

Überlebenszeitanalyse und Risikoprädiktion

Dr. Turid Frahnw

Wintersemester 2020/2021

E-mail: turid.frahnw@uni-bielefeld.de
Bürozeiten: Nach Vereinbarung
Büro: UHG V9-122

Kursart: Vorlesung ohne Übung
Kurszeit: Freitag, 10:15 - 11:45 Uhr
Kursraum: Zoom (digital)
Erster Termin: 06.11.2020

Prognosen sind eine schwierige Sache. Vor allem, wenn sie die Zukunft betreffen.

(vmtl. Mark Twain)

Motivation

Jeder von uns fällt tagtäglich Entscheidungen auf Grund einer meist un(ter)bewusst ablaufenden Risikoprädiktion. „Ist es für mich sicher oder soll ich lieber den Bus noch vorbeifahren lassen, bevor ich die Straße überquere?“ Die wenigsten sind sich dabei bewusst, dass sie hierfür automatisch verschiedene Rahmenbedingungen (Faktoren) berücksichtigen. Dazu gehören etwa die Masse des Busses, die Geschwindigkeit und Entfernung des Fahrzeuges, die Breite der Straße, die eigene Geschwindigkeit und Erfahrungswerte („Zusammenstöße mit Bussen tun mir deutlich mehr weh als dem Bus.“). Ebenso automatisch sortieren wir bereits diese Faktoren nach ihrer Bedeutung (Einflußgröße) und entscheiden z.B. eher nach der Entfernung des Fahrzeuges, als nach der vorherrschenden relativen Luftfeuchtigkeit.

Das was unser Gehirn in dieser und in hunderten ähnlichen Situationen tagtäglich ganz von allein macht, kann in der Statistik in ein geordnetes Geflecht von Analyseschritten gebracht werden. Die Rahmenbedingungen und Forschungsfrage lassen sich über ein statistisches Modell beschreiben, die Faktoren können identifiziert, nach ihrer Relevanz ausgewählt oder sortiert werden, um dann statistisch solide auch in deutlich komplexeren Situationen eine begründete Entscheidung zu treffen. Dabei sind die Anwendungen solcher *risk prediction models* mannigfaltig und beziehen sich nicht nur auf das klassische Verständnis des Begriffes „Risiko“.

Ähnlich geht es in der Überlebenszeitanalyse nicht zwingend immer um Mortalität und Überleben. Produkt- oder Therapievergleiche wie in der Pharmaindustrie („Das neue Medikament XYZ wirkt besser als der Vorgänger ABC“) sind nur eine der Anwendungen, ebenso können auch der Einsatz von Werbung („Einkaufsverhalten“), Jagdstrategien im Tierreich und/oder Haltbarkeiten von Autoersatzteilen untersucht und verglichen werden. Die sogenannten Events, die über einen begrenzten Zeitverlauf beobachtet und gemessen werden, werden wiederum in eine akkurate mathematische Form gebracht und durch ein statistisches Modell beschrieben.

Kursbeschreibung

Die Veranstaltung „Überlebenszeitanalysen und Risikoprädiktion“ vermittelt den grundlegenden Umgang mit Zeit-Event-Datensätzen und erweitert das Repertoire von Analysemethoden und -modellen der Überlebenszeitanalyse (z.B. *Cox proportional hazard* und Kaplan-Meier). Zudem wird die Risikoprädiktion im Zentrum der Veranstaltung stehen und den Regressionsbegriff weiter ausbauen. Hierbei soll insbesondere der Umgang mit dem $n \ll k$ -Problem geübt werden. Weitere Komponenten wie die Regularisierung und (Kreuz-)Validierung der generierten Modelle, die nicht nur in der wissenschaftlichen Praxis auf einer täglichen Basis verwendet werden, werden in der Veranstaltung genauer erläutert und erprobt.

Anhand von Beispielen aus Wissenschaft und Wirtschaft sollen gängige Stolperfallen und ihre Lösungsmöglichkeiten visualisiert und begreiflich gemacht werden.

Die Theorie der so erlernten Methoden wird mittels eines Exkurses in die Programmiersprache R unterfüttert und über einen Ausflug in die Bayesianische Statistik kontextualisiert.

Ziele der Lehrveranstaltung

Studierende sollten nach Besuchen der Lehrveranstaltung folgende Kenntnisse und Fähigkeiten erlangt haben:

- sicherer Umgang mit Zeit-Event-Datensätzen,
- grundlegende Kenntnisse der Risikoprädiktion,
- weiterführende Programmierkenntnisse in R.

