

Simulation – Datenauswertung

Simulation

1. Einführendes Beispiel: Lottozahlen
2. Verallgemeinerung: Stochastischer Prozess
3. Typische Prozesse

Datenauswertung

4. Einführendes Beispiel: Lottozahlen
5. Verallgemeinerung: Stochastischer Prozess
6. Typische Prozesse

Simulation

1. Einführendes Beispiel

Lottozahlen

Je Spielrunde werden im schweizerischen Zahlenlotto zufällig 6 aus 40 (ursprüngliche Version) möglichen Zahlen gezogen. Derjenige, welcher alle Zahlen richtig voraus-sagt, ist der Gewinner.



Technische Realisierung:

40 ununterscheidbare Kugeln werden von 1 bis 40 nummeriert und in eine Urne gelegt. Nach Mischen wird zufällig eine Kugel herausgezogen und auf die Seite gelegt. Dieser Vorgang wird fünfmal wiederholt.

2. Verallgemeinerung

Lotto-Maschine **Verallgemeinerung** → Stochastischer Prozess

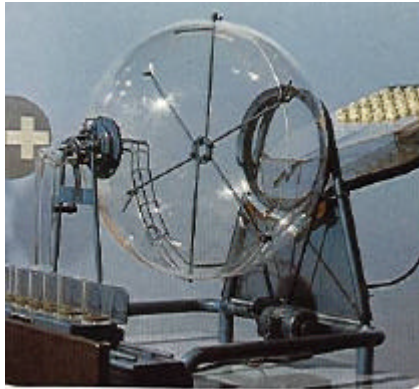
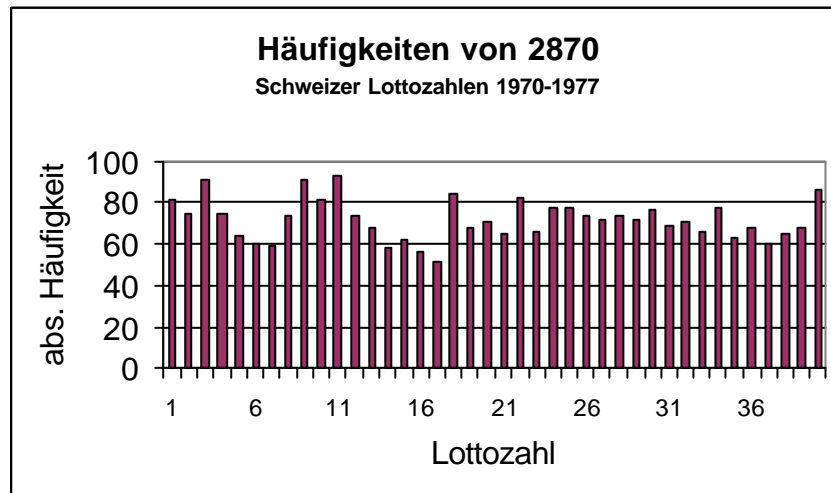
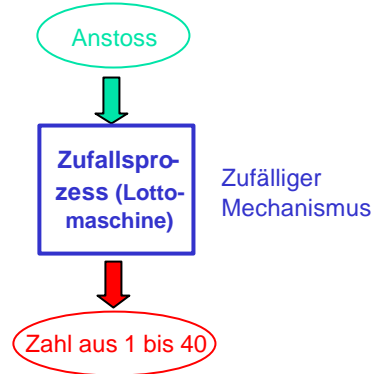


Bild: H. Riedwyl, Schweizer Zahlenlotto, Bern 1979.



Quelle: H. Riedwyl, Schweizer Zahlenlotto, Bern 1979, S. 21.

