

Übung 6. – Anforderungsanalyse, explorative Datenanalyse

1 Datenanalyse von Leistung eines Servers

Wir haben an einem Server die folgenden Leistungsparameter gemessen:

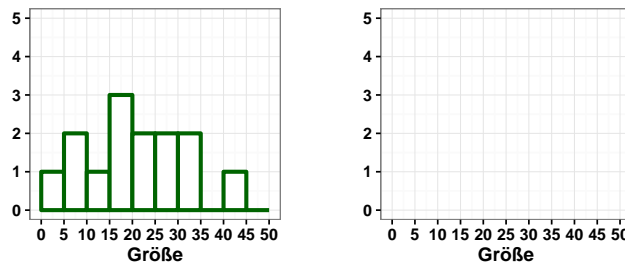
Zeitpunkt der Probe [ms]	500	600	700	800	900
Anzahl der in den letzten 100ms bedienten Anfragen	11	12	21	18	20
Durchschnittsbedienzeit der letzten 100ms [ms]	15	20	21	25	27
Prozessorauslastung der letzten 100ms [%]	12	13	16	17	19
Festplatten-I/O-Auslastung der letzten 100ms [%]	55	63	87	61	73

- Stellen Sie die Anzahl der bedienten Anfragen und die Prozessorauslastung mit Hilfe eines Streudiagrammes (scatterplot) dar. Interpretieren Sie das Diagramm.
- Wie hoch ist der Durchsatz zu der Zeit der ersten Probe? Anhand der Ergebnisse der 5 Abtastungen, wie hoch ist der gemessene Durchschnitt und Median des Durchsatzes? Welche Messwerte fallen in das 40%-Quantil?
- Zwischen welche gemessenen Parameter kann man Grund-Kausal Beziehung ahnen?

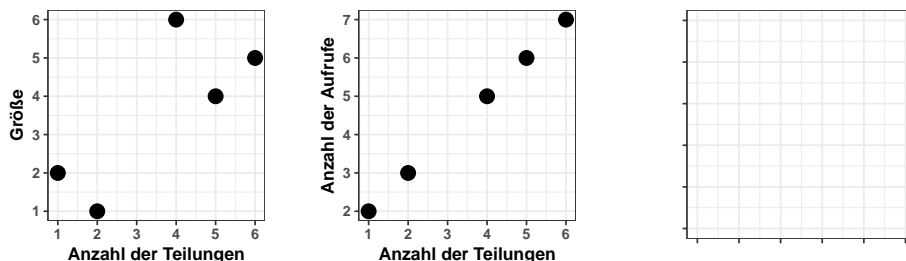
2 Bildgalerie – Datenanalyse

In unserer Bildgalerie können die Benutzer aufgrund der Suche die für einen Suchausdruck passende Bilder anschauen.

- Wir haben die Verteilung der Größen der Alben im nachfolgenden Histogramm dargestellt. Da es bei der wirksamen Organisation der Datenbank genug ist nur das zu wissen, wie viele Alben wir mit weniger als 10, zwischen 10 und 20, usw. Bildern haben, möchten wir ein Histogramm haben, das zweimal so breite Spalten hat, als das vorige (und berechnet die Spalten ebenfalls von der Größe 0). Zeichnen wir die Abbildung!



- Wir haben die Größen der 5 ausgewählte Alben und die Anzahl der Aufrufe mit der Anzahl der Teilungen verglichen an einem Punkt-Punkt Diagramm (Scatterplot) dargestellt. Ist es wahr, dass je größer ein Album ist, desto mehr Leute es besichtigen? Beantworten sie die Frage an einem dritten Punkt-Punkt Diagramm, das die Anzahl der Aufrufe gemäß der Größe darstellt!