Darüberhinaus sollten die Kenntnisse und Fähigkeiten in folgenden *soft skills* vertieft worden sein:

- kritischer Umgang mit Analyseergebnissen,
- Transfer von bekannten Inhalten auf neue Kontexte,
- Verständnis für wissenschaftliche Praxis.

Voraussetzungen und Erwartungen

Teilnehmer der Veranstaltung sollten eingeschriebene Master-Studierende der Universität Bielefeld sein. Ein Bachelor-Abschluss in einer mathematischen/statistischen Disziplin ist förderlich. Sicherer Umgang mit Datensätzen/-formaten sowie Interesse an Modellierung von umweltwissenschaftlichen, biomedizinischen und/oder wirtschaftswissenschaftlichen Zusammenhängen ist von Vorteil. Kenntnisse der Programmierung in einer der statistischen Programmiersprachen (wozu SPSS nicht gehört) wie zum Beispiel R sind hilfreich. Ein Grundverständnis von Regressionsmodellen wird vorausgesetzt.

Materialien und Literaturhinweise

Das Vorlesungsskript wird - im Normalfall - bereits vor Beginn der Veranstaltung im zugehörigen LernraumPlus/Moodle verfügbar sein. Die Dateien sind individuell passwortgeschützt, wobei das Passwort der jeweiligen Skriptdatei am Anfang der Veranstaltung herausgegeben wird.

Zu den Themen der Veranstaltung existiert eine Vielzahl von teils in ihrem Aufbau sehr unterschiedlicher und mehr oder minder ausführlicher Literatur. Die hier angegebene Literatur ist also nur als Vorschlag zu verstehen und ist durch persönliche Präferenzen eingeschränkt. Insbesondere habe ich versucht, Literatur zur Vorbereitung zu nutzen, die auch (teilweise sogar online) in der Universitätsbibliothek verfügbar ist. Sofern weiterführende Literatur genutzt wurde, ist dies im Skript vermerkt.

Literatur:

- Klein, John P. (2003). *Survival analysis. Statistics for biology and health (2. ed.)*. New York [u.a.]: Springer.
- Selvin, Steve. (2008). *Survival analysis for epidemiologic and medical research. Practical guides to biostatistics and epidemiology*. Cambridge [u.a.]: Cambridge Univ. Press.
- Moore, Dirk F. (2016). *Applied Survival Analysis Using R. Use R!* Cham: Springer International Publishing.

Leistungsnachweis/Prüfung

Die Veranstaltung ist den Modulen 31-MM15(-WiMa), 31-M-ASM1/2 und 31-SW-AKStat zugeordnet. Es gelten die für diese Module bekannten Prüfungsordnungen und Prüfungsleistungen. Insbesondere bedeutet dies, dass am Ende des Semesters eine mündliche (Teil-)Prüfung abgelegt werden kann, um sich die Teilnahme anrechnen zu lassen.

Eine Studienleistung existiert nicht.

Zu beachten ist, dass für die Anerkennung als Teilleistung in 31-SW-AKStat ein Formblatt ausgefüllt werden soll, das von mir gegengezeichnet werden muss, ehe es von Herrn Prof. Dr. Langrock akzeptiert werden kann. Dies sollte möglichst frühzeitig geschehen, wobei sich kurz vor oder direkt nach einer der Veranstaltung zum Unterzeichnen anbietet.

Zeitplan

Der Zeitplan ist vorläufig und kann sich ändern. Insbesondere beziehen sich die angegebenen Inhalte auf das (digitale) Präsenzstudium. Die folgenden Stichpunkte sind dabei als Schlüsselbegriffe zu verstehen, mit denen Studierende nach der Veranstaltung (+ Selbststudienzeit) sicher umgehen können sollten. Die Veranstaltungen bauen dabei tendenziell aufeinander auf, sodass bei Verpassen einer Veranstaltung anzuraten ist, sich die Inhalte selbstständig anzueignen. Grundsätzlich sind alle Inhalte der Veranstaltung prüfungsrelevant (vgl. Leistungsnachweis/Prüfung).

Veranstaltung 01, 06.11.2020 Erste Schritte...

- Organisatorisches
- Motivation

Veranstaltung 02, 13.11.2020 Regressionsmodelle

- Wiederholung: Regressionsmodelle
- Erweiterung auf Zeit-Event-Datenformate
- „n«k Problem“

Veranstaltung 03, 20.11.2020 Überlebenszeitanalysen

- Arten von Modellen
- Überlebensfunktion
- Ereignisdichtefunktion

Veranstaltung 04, 27.11.2020 Kaplan-Meier Modelle

Veranstaltung 05, 04.12.2020 *Cox proportional hazard models*

Veranstaltung 06, 11.12.2020 Risikoprädiktion

- Beispiel: binär logistische Regression
- „n«k Problem“
- *subsetting*
- Kolinearität

Veranstaltung 07, 18.12.2020 Regularisierung

- LASSO
- Ridge

Veranstaltung 08, 25.12.2020 entfällt (Fröhliche Weihnachten!)

