die *Lehre* aufzuwenden. *Ingenieurwissenschaftler* investieren im Vergleich zu den übrigen Fächergruppen die meiste Zeit in die Betreuung des *wissenschaftlichen*

Nachwuchses, in die *Administration* und in die Beantragung von *Drittmitteln*. Der Aufwand von Professorinnen und Professoren scheint sich nicht deutlich zu unterscheiden. In Abschnitt 6.5 testen wir die mittleren Ergebnisse auf signifikante Unterschiede. Auffällig ist die breite Streuung der Ergebnisse. Mögliche Gründe sowie weitere Aspekte, die bei der Interpretation dieser Ergebnisse zu berücksichtigen sind, werden auch in Abschnitt 7 diskutiert.

6.4.2 Bayesianische Analyse

Bei der Bayesianischen Analyse ergeben sich prädiktive a-posteriori Verteilungen und daraus prädiktive a-posteriori Mittelwerte und Prognoseintervalle auf der Basis der a-priori Festlegungen in Abschnitt 5. Beispielhaft sind in Abbildung 4 die a-priori Dichte für den summierten wöchentlichen Gesamzeitaufwand über alle Fächergruppen, die prädiktive a-posteriori Wahrscheinlichkeitsdichte sowie das zugehörige 95 %-hpd-Intervall dargestellt.

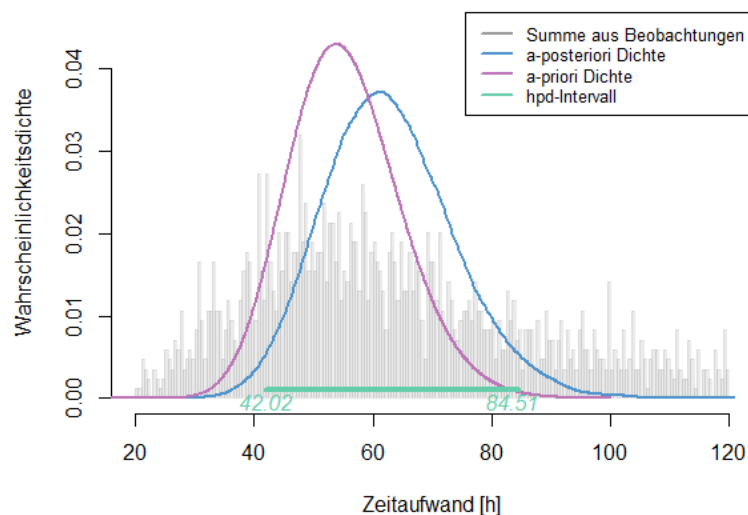


Abb. 4: Vergleich von prädiktiver a-posteriori Dichte mit a-priori Dichte sowie 95 %-hpd-Intervall für den summierten wöchentlichen Gesamzeitaufwand für alle Fächergruppen zusammen.

Die Verteilung ist, wie erwartet, leicht rechtsschief. Es fällt auf, dass die Dichten deutlich schmäler sind als die Streuung der realisierten Werte. Allerdings bewirken die Realisierungen eine Verschiebung der a-priori Dichte und eine leichte Vergrößerung der Streuung bei der prädiktiven a-posteriori Dichte. Die summierten Gesamtarbeitszeiten erscheinen außerdem ungefähr 7 h größer als bei der Selbstschätzung der Gesamtarbeitszeit. Bei der Bayesianischen Analyse zeigt sich also, dass Messungen für die Gesamtarbeitszeit auf der Basis von unabhängig voneinander ermittelten Arbeitszeiten für Teilaufgaben nur dann verlässlich sind, wenn Vorinformationen zu der Variation der Gesamtarbeitszeit vorliegen, weil die Summe der Teilarbeitszeiten nicht nur tendenziell größer ist als die Gesamtarbeitszeitschätzung, sondern auch deutlich stärker variiert. Ein Grund dafür könnte die fehlende Übersicht über die insgesamt angegebene Arbeitszeit sein während der Angaben der Teilarbeitszeiten. Ein Mitführen eines Summenzählers während des Ausfüllens des Fragebogens könnte hier Abhilfe schaffen.

Tabelle 7 beinhaltet den prädiktiven a-posteriori Mittelwert und das 95 %-hpd-Intervall als Prognoseintervall für die Summen des geschätzten wöchentlichen Zeitaufwands der einzelnen Fächergruppen und

insgesamt sowie von Professorinnen und Professoren. Besonders auffällig sind die deutlich kleineren Prognoseintervalle als im frequentistischen Fall, der kleine Anteil der *Lehre* bei den *Lebenswissenschaften*, der Betreuung *wissenschaftlichen Nachwuchses* bei den *Geistes- und Sozialwissenschaften* und der *Forschung* bei den *Ingenieurwissenschaften*. In Abschnitt 6.5 testen wir die mittleren Ergebnisse auf signifikante Unterschiede.