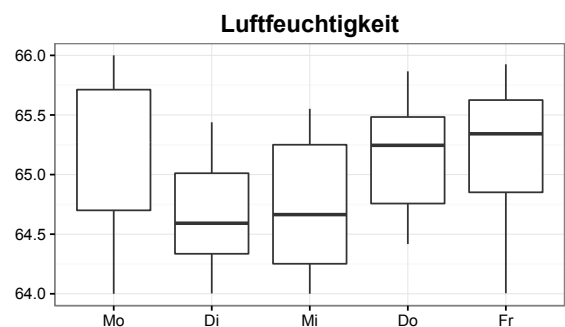
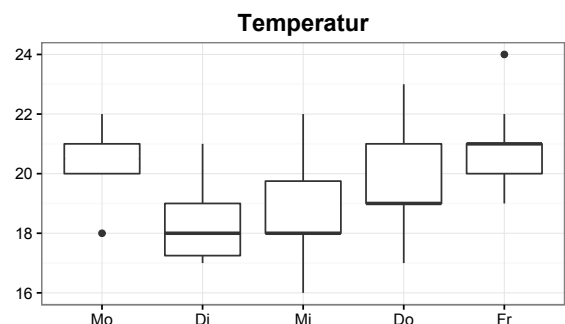
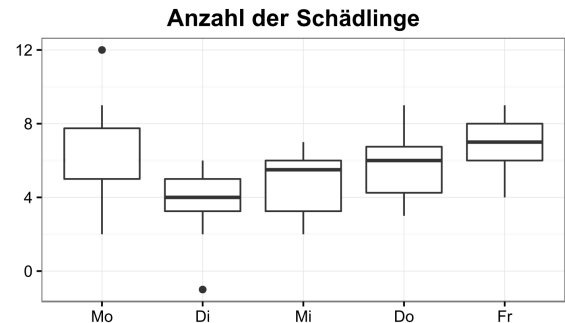
- Wir wollen die typische Popularität der Alben bestimmen, deshalb haben wir aufgrund der Punkt-Punkt Diagrammen den Durchschnitt und den Median der Anzahlen der Aufrufe berechnet. Kann man das im Allgemeinen aufgrund eines Punkt-Punkt Diagrammes tun? Wie ändern sich diese Mittelwerte, falls wir ein neues Album aufladen, dessen Anzahl der Aufrufe 40 ist?

3 Netzwerk von Sensoren (frühere Klausuraufgabe) – Datenanalyse

Sei ein Netzwerk von Landwirtschaftssensoren gegeben, mit dem Freilandstücke, Glashäuser und Folien überwacht werden. Die Sensoren messen Daten wie Temperatur, Luftfeuchtigkeit, Helligkeit, Windstärke, Anzahl der Schädlinge, usw.

Datum	Temp. [°C]	Luftf. [%]	Schädl. [Stück]
4. Mai 2015 08:00	18	66,00	3
4. Mai 2015 09:00	20	65,75	6
4. Mai 2015 10:00	20	65,75	8
4. Mai 2015 11:00	20	65,50	9
4. Mai 2015 12:00	20	65,50	5
4. Mai 2015 13:00	21	65,00	12
4. Mai 2015 14:00	21	64,70	5
4. Mai 2015 15:00	21	64,70	6
4. Mai 2015 16:00	21	64,60	7
4. Mai 2015 17:00	22	64,00	2

- Leider sind die Mittelwerte (Median) von Montag, dem 4. Mai von den Bildern verschwunden, zeichnen Sie sie anhand der Daten in der Tabelle ein.
- Interpretieren Sie die Diagramme: Von welchen Variablen weisen die 1. Quartile eine streng monotone Änderung mit der Zeit auf?
- (Ergänzende Aufgabe.) Die Temperaturwerte und die Anzahldaten der Schädlinge von Montag, dem 4. Mai sollen an einem Diagramm mit parallelen Koordinaten verglichen werden.



4 Netzwerk von Sensoren (frühere Klausuraufgabe)

– Leistungsanalyse

(Zu Aufgabe 3. gehörende Aufgaben für Analyse von Leistungen.)