Einfachste Variante: Mit **Würfel**,
z.B. Summe von 9 Würfeln vermindert um 8
(Zahlen >42 weglassen)

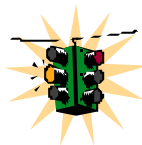
Einfache el. Variante: Mit **Excel-Funktion**:
-> Einfügen
-> Funktion
-> Math. & Trig.
-> Zufallsbereich

Allgemeine Variante: Mit **Random-Generator**:
In jeder Programmiersprache können z.B. durch
folgende Anweisungen Zufallszahlen zwischen 0
und 1 generiert werden:

$$x_0 = 1, 2, 3, \dots, \text{ oder } 2^{31}-1$$

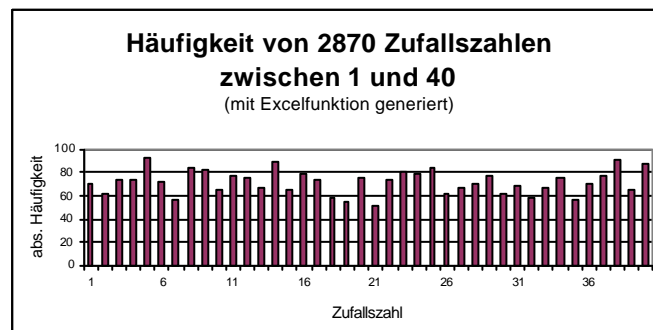
$$x_n = 65539 * x_{n-1} \text{ mod } 2^{31}$$

$$z_n = x_n / 2^{31}$$



Bisherige Zufallsprozesse liefern sogenannte
gleichverteilte Zufallszahlen.

Jede der möglichen Zahlen kommt bei einer grossen
Zahl von Wiederholungen des Zufallsprozesses gleich
oft vor!



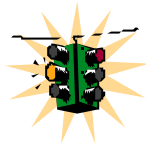
3. Typische Prozesse

Zweipunktprozess – Bernoulli-Prozess

Damit wird eine **Zufallsvariable** X modelliert, die nur zwei Werte annehmen kann:

$$X = \begin{cases} 0 & \text{bzw. Misserfolg} \\ 1 & \text{bzw. Erfolg} \end{cases}$$

Mit Münze praktisch realisierbar.



Die Häufigkeiten müssen **nicht gleich verteilt** sein!
z.B. Wurf eines halben Tennisballs

Binomialprozess

Der Binomialprozess wird aus mehreren (nicht notwendig gleichverteilten) Zweipunktprozessen aufgebaut. Die **Summe von n Zweipunktprozessen** ergibt einen **Binomialprozess**. Praktisch durch Wiederholung von Münzenwürfen realisierbar.

Poissonprozess

Typische Situation für einen Schalter (Post, Bank) oder allgemeiner einem sogenannten „Ankunftsprozess“:



Die Zufallsvariable X beschreibt die Anzahl der in einem gegebenen Zeitintervall der Länge T registrierten Ankünfte. Die Modellierung erfolgt durch den **Poissonprozess**.

Simulation von stochastischen Prozessen

Reale, zufällige Vorgänge lassen sich durch stochastische Prozesse modellieren. Excel stellt entsprechende Funktionen bereit.

-> Extras -> Analysefunktionen -> Zufallszahlengenerierung

Die einfachsten Prozesse sind:

- | | |
|--------------------|--|
| - Zweipunktprozess | Verteilung: Bernoulli
Parameter: p - rel. Häufigkeit für Erfolg |
| - Binomialprozess | Verteilung: Binomial
Parameter: p - rel. H'gkeit für Einzelerfolg |
| - Poissonprozess | Verteilung:
Parameter: λ – mittlere Anzahl Ankünfte |

Simulation – Datenauswertung

Simulation

1. Einführendes Beispiel: Lottozahlen
2. Verallgemeinerung: Stochastischer Prozess
3. Typische Prozesse

Datenauswertung

4. Einführendes Beispiel: Körpermasse
5. Masszahlen
6. Graphische Darstellungen

Die Menge aller statistischen Einheiten oder Merkmalsträger, die für eine zu untersuchende Fragestellung relevant sind, bildet eine **Grundgesamtheit**. Letztere wird gelegentlich auch als Population angesprochen.

Daten beziehen sich auf **Merkmalsträger** (beliebige Objekte). Für diese verwendet man in der Statistik auch die neutrale Bezeichnung Untersuchungseinheit oder statistische Einheit. Statistische Einheiten müssen sachlich, räumlich und zeitlich eindeutig identifizierbar sein.

