

Projekt 2

Planung und Analyse von intervallzensierten Daten

In diesem Projekt sollen Methoden für intervallzensierte Daten per Simulation und anhand eines kleinen Datensatzes miteinander verglichen werden. Bei dem Datensatz handelt es sich um Diamantausbrüche beim Bohren in Beton mit einem diamantimprägnierten Werkzeug. Abbildung 1 zeigt ein Bohrsegment zu zwei aufeinander folgenden Zeitpunkten. Die Diamanten, die ausbrechen, sind rot gekennzeichnet. Gleichzeitig treten aber auch zwei neue Diamanten auf, die blau gekennzeichnet sind.



Abbildung 1: Darstellung eines Bohrsegments zu zwei aufeinander folgenden Zeitpunkten, zwischen denen zwei Diamanten ausgebrochen und zwei Diamanten neu hinzugekommen sind

Da der Diamantausbruch ein relativ seltenes Ereignis ist, wird zwischen zwei Zeitpunkten oft überhaupt kein Diamantausbruch beobachtet. Gibt es zu viele Intervalle ohne Diamantausbruch, so wird die statistische Analyse schwieriger. Hinzu kommt, dass das Unterbrechen des Bohrprozesses und die Analyse der herausgebrochenen und hinzugekommenen Diamanten sehr aufwendig ist. Das spricht dafür wenige große Abstände für die Analysezeitpunkte zu wählen. Wählt man aber die Abstände der Zeitpunkte zu groß, dann werden zu viele Diamantausbrüche beobachtet, womit wieder die statistische Analyse ungenau wird. Damit ist es sinnvoll, die Analysezeitpunkte in Abhängigkeit zu dem, was schon beobachtet wurde, zu wählen. Das wird mit sogenannten adaptierten Versuchsplänen erreicht. Per Simulation sollen daher verschiedene adaptierte Versuchspläne für die Wahl der Analysezeitpunkte miteinander verglichen werden. Außerdem soll anhand der vorhandenen Daten Vorschläge erarbeitet werden, in welchen Abständen in Zukunft die Bohrprozesse unterbrochen werden sollen.

Versuchsbeschreibung

Die vorliegenden Daten stammen aus vier Versuchen im Zeitraum September bis Oktober 2016, bei denen mit einem diamantimprägnierten Bohrsegment in Beton gebohrt wurde. Dabei wurde jeweils nur ein Segment einer Bohrkronen verwendet. Diese Segmente besitzen eine Metallmatrix, in der sich zufällig verteilte Diamanten befinden, von denen nur ein Teil davon sichtbar ist und somit zum Einsatz kommen. Jeweils zwei Versuche wurden mit normalen Beton C20 und zwei Versuche mit dem besonders druckfesten, aber mehr homogenen Beton C100 durchgeführt. Je einer dieser zwei Versuche erfolgte mit Diamanten der Siebgröße 40/50 *us-mesh*, was einer Diamantgröße von 297-400 μm entspricht, und je einer der zwei Versuche mit Diamanten der Siebgröße 20/30 *us-mesh*, was einer Diamantgröße von 595-841 μm entspricht. Diese Bohrversuche waren sogenannte wegeregeltete Versuche, bei denen vorgegeben wurde, dass in einer Minute 4 mm tief gebohrt werden soll. Bei 616 Rotationen pro Minute bedeutet das, dass pro Umdrehung 6.49 μm tief gebohrt wurde, so dass die Vorschubgeschwindigkeit 6.49 $\mu\text{m}/U$ ist. Nach jeder Minute, d.h. nach je weiteren 4 mm, wurde das Bohren unterbrochen und im Bohrsegment wurden die sichtbaren Diamanten ermittelt. Auch Fotos wie die in Abbildung 1 wurden davon gemacht.

Datenbeschreibung

Die xls-Datei `MA_Shi_fuerStatistik_neu.xls` mit den Daten finden Sie auf der Homepage zu dieser Veranstaltung. Diese enthält vier Blätter mit den Namen `Drilling_C100_b29`, `Drilling_C20_b28`, `Drilling_C20_b19` und `Drilling_C100_b18`, wobei jedes Blatt die Daten von eins der vier Versuche enthält. Die Bezeichnungen C20 und C100 stehen für die beiden Betonsorten, b29 und b28 bezeichnen die kleineren Diamanten hergestellt mit der Siebgröße 40/50 sowie b19 und b18 die größeren Diamanten, die mit der Siebgröße 20/30 produziert wurden. Jedes Blatt enthält einen Kopf, welche die Betonart und Siebgröße enthält. Der Datensatz besteht aus 21 Spalten, wobei viele Spalten die Bedingungen des Experimentes beschreiben, die bei allen vier Versuchen gleich waren. Von besonderer Bedeutung sind hier die Spalten R, V und W, die die Anzahlen der herausgebrochenen Diamanten, die Anzahl der neu aufgetauchten Diamanten beziehungsweise die Gesamtanzahl aller sichtbaren Diamanten liefern.