Tab. 7: Prädiktiver a-posteriori Mittelwert und 95 %-*hpd* Prognoseintervall für den wöchentlichen Zeitaufwand in *h*

| | Gesamt | Lehre | Wiss.Nachwuchs | Administration |
|-----------------------|--------------------|-------------------|---------------------|-------------------|
| Alle Fächergruppen | 62,7 [42,0 , 84,5] | 14,0 [5,4 , 23,4] | 7,0 [0,0 , 19,1] | 10,8 [2,3 , 21,0] |
| Geistes-/Sozial-Wiss. | 62,5 [41,9 , 84,0] | 16,7 [6,9 , 27,9] | 4,6 [0,0 , 12,1] | 10,2 [2,3 , 19,2] |
| Lebens-Wiss. | 59,1 [39,9 , 79,5] | 9,3 [4,1 , 15,2] | 7,7 [0,0 , 19,7] | 10,6 [2,7 , 19,8] |
| Natur-Wiss. | 61,1 [41,1 , 81,8] | 12,9 [5,5 , 21,2] | 8 [0,0 , 20,4] | 10,1 [2,3 , 19,4] |
| Ingenieur-Wiss. | 62,9 [41,9 , 83,6] | 14,7 [6,4 , 24,2] | 9,0 [0,0 , 23,6] | 12,0 [2,9 , 22,3] |
| Professorinnen | 60,4 [40,6 , 81,0] | 15,6 [6,2 , 25,9] | 5,6 [0,0 , 14,7] | 10,0 [2,3 , 18,8] |
| Professoren | 62,3 [41,9 , 83,9] | 13,5 [5,2 , 22,5] | 7,6 [0,0 , 20,7] | 10,5 [2,0 , 20,3] |
| | Drittmittel | Forschung | Tagungen & Vorträge | Sonstiges |
| Alle Fächergruppen | 5,1 [1,0 , 9,9] | 18,2 [4,5 , 34,4] | 6,5 [0,2 , 15,2] | 1,2 [0,0 , 4,8] |
| Geistes-/Sozial-Wiss. | 3,8 [0,8 , 7,4] | 19,3 [6,1 , 34,9] | 7,1 [0,3 , 16,7] | 0,8 [0,0 , 2,8] |
| Lebens-Wiss. | 5,6 [1,6 , 10,1] | 18,4 [5,4 , 33,2] | 5,9 [0,5 , 12,2] | 1,5 [0,0 , 5,1] |
| Natur-Wiss. | 5,4 [1,4 , 10,1] | 18,1 [5,6 , 32,4] | 5,9 [0,5 , 12,9] | 0,8 [0,0 , 3,0] |
| Ingenieur-Wiss. | 7,4 [2,1 , 13,3] | 12,8 [4,6 , 22,7] | 5,9 [0,6 , 12,6] | 1,1 [0,0 , 3,4] |
| Professorinnen | 4,3 [1,0 , 8,1] | 17,5 [5,0 , 31,9] | 6,4 [0,3 , 14,5] | 1,2 [0,0 , 4,0] |
| Professoren | 5,1 [1,1 , 9,8] | 18,1 [4,6 , 34,1] | 6,4 [0,2 , 14,8] | 1,1 [0,0 , 4,3] |

6.4.3 Prozentanteile

Betrachtet man die Prozentanteile der verschiedenen Teilaufgaben in Tabelle 8, so fallen der kleine Anteil der Betreuung des *wissenschaftlichen Nachwuchses* bei den *Geistes- und Sozialwissenschaften* und die kleine Summe von *Drittmitteln* und *Forschung* bei den *Ingenieurwissenschaften* besonders auf. Die Bayesianisch geschätzten Anteile der Teilaufgaben an der Gesamtarbeitszeit unterscheiden sich nur wenig von den frequentistisch bestimmten. Die Unsicherheitsbereiche der Prozentsätze zeigen, dass die wahren Werte im Allg. um höchstens rund 10 % vom Mittelwert abweichen. Die größten Abweichungen ergeben sich allgemein bei der Kategorie *Sonstiges* wegen der kleinen Prozentwerte und abgesehen davon speziell bei der Fächergruppe *Lebenswissenschaften*, wo die kleinste Stichprobe vorliegt, maximal mit einer Halbbreite des Unsicherheitsbereichs von 10,1 % des Mittelwerts (bei der Kategorie *Wiss.Nachwuchs*). Der direkte Forschungsanteil an der Arbeitszeit ist mit im Mittel 29 – 30 % deutlich höher als der Anteil der Lehre mit 22 – 24 %. Man beachte, dass sich die Bayesianisch ermittelten Anteile deutlich von den verwendeten a-priori Anteilen unterscheiden (s. Abschnitt 5.3), z.B. wurden in unserer Analyse 23,0 % für Lehre und 28,3 % für Forschung ermittelt im Gegensatz zu 33,1 % für Lehre und 19,4 % für Forschung a-priori. Das zeigt die Flexibilität des Bayes Ansatzes.

Insgesamt umfassen forschungsnahe Tätigkeiten wie Betreuung des Wissenschaftlichen Nachwuchses, Drittmittelinwerbung, Forschung und Publikation sowie Ausrichten und Besuch von Tagungen rund 60 % der Arbeitszeit, rund 23 % der Arbeitszeit werden für Lehre und Betreuung und Prüfung von Studierenden verwendet und rund 17 % für administrative Tätigkeiten.