Die Eigenschaften einer statistischen Einheit heißen **Merkmale**. Häufig werden an einer statistischen Einheit mehrere Merkmale gleichzeitig beobachtet. Die beobachteten Merkmalswerte heißen Merkmalsausprägungen oder Realisationen. Hinsichtlich der Menge der möglichen Merkmalswerte unterscheidet man zwischen stetigen Merkmalen und diskreten Merkmalen. Die Werte, die ein Merkmal einer statistischen Einheit annehmen kann, heißen Merkmalsausprägungen oder Realisationen.

Die Werte, die ein Merkmal einer statistischen Einheit annehmen kann, heißen **Merkmalsausprägungen** oder **Realisationen**.

Ein Merkmal, das nur endlich oder abzählbar unendlich viele Ausprägungen aufweist, heißt **diskret**.

Ein Merkmal, dessen Ausprägungen durch ein Intervall repräsentiert werden, heißt **stetig**. Für je zwei Merkmalsausprägungen eines stetigen Merkmals gilt, dass auch alle Zwischenwerte angenommen werden können.

Bei der Datenerhebung für ein Merkmal kann man mit einer **Nominalskala**, einer **Rangskala** (Ordinalskala) oder einer **metrischen Skala** (Kardinalskala) arbeiten.

Bei der **Nominalskala** stellen die Ausprägungen eines Merkmals nur Kategorien (Namen) dar.

Bei der **Rangskala** stellen die Ausprägungen eines Merkmals Kategorien dar, für die - anders als bei nominalskalierten Merkmalen - eine natürliche Rangordnung existiert.

Bei einer **metrischen Skala** sind für die Ausprägungen eines Merkmals neben einer natürlichen Rangordnung zumindest Abstände definiert (Differenzenbildung).

Für metrische Skalen können charakteristische Werte bestimmt werden.

- Durchschnittswert (Mittelwert)

Der mittlere oder durchschnittliche Wert einer „Zahlenmenge“, d.h. einer Stichprobe oder Grundgesamtheit, ist gleich dem arithmetischen Mittel aller Elemente.

Stichprobe: 1, 4, 5, 4, 6, 2, 6

Mittelwert: $\frac{1+4+5+4+6+2+6}{7} = 4$

Stichprobe: $x_1, x_2, x_3, \dots, x_n$

Mittelwert: $\frac{x_1 + x_2 + x_3 + \dots + x_n}{n} = \bar{x}$

- Zentralwert (Median)

Der zentrale Wert einer Stichprobe ist oft der geeignetere „Durchschnitt“ als der Mittelwert. Er liegt in der **SORTIERTEN** Stichprobe im Zentrum.

Stichprobe: 1, 4, 5, 4, 6, 2, 6

Zentralwert: 1, 2, 4, 4, 5, 6, 6

↑
mittlere Position

Stichprobe: $x_1, x_2, x_3, \dots, x_n$

Median: $x_{j_1}, x_{j_2}, \dots, x_{j_h}, \dots, x_{j_n}$

↑
mittlere Position

- Häufigster Wert (Modus, Modalwert)

Der häufigste Wert einer Stichprobe kann manchmal den durchschnittlichen Wert besser beschreiben als der Mittelwert.

Stichprobe: 1, 4, 5, 3, 6, 2, 6

Modus: 6

Stichprobe: $x_1, x_2, x_3, \dots, x_n$

Modus: x_{j_m}

- Spannweite (Streuungsmass)

Die Werte einer Stichprobe sind meist verschieden, sie „streuen“. Zwischen dem kleinsten und grössten Wert liegt die Spannweite.

Stichprobe: 1, 4, 5, 3, 6, 2, 6

Spannweite: grösster Wert – kleinster Wert = 6 – 1 = 5

Stichprobe: $x_1, x_2, x_3, \dots, x_n$

Spannweite: $\max(x_i) - \min(x_i)$

- Standardabweichung (Streuungsmaß)

Die Werte einer Stichprobe sind meist verschieden, sie „streu“.
 Die einen Werte sind kleiner, die anderen sind grösser als der Mittelwert.
 Der Durchschnitt dieser Abweichungen wäre ein Mass für die mittlere
 Streuung, sofern der absolute Betrag dieser Abweichungen berücksichtigt
 würde. Zur Vereinfachung der Rechnung wird jedoch meist die mittlere
 quadratische Abweichung bestimmt und daraus die Wurzel gezogen.