Die verschiedenen Sensoren liefern entsprechende Daten von ihrer 100 m-Umgebung, die sie samt Zeitstempel über dem drahtlosen Netzwerk zu der Zentrale schicken. Der Prozessor des zentralen Rechners verarbeitet diese Nachrichten, dann schreibt er sie auf eine Platte zur Langzeit-speicherung. Unser Betrieb hat insgesamt 4500 Sensoren installiert, die je eine Nachricht pro Minute schicken. Unsere Infrastruktur bedient die aktuelle Belastung erfolgreich. Das drahtlose Netzwerk kann 100 Messdatenpakete pro Sekunde übertragen. Der Prozessor des zentralen Rechners läuft zu 75% leer. Die Platte braucht 8 ms für das Speichern einer Nachricht.

- Wie viele Nachrichten pro Sekunde ist der aktuelle Durchsatz der Infrastruktur?
- Wie hoch sind der aktuelle Durchsatz, der Grenzdurchsatz und die Auslastung des Netzwerks, des Prozessors, bzw. der Platte?
- Um die Messgenauigkeit zu erhöhen, wollen wir weitere Sensoren installieren. Wie viele dürfen wir ohne Verstärkung der Infrastruktur? Nehmen Sie eine lineare Skalierbarkeit an.
- Mit geschickter Kodierung unterstützt das Netzwerk, dass mehrere Sensoren gleichzeitig Messdaten schicken können. Im Durchschnitt wie viele Sensoren senden gleichzeitig im Netzwerk (wie hoch ist die Überlappung im Netzwerk) aktuell bzw. bei der maximalen Auslastung des Netzwerks, wenn das Abschicken einer Nachricht 40 ms dauert?

5 Überprüfung der Anforderungen bei einem Eisenbahnenversicherungssystem

Wir entwickeln ein Versicherungssystem für Eisenbahnen. Das wichtigste Ziel des Systems ist, das Zusammenstoßen der Züge zu verhindern. Um ein entsprechendes System zu entwerfen brauchen wir eine Anforderungsspezifikation guter Qualität, weil wir anhand der Anforderungen später die Testfälle und die weitere Überprüfungen ausführen sollen.

Tabelle 1: Die Anforderungen des Eisenbahnenversicherungssystems (Teilausschnitt)

R1	Sicherheit	Die Züge auf dem beaufsichtigten Bahnsystem dürfen nicht zusammenstoßen.
R2	Funktionalität	Die Züge müssen ihre Ziele erreichen können.
R3	Optimalität	Die Reisezeiten sollen minimalisiert werden.
R4	Aufsicht der Bahnabschnitte	Die Bahnstrecken sollen auf Abschnitte aufgeteilt werden, auf deren gleichzeitig nur ein Zug sich aufhalten darf.
R5	Aufteilung auf Abschnitte	Das Bahnsystem soll auf Abschnitte aufgeteilt werden.
R6	Besetzbarkeit	Auf einem Abschnitt darf gleichzeitig nur ein Zug sich aufhalten.
R7	Feststellung des Besetztseins	Irgendwie muss festgestellt werden, ob es auf einem Abschnitt einen Zug gibt, oder nicht.
R8	Fehlertoleranz	Auf dem Ausfall der Komponenten soll man sich vorbereiten.
R9	Sensoren zu Besetzzeitn	Das Besetztsein soll mit verschiedenen Sensoren redundant empfunden werden.
R10	Schienenensoren	In die Schiene sollen auf allen Strecken Sensoren eingebaut werden, die melden können, ob auf dem gegebenen Abschnitt ein Zug sich aufhält, oder nicht.
R11	Kamerasystem	Wo es möglich ist, sollen Kameras angesiedelt werden, um die Abschnitte zu beobachten.
R12	Positionsbestimmung	Die Züge sollen ihre Position kontinuierlich zu der Leitzentrale schicken.
R13	GPS Teilsystem	Die Züge sollen mit GPS Teilsystem ausgerüstet werden.
R14	Drahtloser Anschluss	Es muss versichert werden, dass die Züge ihre Position über drahtlosem Netzwerk zur Zentrale schicken könnten.
R15	Steuerung der Züge	Es muss verhindert werden, dass auf einen besetzten Abschnitt noch ein Zug einfahren kann.
R16	Züge anhalten	Die Zentrale muss den Zug gleich zum Stillstand bringen können.
R17	Unterstützung von Lokomotivtypen	Das System soll alle Lokomotivtypen unterstützen, die auf diesen Strecken fahren können.
R18	Lokomotive nicht verändern	Keine Lösung kann verwendet werden, zu welcher die Steuerung der Loks verändert werden muss.