Tab. 8: Mittlere prozentuale Anteile an der wöchentlichen Gesamtarbeitszeit

| frequentistisch (in %) | Lehre | Wiss.Nachwuchs | Administr. | Drittmittel | Forschung | Tagungen | Sonstiges |
|------------------------|----------------------|----------------------|----------------------|----------------------|----------------------|----------------------|-------------------|
| Alle Fächergruppen | 22,8 [22,2, 23,4] | 9,4 [9,0, 9,8] | 17,2 [16,7, 17,8] | 9,6 [9,2, 9,9] | 29,6 [28,9, 30,3] | 9,5 [9,2, 9,8] | 2,0 [1,7, 2,3] |
| Geistes-/Sozial-Wiss. | 25,3 [24,3, 26,4] | 6,6 [6,1, 7,1] | 16,0 [15,2, 16,7] | 7,7 [7,2, 8,1] | 33,0 [31,8, 34,2] | 10,0 [9,4, 10,5] | 1,5 [1,2, 1,9] |
| Lebens-Wiss. | 17,7 [16,1, 19,3] | 10,8 [9,7, 11,9] | 18,3 [16,8, 19,9] | 10,9 [9,8, 12,0] | 29,5 [27,7, 31,4] | 9,3 [8,4, 10,1] | 3,7 [2,4, 5,0] |
| Natur-Wiss. | 22,0 [21,0, 23,0] | 11,1 [10,4, 11,8] | 16,9 [16,0, 17,8] | 9,6 [9,0, 10,1] | 29,7 [28,5, 30,9] | 9,3 [8,8, 9,8] | 1,6 [1,2, 2,0] |
| Ing.-Wiss. | 23,1 [21,5, 24,6] | 11,5 [10,5, 12,6] | 19,9 [18,5, 21,4] | 12,9 [11,9, 13,9] | 21,2 [19,7, 22,6] | 9,1 [8,3, 9,8] | 2,5 [1,7, 3,2] |
| Professorinnen | 24,4 [23,1, 25,6] | 8,5 [7,7, 9,2] | 16,7 [15,7, 17,6] | 9,1 [8,5, 9,7] | 29,4 [28,0, 30,8] | 9,7 [9,0, 10,3] | 2,5 [1,9, 3,1] |
| Professoren | 22,3 [21,6, 23,0] | 9,7 [9,3, 10,2] | 17,4 [16,8, 18,1] | 9,7 [9,3, 10,1] | 29,6 [28,8, 30,5] | 9,5 [9,1, 9,8] | 1,8 [1,5, 2,1] |
| Bayesianisch (in %) | | | | | | | |
| Alle Fächergruppen | 23,0 [22,5, 23,4] | 10,9 [10,5, 11,3] | 17,3 [16,9, 17,8] | 8,1 [7,9, 8,4] | 28,3 [27,7, 28,9] | 10,4 [10,1, 10,7] | 1,9 [1,8, 2,0] |

6.4.4 Sensitivitätsanalyse für das Bayes-Modell

Anhand einer Sensitivitätsanalyse wurde untersucht, wie sensitiv die Auswertungsergebnisse für den summierten wöchentlichen Gesamtzeitaufwand Y_{Summe} auf die informativen Einstellungsparameter fsd_2 , fsd_{11} , fsd_{20a} , fsd_{20b} , $Var(\alpha_2)$, $Var(\alpha_{11})$, $Var(\alpha_{20a})$, $Var(\alpha_{20b})$ und $Var(\alpha_{Gesamt})$ reagieren. Es wird ein Plackett-Burman-Plan mit 12 Versuchen verwendet. Die Niveaus der fsd -Parameter werden auf $\pm 0,1$ gegenüber dem in Abschnitt 6.4.2 verwendeten Niveau festgelegt, die Niveaus der anderen Parameter auf 0,01 und 0,03. In dem Übereinstimmungsmaß d werden als Referenzwerte die Ergebnisse aus Abschnitt 6.4.2 verwendet:

$$d = \left| 1 - \frac{\mu_z}{\mu_{ref}} \cdot \frac{hpd_r - hpd_l}{hpd_{r,ref} - hpd_{l,ref}} \right|. \quad (8)$$

Je näher der Wert für d bei 0 liegt, desto größer die Übereinstimmung der beiden betrachteten prädiktiven a-posteriori Verteilungen. Die Ergebnisse für d liegen alle zwischen 0,0 und 0,032. Keiner der Faktoren hat sich dabei auf dem 5 %-Niveau als statistisch signifikant erwiesen (kleinster p-Wert ist $\approx 15\%$ für $Var(\alpha_{Gesamt})$).

Die Variation der Ergebnisse für den wöchentlichen Zeitaufwand lässt sich auch anhand der Variationskoeffizienten cv_{mean} für den a-posteriori Mittelwert ($< 0,4\%$) und cv_{hpd} für die hpd -Breite ($< 1,3\%$) charakterisieren. Eine Veränderung der Einstellungsparameter für die informativen Variablen führt also im Rahmen der im Versuchsplan aufgestellten Grenzen zu einer Variation der Auswertungsergebnisse zum wöchentlichen Gesamtzeitaufwand von weniger als 1,5 %, gemessen an den prädiktiven a-posteriori Mittelwerten und den hpd -Breiten.

6.5 Signifikante Unterschiede beim mittleren wöchentlichen Zeitaufwand

6.5.1 Fächergruppen

Zur Bestimmung signifikanter Unterschiede zwischen den Erwartungswerten zweier Fächergruppen wurden Erwartungswertdifferenzen gegen Null getestet. Dabei wurden sowohl im frequentistischen als auch im Bayesianischen Fall Konfidenzintervalle der Mittelwertdifferenzen als Unsicherheitsintervalle verwendet, weil Mittelwerte tendenziell normalverteilt sind. Falls das Konfidenzintervall die Null nicht enthält, kann man davon ausgehen, dass die Erwartungswerte der Fächergruppen signifikant voneinander abweichen. Im Bayesianischen Fall werden Mittelwerte und Konfidenzbereiche der prädiktiven a-posteriori Verteilung verwendet.

Die Differenzen der Mittelwerte des zeitlichen Gesamtaufwands pro Woche wurden auf Basis der Gesamtschätzungen (vgl. Tabelle 5) und auf Basis der Summen der Teilarbeitszeiten (vgl. Tabellen 6,7) berechnet und die Ergebnisse sind der Größe nach geordnet in den Tabellen 9 und 10 aufgeführt, sowohl für den frequentistischen als auch den Bayesianischen Fall.