Veranstaltung 09, 01.01.2021 entfällt (Guten Rutsch!)

Veranstaltung 10, 08.01.2021 Validierung

- Kreuzvalidierung
- *leave-one-out*
- Informationskriterien / Güte

Veranstaltung 11, 15.01.2021 Erweiterungen, Interpretation und Stolperfallen

- *hazard ratio*
- *sensitivity and specificity*
- ROC-Kurven
- *overfitting*

Veranstaltung 12, 22.01.2021 Anwendung und Beispiele

- wissenschaftliche Praxis

Veranstaltung 13, 29.01.2021 Exkurs R & Bayesianischer Ansatz

- programmierende Praxis
- „Modelle ohne p-Wert“

Veranstaltung 14, 05.02.2021 Wunschthema

Veranstaltung 15, 12.02.2021 Zusammenfassung

(Wichtig: Bitte habt ein (weiteres) internetfähiges Gadget zur Hand!)

- Wiederholung
- (offene) Fragerunde
- Prüfungsvorbereitung

Richtlinien

Während der Veranstaltung

Die Unterrichtsmaterialien sollten im Normalfall bereits vor Beginn der Veranstaltung verfügbar sein. Da es sich hauptsächlich um *pdf-files* handeln wird und die elektronische Aufzeichnung von Notizen ggf. bevorzugt sein könnte, wird die Nutzung von Computern, sprich Laptops oder Tablets, zugelassen. Während des Präsenzstudiums sollten die Computer nicht für Aktivitäten genutzt werden, die nichts mit den Studieninhalten zu tun haben (z.B. Steam oder E-mail-Korrespondenz). Smartphones und Handys bringen keinen Mehrwert im Präsenzstudium und sollten daher mindestens auf stumm geschaltet sein (dies gilt auch für den Vibrationsalarm). Essen und Trinken sind im Präsenzstudium erlaubt, solange die anderen Teilnehmer nicht gestört werden.

Durchgängige Teilnahme

Die Studieninhalte der Veranstaltungen bauen aufeinander auf, daher ist eine Teilnahme an allen Veranstaltungen anzuraten und wird darüber hinaus auch erwartet. Eine rechtliche Anwesenheitspflicht gilt jedoch nicht. Entschuldigungen/Abmeldungen bei Abwesenheit sind nicht notwendig. Bei Zuspätkommen sollte darauf geachtet werden, andere Teilnehmer nicht zu stören und möglichst auch nicht das Fortsetzen der Veranstaltung unnötig zu verzögern.

Abgabefristen

Sofern im Kurs ggf. eine Abgabefrist für Studieninhalte vereinbart wurde, ist diese einzuhalten. Eine Verlängerung einer Abgabefrist ist nur nach frühzeitiger Absprache mit mir möglich. Bei Verlängerung ist auch diese Frist bindend. Verspätete Abgaben werden weder akzeptiert noch berücksichtigt.

Akademische Integrität und Ehrlichkeit

Teilnehmer der Veranstaltung sollten als Studierende der Universität Bielefeld den Richtlinien zur guten wissenschaftlichen Praxis folgen¹.

Allgemeine Beeinträchtigungen

Sofern nachweislich Beeinträchtigungen auf Seiten eines Studierenden vorliegen, die die Teilnahme an der Lehrveranstaltung erschweren, sollte dies frühzeitig angezeigt werden, damit mögliche Alternativen bzw. Anpassungen besprochen und vorgenommen werden können.

Heterogenität in der Lehre

Wissenschaft lebt von Vielfalt! Diskriminierung und Belästigung („Mobbing“) aufgrund von Äußerlichkeiten, Beeinträchtigungen, Geschlecht, Alter, sexueller Orientierung, Selbstverständnis, Religion, kulturelle Herkunft, Geburtsort und anderen Faktoren des persönlichen Lebens sind eine Verletzung des Grundrechts jedes Individuums und werden in der Lehrveranstaltung und dem Umfeld nicht toleriert.

Datenschutz

Personenbezogene Daten werden von mir nur im an einer Universität üblichen Umfang verwendet (Teilnehmerlisten, E-mail-Korrespondenz, Leistungsüberprüfung). Die Daten werden nicht an universitätsferne Dritte weitergegeben oder veröffentlicht, nicht für Forschung genutzt oder nach Ende der Lehrveranstaltung (bzw. Leistungserbringung) auf universitätsfernen Medien gespeichert.

¹http://www.uni-bielefeld.de/gute_wiss_praxis/