Die xls-Dateien `UpdateB19.xlsx` und `UpdateB28.xlsx` enthalten Updates für die Versuche B19 und B28, die ein zweites Mal anhand der Bilder, die in den Dateien `FotoalbumB19.pptx` und `FotoalbumB28.pptx` enthalten sind, vorgenommen wurden. Dabei ergaben sich insbesondere beim Versuch B28 stark abweichende Ergebnisse. Die Dateien `B19_Matrix.xlsx` und `B28_Matrix.xlsx` liefern zusätzlich das Auftreten und Verschwinden der Diamanten getrennt nach den einzelnen Diamanten.

Aufgaben

- 1) Nehmen Sie an, dass die Zeit, bis ein Diamant herausbricht, durch die gedächtnislose Exponentialverteilung gegeben ist. Schätzen Sie bei jedem der vier Versuche den unbekannt Parameter der Exponentialverteilung mit mindestens zwei verschiedenen Methoden. Wie unterscheiden sich die Schätzungen beim Versuch B28 bei den zwei verschiedenen Auszählungen?
- 2) Nehmen Sie an, dass als Daten nur die Dateien `Drilling_C100_b29`, `Drilling_C20_b28`, `Drilling_C20_b19`, `Drilling_C100_b18` sowie `UpdateB19.xlsx` und `UpdateB28.xlsx` vorhanden sind. D.h. die Information, wie lange die einzelnen Diamanten sichtbar sind, ist nicht gegeben. Finden Sie auch für diese Situation unter geeigneten Zusatzannahmen eine Schätzmethode für den Parameter der Exponentialverteilung. Wie unterscheiden sich diese Schätzungen bei den verschiedenen Versuchen und Auszählungen?
- 3) Geben Sie für die Situationen in 1) und 2) Schätzungen der Wahrscheinlichkeit für einen Diamantausbruch innerhalb von 30 Minuten an.
- 4) Geben Sie für die Situationen in 1) und 2) Schätzungen des Zeitpunktes an, bei dem die Wahrscheinlichkeit für einen Diamantausbruch bis zu diesem Zeitpunkt bei 50% liegt.
- 5) Vergleichen Sie die Schätzungen per Simulation.
- 6) Geben Sie mindestens zwei Methoden an, um optimale Inspektionszeitpunkte adaptiv festzulegen, wenn die Anzahl der Inspektionen vorgegeben ist. Wie sollten die Inspektionszeitpunkte bei einer Inspektion, bei zwei, 10 und 20 Inspektionen gewählt werden, wenn die vier Versuche unter den gleichen Bedingungen noch einmal durchgeführt werden?
- 7) Geben Sie mindestens zwei Methoden an, um optimale Inspektionszeitpunkte adaptiv festzulegen, wenn die maximale Inspektionszeit vorgegeben ist. Wie sollten die Inspektionszeitpunkte bei maximaler Inspektionszeit von 20 und 30 Minuten festgelegt werden, wenn die vier Versuche unter den gleichen Bedingungen noch einmal durchgeführt werden und es maximal 30 Inspektionen geben darf?
- 8) Vergleichen Sie die zwei Methoden zum Festlegen der Inspektionszeitpunkte per Simulation. Dabei reicht es, wenn Sie nur eine der Fragestellungen aus entweder 6) oder 7) betrachten.

Literatur

- 1) Boruvka, A. und Cook, R.J. (2015). A Cox-Aalen model for interval-censored data. *Scandinavian Journal of Statistics* **42**, 414-426.

- 2) Cook, R.J. und Lawless, J.F. (2007). *The Statistical Analysis of Recurrent Events*. Springer, New York.
- 3) Klein, J.P. und Moeschberger, M.L. (2003). *Survival Analysis. Techniques for Censored and Truncated Data*. Springer, New York.
- 4) Konstantinou, M. (2013). *Locally Optimal and Robust Designs for Two-parameter Nonlinear Models with Application to Survival Models*. Ph.D. thesis, University of Southampton.
- 5) Müller, Ch.H. (2017). Estimators and designs for interval censored data. *Manuskript*.
- 6) Müller, Ch.H. (2013). D-optimal designs for lifetime experiments with exponential distribution and censoring. In: D. Ucinski, A. C. Atkinson, M. Patan (Hrsg.) *mODa 10 - Advances in Model-Oriented Design and Analysis*, Physica-Verlag, Heidelberg, 179-186.
- 7) Rivas-López, M.J., López-Fidalgo, J. und del Campo, R. (2014). Optimal experimental designs for accelerated failure time with type I and random censoring. *Biometrical Journal* **56**, 819-837.
- 8) Tillmann, W., Biermann, D., Weihs, C., Ferreira, M., Nellesen, J., Kansteiner, M. und Herbrandt S. (2015). Influence of the diamond size and diamond distribution on the wear behavior of sintered diamond-metal composites during the machining of reinforced concrete. In: European Powder Metallurgy Association (Hrsg.) *Euro PM2015 Proceedings: HM - Modelling and Characterisation*, EuroPM.
- 9) Tutz, G. und Schmid, M. (2016). *Modeling Discrete Time-to-Event Data*. Springer. New York.

Abgabetermin

Abgabe am **Montag, dem 7.8.2017**, bei Frau Große-Oetringhaus (Sekretariat Müller).