Auf der Basis der Gesamtschätzungen ergeben sich sowohl im frequentistischen als auch im Bayesianischen Fall nur bei den Vergleichen von *Lebenswissenschaften* mit *Geist-/Sozial-* und *Naturwissenschaften* signifikante Unterschiede.

Auf der Basis von Summen von Teilarbeitszeiten ergeben sich im frequentistischen Fall nur bei den Vergleichen von *Geistes- und Sozial-* mit *Lebens-* bzw. *Naturwissenschaften* signifikante Unterschiede. Das Verfahren von Ryan (s. Abschnitt 2.3) findet die homogene Gruppe $\{\text{Lebens-, Ingenieur- und Naturwissenschaften}\}$, deren Fächergruppen sich also nicht signifikant unterscheiden. Bei der Bayesianischen Analyse sind im Gegensatz zur frequentistischen Analyse alle Differenzen signifikant außer der zwischen *Ingenieur-* und *Geistes/Sozialwissenschaften*.

Auffällig ist, dass die *Geistes-/Sozial-* und *Lebenswissenschaften* bei Gesamtschätzungen und Summen der Teilarbeitszeiten insofern die Rollen tauschen, dass einmal die *Lebenswissenschaften* den signifikant größten Mittelwert aufweisen und die *Geistes-/Sozialwissenschaften* den kleinsten, während das Ergebnis bei den Summen umgekehrt ist. Dabei fällt auf, dass bei den *Lebenswissenschaften* Gesamtschätzung und Summe relativ gut übereinstimmen, während sie bei den anderen Fächergruppen, insbesondere bei den *Geistes-/Sozialwissenschaften*, deutlich stärker abweichen (vgl. auch Tabelle 12).

Tab. 9: Mittelwert und Konfidenzintervall für den Unterschied zwischen jeweils zwei Fächergruppen für den wöchentlichen Gesamtzeitaufwand in h auf der Basis der Gesamtschätzungen

| frequentistisch | | | Bayesianisch | | |
|------------------------------------|--------------------|---------------------------------|------------------------------------|--------------------|---------------------------------|
| Fächergruppe 1 - Fächergruppe 2 | Mittelwert (h) | 95 %-Konfidenzintervall (h) | Fächergruppe 1 - Fächergruppe 2 | Mittelwert (h) | 95 %-Konfidenzintervall (h) |
| Leben - Geist/Sozial | 2,1 | [0,7 , 3,4] | Leben - Geist/Sozial | 2,1 | [0,4 , 3,7] |
| Leben - Natur | 1,7 | [0,4 , 3,1] | Leben - Natur | 1,8 | [0,1 , 3,4] |
| Leben - Ing. | 1,4 | [-0,4 , 3,2] | Leben - Ing. | 1,4 | [-0,8 , 3,5] |
| Ing. - Geist/Sozial | 0,7 | [-0,6 , 1,9] | Ing. - Geist/Sozial | 0,7 | [-0,8 , 2,2] |
| Natur - Geist/Sozial | 0,4 | [-0,8 , 1,5] | Ing. - Natur | 0,4 | [-1,4 , 2,2] |
| Ing. - Natur | 0,3 | [-1,1 , 1,8] | Natur - Geist/Sozial | 0,3 | [-1,0 , 1,6] |

Tab. 10: Mittelwert und Konfidenzintervall für den Unterschied zwischen jeweils zwei Fächergruppen für den wöchentlichen Gesamtzeitaufwand in h auf der Basis der summierten Teilarbeitszeiten

| frequentistisch | | | Bayesianisch | | |
|---------------------------------|--------------------|---------------------------------|---------------------------------|--------------------|---------------------------------|
| Fächergruppe 1 - Fächergruppe 2 | Mittelwert (h) | 95 %-Konfidenzintervall (h) | Fächergruppe 1 - Fächergruppe 2 | Mittelwert (h) | 95 %-Konfidenzintervall (h) |
| Geist/Sozial - Leben | 5,0 | [1,6 , 8,4] | Ing. - Leben | 3,8 | [2,0 , 5,7] |
| Geist/Sozial - Natur | 3,5 | [0,9 , 6,0] | Geist/Sozial - Leben | 3,5 | [1,9 , 5,0] |
| Ing. - Leben | 2,6 | [-1,5 , 6,6] | Natur - Leben | 2,1 | [0,3 , 3,8] |
| Geist/Sozial - Ing. | 2,5 | [-1,3 , 6,3] | Ing. - Natur | 1,8 | [0,2 , 3,4] |
| Natur - Leben | 1,6 | [-2,4 , 5,5] | Geist/Sozial - Natur | 1,4 | [0,0 , 2,8] |
| Ing. - Natur | 1,0 | [-2,9 , 4,8] | Ing. - Geist/Sozial | 0,4 | [-1,4 , 2,2] |

Des Weiteren wurden die Mittelwertdifferenzen jeweils für die Tätigkeitsfelder Lehre und Forschung verglichen (vgl. Tabelle 11). Auffällig ist, dass für die *Lehre* alle Differenzen signifikant sind außer im frequentistischen Fall zwischen *Ingenieur-* und *Naturwissenschaften*. Für die *Forschung* ergeben sich Nichtsignifikanzen im frequentistischen Fall zwischen *Natur-* und *Lebenswissenschaften* und im Bayesianischen Fall zwischen *Lebenswissenschaften* und *Geistes-/Sozial-* sowie *Naturwissenschaften*. Es gibt also nur kleine Unterschiede zwischen frequentistischer und Bayesianischer Analyse.