- Sammeln wir die Teilnehmer, die bei Entwurf eines solchen Systems wichtig sind, also die, welche Anforderungen gegenüber des zukünftigen Systems stellen können (sog. *stakeholder*).
- Nach der Erkundung der „stakeholder“ haben wir die entsprechende Anforderungen auch gesammelt; ein Teil von denen gibt die Tabelle 1 an. Zeichnen wir die Zusammenhänge zwischen den Anforderungen mit Hilfe eines Graphen! Es gibt eine gerichtete Kante von A nach B , falls (1) die Anforderung A Teil der Anforderung B ist (*Komposition*), (2) die Anforderung A eine Verfeinerung von Anforderung B ist (*refines* Beziehung), bzw. (3) die Anforderung A von Anforderung B stammen kann (*derive* Beziehung). Beschäftigen wir uns nicht damit, welche Beziehung von den obigen gilt; nur die Existenz der Beziehung ist wichtig.

- c. Welche von den aufgezählten Anforderungen sind funktionale Anforderungen, bzw. welcher Typ die nichtfunktionalen Anforderungen haben (Sicherheit, Leistung, Zuverlässigkeit, usw.)?
- d. Prüfen wir, ob das gegebene Anforderungssystem konsistent ist! Falls nein, geben wir ein Gegenbeispiel!
- e. Geben wir ein Beispiel für direkt überprüfbare Anforderungen!

Optionale Aufgaben

6 Gemeinschaftsseite

Wir betreiben eine Gemeinschaftsseite am Internet. In der letzten Zeit ist die Seite viel populärer geworden, deswegen ist aber die Antwortzeit auch gewachsen. Unser Geschäftsziel ist es, dass die Homepage in der Spitzenzeit auf einmal 1500 Benutzer mit 4 Sekunden durchschnittlicher Antwortzeit bedienen kann.

- a. Mindestens wie groß muss man den Durchsatz der Bedienungsinfrastruktur planen, falls die Zeit aller anderen Verzögerungen (Netzverkehr, HTML-Erscheinung beim Klienten) als eine Sekunde betrachtet werden kann?
- b. Entsprechend den Messungen an der wiederentworfenen Webseite braucht die Bedienung einer Anfrage durchschnittlich 20 ms CPU-Zeit am Webserver, und es besetzt einen Datenbankserver für 12,5 ms. Zurzeit bedienen 15 Webserver die Anfragen, und die Datenbank ist für 5 Bediener repliziert. Welche Art und mindestens wie viele Computer müssen wir noch kaufen, um das obige Ziel zu erfüllen, falls wir eine lineare Skalierbarkeit annehmen?
- c. (*) Wie groß wird die Auslastung im erweiterten System? Wie fern müsste das System erweitert werden, falls wir auch in der Spitzenzeit nur höchstens 50% Auslastung erlauben wollen?
- d. Betrachten wir nur die 2 Webserver und die 3 Datenbanken. Konstruieren wir zustandsbasierte Modelle, die die Ressourcen der Infrastruktur nach Verfügbarkeit (frei/besetzt) modellieren! Welche Designerentscheidungen treffen wir? Was sind die Vor- und Nachteile der verschiedenen Möglichkeiten?