Tab. 11: Mittelwert und Konfidenzintervall für den Unterschied zwischen jeweils zwei Fächergruppen für den wöchentlichen Zeitaufwand für Lehre bzw. Forschung in h

| frequentistisch | | | | | |
|---------------------------------|--------------------|---------------------------------|---------------------------------|--------------------|---------------------------------|
| Fächergruppe 1 - Fächergruppe 2 | Lehre | | Fächergruppe 1 - Fächergruppe 2 | Forschung | |
| | Mittelwert (h) | 95 %-Konfidenzintervall (h) | | Mittelwert (h) | 95 %-Konfidenzintervall (h) |
| Geist/Sozial - Leben | 5,9 | [4,5 , 7,4] | Geist/Sozial - Ing. | 8,6 | [7,1 , 10,1] |
| Ing. - Leben | 3,7 | [2,0 , 5,3] | Natur - Ing. | 4,6 | [3,2 , 6,0] |
| Natur - Leben | 3,2 | [1,6 , 4,9] | Leben - Ing. | 4,3 | [2,3 , 6,3] |
| Geist/Sozial - Natur | 2,7 | [1,5 , 4,0] | Geist/Sozial - Leben | 4,3 | [2,6 , 6,1] |
| Geist/Sozial - Ing. | 2,3 | [0,6 , 4,0] | Geist/Sozial - Natur | 4,0 | [2,4 , 5,6] |
| Ing. - Natur | 0,5 | [-1,2 , 2,1] | Natur - Leben | 0,3 | [-1,6 , 2,2] |
| Bayesianisch | | | | | |
| Fächergruppe 1 - Fächergruppe 2 | Lehre | | Fächergruppe 1 - Fächergruppe 2 | Forschung | |
| | Mittelwert (h) | 95 %-Konfidenzintervall (h) | | Mittelwert (h) | 95 %-Konfidenzintervall (h) |
| Geist/Sozial - Leben | 7,3 | [6,8 , 7,9] | Geist/Sozial - Ing. | 6,6 | [5,7 , 7,4] |
| Ing. - Leben | 5,3 | [4,7 , 6,0] | Leben - Ing. | 5,7 | [4,5 , 6,8] |
| Geist/Sozial - Natur | 3,8 | [3,3 , 4,4] | Natur - Ing. | 5,3 | [4,3 , 6,2] |
| Natur - Leben | 3,5 | [3,0 , 4,1] | Geist/Sozial - Natur | 1,3 | [0,4 , 2,1] |
| Geist/Sozial - Ing. | 2,0 | [1,2 , 2,8] | Geist/Sozial - Leben | 0,9 | [-0,4 , 2,2] |
| Ing. - Natur | 1,8 | [1,1 , 2,6] | Leben - Natur | 0,4 | [-0,9 , 1,7] |

Auch Erwartungswerte von Kombinationen von Teilaufgaben zeigen signifikante Abweichungen, aber auch lineare oder nichtlineare Zusammenhänge. Zum Beispiel sieht man grafisch deutliche signifikante Unterschiede und einen fast linearen Zusammenhang bei den Mittelwerten von *Administration + Dritt-*

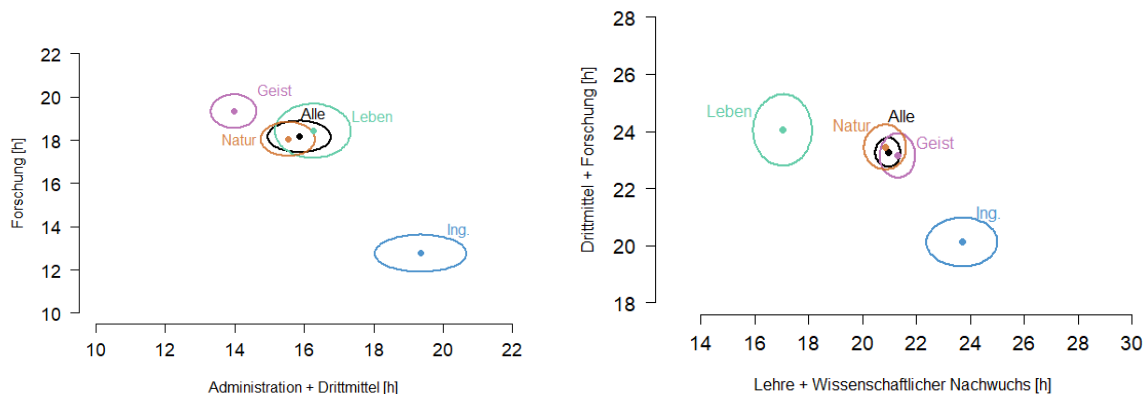


Abb. 5: Zusammenhang zwischen den Mittelwerten (mit 95 %-Konfidenzbereichen) von *Administration + Drittmittel* und *Forschung* (links) sowie *Lehre + Wiss.Nachwuchs* und *Drittmittel + Forschung* (rechts) (jeweils Bayesianisch).

mittel und *Forschung* (vgl. Abbildung 5 (links)) bzw. einen nichtlinearen Zusammenhang bei *Lehre + Wiss.Nachwuchs* und *Drittmittel + Forschung* (vgl. Abbildung 5 (rechts)). Beachte die immer deutlich abweichende Stellung der *Ingenieurwissenschaften* mit relativ wenig *Forschung* sowie viel *Administration*, *Drittmittelbeschaffungsaufwand*, *Lehre* und *Betreuung von Wissenschaftlichem Nachwuchs* (vgl. auch Tabelle 7).

6.5.2 Professorinnen und Professoren

Auch für Professorinnen und Professoren wurden jeweils Mittelwert und Konfidenzintervall der Differenzen der Erwartungswerte gemäß Abschnitt 2.3 berechnet und auf der Basis der Gesamtschätzungen bzw. der aufsummierten Teilarbeitszeiten gegen Null getestet. Der frequentistische Mittelwert für den Unterschied zwischen Professoren und Professorinnen liegt bei $1,3 h$. Das 95 %-Konfidenzintervall beträgt $[0,3, 2,2] h$ auf Basis der Gesamtschätzungen bzw. $[-1,3, 3,8] h$ auf Basis der aufsummierten Teilarbeitszeiten. Die Verwendung der Gesamtschätzungen ist also trennschärfer. Bei der Bayes-Analyse liegt die a-posteriori Mittelwertdifferenz auf der Basis der Gesamtschätzungen bei $1,3 h$ mit einem 95 %-Prognoseintervall von $[0,2, 2,4] h$ und auf der Basis von aufsummierten Teilarbeitszeiten bei $1,9 h$ mit einem 95 %-Prognoseintervall von $[0,7, 3,1] h$, was beides für einen signifikanten Unterschied spricht. Insgesamt scheinen Professorinnen möglicherweise im Mittel ein bis zwei Stunden weniger zu arbeiten als Professoren.

6.6 Vergleich von Gesamtschätzung und Summe

Schließlich sollen noch die Gesamtschätzungen der wöchentlichen Arbeitszeit mit den Summationsergebnissen verglichen werden (s. Tabelle 12). Die Abweichung ist bei den *Lebenswissenschaften* am kleinsten und bei den *Geistes-/Sozialwissenschaften* deutlich am größten (vgl. Abschnitt 6.5).

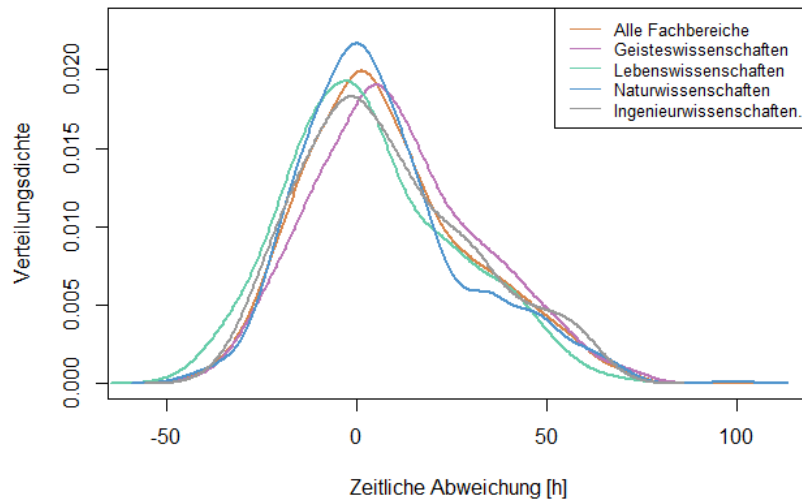


Abb. 6: Abweichung der Summe von der Gesamtschätzung (Dichteschätzungen mit Gaußkern in R (2016)).

Tab. 12: Mittelwert der Gesamtschätzung, Abweichung der Mittelwerte und Mittelwert der Abweichungen (jeweils Summe - Gesamtschätzung) in h

| | Mittelwert | Abweichung der Mittelwerte | Mittelwert der Abweichungen |
|-----------------------|------------|----------------------------|-----------------------------|
| Alle Fächergruppen | 56,4 | 6,8 | 8,2 |
| Geistes-/Sozial-Wiss. | 55,9 | 9,5 | 10,8 |
| Lebens-Wiss. | 58,0 | 2,4 | 3,9 |
| Natur-Wiss. | 56,2 | 5,6 | 6,8 |
| Ingenieur-Wiss. | 56,6 | 6,3 | 8,3 |
| Professorinnen | 55,5 | 6,7 | 7,9 |
| Professoren | 56,7 | 6,8 | 8,3 |

Die Verteilung der Abweichungen ist in Abb. 6 dargestellt (vgl. Spalte *Mittelwert der Abweichungen* in Tabelle 12). Es wird deutlich, dass die Gesamtschätzung bei vielen Teilnehmern nicht mit dem Summationsergebnis übereinstimmt. Zu beobachten ist eine leicht rechts-schiefe Verteilung, d.h. in der Gesamtschätzung der Arbeitszeit kommen Überschätzungen seltener vor als Unterschätzungen. Tatsächlich wird der wöchentlich investierte Zeitaufwand in der Gesamtschätzung teilweise sehr deutlich unterschätzt bzw. die Einzelaufwände deutlich überschätzt. Die Einzelaufwände könnten insbesondere deshalb überschätzt werden, weil in dem Fragebogen keine Summenkontrolle der Einzelaufwände vorgesehen war!

7 Zusammenfassung und Diskussion

Die Analyse der gültigen Fragebögen von aktiven Vollzeit arbeitenden Universitätsprofessorinnen und -professoren ergibt bei der Gesamtschätzung der durchschnittlichen wöchentlichen Arbeitszeit eine mittlere Arbeitszeit von $56 h$ mit 95%-Prognoseintervallen zwischen $35 h$ und $80 h$. Frequentistische und

Bayesianische Analyse führen zu sehr ähnlichen Ergebnissen. Auch die betrachteten Fächergruppen und die Geschlechter unterscheiden sich nur wenig.

Die Summen der Arbeitszeiten der Teilaufgaben sind deutlich größer mit im Mittel 63 h. Die 95%-Prognoseintervalle unterscheiden sich deutlich im frequentistischen Fall mit [27,5, 113,1] h und Bayesianischen Fall mit [42,0, 84,5] h. Offenbar bewirkt die Vorinformation der Varianz der Selbstauskunft eine deutliche Verkleinerung der Variation der geschätzten Summen! Die summierten Gesamtarbeitszeiten erscheinen aber ungefähr 7 h größer als bei der Selbstauskunft angegeben.

Bei dem Vergleich von frequentistischer und Bayesianischer Analyse zeigt sich also, dass Messungen für die Gesamtarbeitszeit auf der Basis von unabhängig voneinander ermittelten Arbeitszeiten für Teilaufgaben nur dann verlässlich sind, wenn Vorinformationen zu der Variation der Gesamtarbeitszeit vorliegen und eine Bayesianische Analyse durchgeführt wird, weil die Summe der Teilarbeitszeiten nicht nur tendenziell größer ist als die realistische Gesamtarbeitszeitschätzung, sondern auch deutlich stärker variiert. Ein Grund dafür könnte die fehlende Übersicht über die insgesamt angegebene Arbeitszeit sein während der Angaben der Teilarbeitszeiten. Ein Mitführen eines Summenzählers während des Ausfüllens des Fragebogens könnte hier Abhilfe schaffen.

Bei den Prozentanteilen der verschiedenen Teilaufgaben fallen der kleine Anteil der Betreuung des *wissenschaftlichen Nachwuchses* bei den *Geistes- und Sozialwissenschaften* und die kleine Summe von *Drittmitteln* und *Forschung* bei den *Ingenieurwissenschaften* besonders auf. Der Anteil forschungsnaher Tätigkeiten an der Arbeitszeit ist mit etwa 60 % deutlich höher als der Anteil der Lehre mit 23 % und der Anteil administrativer Tätigkeiten mit 17 %.

Die Unterschiede in den Erwartungswerten der Fächergruppen und Geschlechter werden auch auf Signifikanz getestet. Dabei fällt auf, dass die größten signifikanten Differenzen immer zwischen den *Geistes-/Sozialwissenschaften* und einer der anderen Fächergruppen auftreten. Der Unterschied zwischen dem erwarteten Gesamtarbeitsaufwand von Professorinnen und Professoren liegt bei 1 – 2 h.

Abschließend sollen unsere Ergebnisse noch mit dem Ergebnis einer älteren Studie verglichen werden, die in Abschnitt 3 kurz vorgestellt, aber nicht als Vorinformation verwendet wurde. Aus Tabelle 13 geht hervor, dass die Angaben in der DHV-Studie aus dem Jahre 2007 tendenziell höher lagen als in der aktuellen Studie. Lediglich der Prozentsatz der Angaben zwischen 51 und 60 h ist ähnlich.

Tab. 13: Vergleich der Gesamtschätzungen dieser Studie und der DHV-Studie

| | Aktuelle Studie (%) | DHV-Studie (%) |
|-----------|---------------------|----------------|
| bis 40 h | 0,6 | 0,38 |
| 41 – 50 h | 35,0 | 11,1 |
| 51 – 60 h | 42,7 | 42,0 |
| 61 – 70 h | 13,3 | 31,7 |
| über 70 h | 5,4 | 14,82 |

Die umfangreiche Datengrundlage dieser Befragung bietet Potenzial für die Untersuchungen diverser weiterer Fragestellungen. Hier wurden als Anwendungsbeispiele nur die sehr grobe Aufteilung in vier Fächergruppen bzw. die beiden Geschlechter und in die sieben Tätigkeitsfelder verwendet, was zu einer hohen Streuung und vermutlich zu einer Vermischung verschiedener Effekte führt. Es gibt sicher noch große Unterschiede innerhalb der vier Fächergruppen. Eine feinere Unterteilung, z.B. in Fächer, würde vermutlich zu engeren Prognoseintervallen führen und weitere Effekte herausstellen wie beispielsweise welche Tätigkeitsfelder in welchen Bereichen eine besondere Rolle spielen (z.B. Labortätigkeiten in der Chemie, der Biologie und den Chemie- und Bioingenieurwissenschaften). Es empfiehlt sich daher,

im Rahmen weiterer Analysen, weitere konkrete Fragestellungen zu betrachten und diese basierend auf einer feineren Fächergruppeneinteilung, einer feineren Fragenkomplexeinteilung oder für spezielle Bundesländer oder andere Hochschularten zu analysieren.

Das Statistische Bundesamt hat 2016 eine Studie mit ähnlicher Zielsetzung durchgeführt. Eine erste (methodische) Veröffentlichung dazu findet sich in Statistisches Bundesamt (2017c).

Danksagung: Wir danken unseren Kooperationspartnern bei der Wochenzeitung Die ZEIT und ZEIT-online, insbesondere Sascha Venohr und Paul Blickle, sowie den Studierenden in den Studiengängen Wissenschaftsjournalismus, Journalistik und Statistik der TU Dortmund, die im Rahmen von Lehrforschungsprojekten an der Vorbereitung der Studie beteiligt waren. Für die Durchführung von qualitativen Pretests danken wir hier insbesondere Marleen Halbach, Alia Khaddour, Franziska Lehnert, Lara Malberger, Eva Mühle und Vanessa Reske. Wir danken ferner Dr. Uwe Ligges (TU Dortmund) für die technische Unterstützung bei der Durchführung der Umfrage und Prof. Dr. Katja Ickstadt (TU Dortmund) für die kritische Kommentierung der Bayes-Analyse.

Literatur

- Bernardo, J., Smith, A.F.M. (2000) *Bayesian Theory*, Wiley
- Böhmer, S., Neufeld, J., Hinze, S., Klode, C., Hornbostel, S. (2011) *Wissenschaftlerbefragung 2010: Forschungsbedingungen von Professorinnen und Professoren an Deutschen Universitäten*, iFQ-Working Paper No. 8; URL www.forschungsinform.de/publikationen/download/working_paper_8_2010.pdf
- Brooks, S.P., Gelman A. (1998): *Alternative Methods for Monitoring Convergence of Iterative Simulations*, Journal of Computational and Graphical Statistics, 7, 434-455.
- Deutscher Bundestag (2014) *Die Grundrechte*, Grundgesetz der Bundesrepublik Deutschland, Artikel 5, Absatz 3, Berlin, URL <http://www.bundestag.de/bundestag/aufgaben/rechtsgrundlagen/grundgesetz/grundgesetz/197094>
- Die Bundesregierung (2014) *Attraktiver Forschungsstandort Deutschland*, Forschung, Berlin, URL <http://www.bundesregierung.de/Content/DE/Infodienst/2014/02/2014-02-26-forschung-und-innovation/infodienst-forschung-innovation.html>
- Gemeinsame Wissenschaftskonferenz (2017): Hochschulpakt 2020. Umsetzung in der zweiten Programmphase. URL <http://www.gwk-bonn.de/fileadmin/Papers/GWK-Heft-54-Hochschulpakt-Umsetzung-Programmphase-2011-2015.pdf>
- Hartmer, M. (2008) *Hochschullehrer wollen forschen und lehren, nicht verwalten*, Forschung & Lehre, Ausgabe 02/2008, Deutscher Hochschulverband, Bonn
- Kruschke, J.K. (2010) *Doing Bayesian Data Analysis: A Tutorial with R and BUGS*, Academic Press, Elsevier, UK
- Lüpsen, H. (2016) *Multiple Mittelwertvergleiche - parametrisch und nichtparametrisch - sowie α -Adjustierungen mit praktischen Anwendungen mit R und SPSS*, Version 0.91, Regionales Rechenzentrum (RRZK), Universität zu Köln
- Marques, F.J. (2012) *On the product of independent Generalized Gamma random variables*, Discussion Paper 19-2012, CMA-FCT-Universidade Nova de Lisboa; URL https://www.cma.fct.unl.pt/sites/www.cma.fct.unl.pt/files/documentos/publicacoes/pdf_2012/CMA_2019-2012.pdf
- Milic, T. (2015) *Wie aussagekräftig sind die Resultate von Online-Umfragen?* ZDA - Zentrum für Demokratie Aarau, Universität Zürich, Fachhochschule Nordwestschweiz, Stadt Aarau, Kanton Aarau; URL http://www.preferencematcher.com/edc/wp-content/uploads/2015/07/Milic_Onlineumfragen.pdf
- Moschopoulos, P.G. (1985) *The distribution of the sum of independent gamma random variables*, Ann. Inst. Statist. Math., 37, Part A, 541-544

- R Development Core Team (2016) *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.
- Roberts, G.O. (1996) *Markov Chain Concepts Related to Sampling Algorithms*, in *Markov Chain Monte Carlo in Practice* (eds Gilks, W. R., Richardson, R. und Spiegelhalter, D. J.), Chapman & Hall, London, UK
- Schomburg, H., Flöther, C., Wolf, V. (2012) *Wandel von Lehre und Studium an deutschen Hochschulen - Erfahrungen und Sichtweisen der Lehrenden*, Projektbericht, Kassel: Internationales Zentrum für Hochschulforschung (INCHER-Kassel), Universität Kassel; URL https://www.hrk-nexus.de/uploads/media/HRK_nexus_LESSI.pdf
- Spiegelhalter, D.J., Abrams, K.R., Myles, J.P. (2004) *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*, John Wiley and Sons Ltd, UK
- statista (2017) *Anzahl der hauptberuflichen Professoren und Professorinnen an deutschen Hochschulen 1999 bis 2015*, URL <https://de.statista.com/statistik/daten/studie/160365/umfrage/professoren-und-professorinnen-an-deutschen-hochschulen/>
- Statistisches Bundesamt (Hrsg.)(2016a), *Promovierende in Deutschland*, Wiesbaden, URL <https://www.destatis.de/DE/Publikationen/Thematisch/BildungForschungKultur/Hochschulen/Promovierende.html>
- Statistisches Bundesamt (Hrsg.) (2017), *Bildung, Forschung, Kultur*, Wiesbaden, URL <https://www.destatis.de/DE/Publikationen/Thematisch/BildungForschungKultur/Hochschulen/BroschuereHochschulenBlick.html>
- Statistisches Bundesamt (Hrsg.)(2017a), *Studierende*, URL https://www.destatis.de/DE/ZahlenFakten/Indikatoren/LangeReihen/Bildung/Irbil01.html?cms_gtp=152374_list%253D1&https=1
- Statistisches Bundesamt (Hrsg.)(2017b), *Personal an Hochschulen*, URL <https://www.destatis.de/DE/ZahlenFakten/GesellschaftStaat/BildungForschungKultur/Hochschulen/Tabellen/PersonalHochschulen.html>
- Statistisches Bundesamt (Hrsg.)(2017c), *Methoden-Verfahren-Entwicklungen. Nachrichten aus dem Statistischen Bundesamt, Ausgabe 1/2017*, URL https://www.destatis.de/DE/Methoden/Methodenpapiere/Download/01_2017.pdf?__blob=publicationFile
- Thomas, A., O'Hara, B., Ligges, U., Sturtz, S. (2006) *Making BUGS Open*, R News 6(1), 12–17
- Weihs, C., Jessenberger, J. (1999): *Statistische Methoden zur Qualitätssicherung und -optimierung*, Wiley-VCH, Weinheim