Stichprobe: 2, 4, 5, 3, 6, 2, 6

Standardabw.:
$$\sqrt{\frac{(2-4)^2 + (5-4)^2 + (3-4)^2 + (6-4)^2 + (2-4)^2 + (6-4)^2}{7}} = 1.6$$

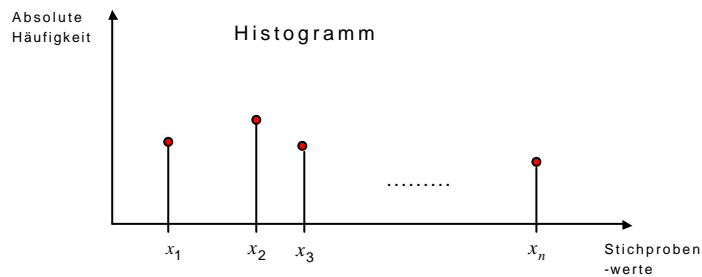
Stichprobe: $x_1, x_2, x_3, \dots, x_n$

Spannweite:
$$\sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n}} = s$$

6. Graphische Auswertung

- Häufigkeits-Diagramm (Histogramm) für diskrete Beobachtungswerte

Die Werte einer Stichprobe sind meist verschieden, sie können mehrmals
 in der Stichprobe auftreten. Jedem möglichen Wert einer Stichprobe ist
 eine absolute Häufigkeit zugeordnet: diskrete Funktion!



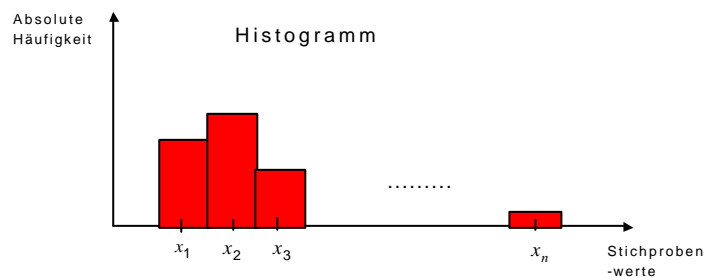
Zugehörige EXCEL-Befehle

- > Diagramm-Assistent
- > Diagrammtyp: Kurs
- > Datenbereich
 - (3 nebeneinanderliegende Spalten, wovon 1. Spalte mit abs. Häufigkeiten, 2. Spalte mit Nullen und 3. Spalte wie 1. Spalte)
 - > Reihe, Rubrikenachse (Spalte mit Merkmalswerten)
 - > Diagramm fertig stellen

- > Rechtecke 2 mal anklicken und Markierung anpassen

- Häufigkeits-Diagramm (Histogramm) für **stetige** Beobachtungswerte

Die Werte einer Stichprobe müssen in Klassen eingeteilt werden. Jeder Klasse ist eine absolute Häufigkeit zugeordnet. Stellvertretend wird die Klassenmitte als Beobachtungswert verwendet.

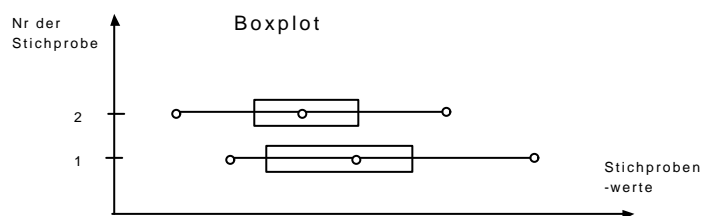


Zugehörige EXCEL-Befehle

- > Extras
 - > Analyse-Funktionen
 - > Histogramm
(Eingabebereich: abs. Häufigkeiten.
Klassenbereich: Klassenmitten
Diagrammdarstellung anklicken)
 - > Rechtecke 2 mal anklicken
 - > Optionen anwählen, Abstand auf 0 stellen
 - > Diagramm einmal anklicken
 - > Diagramm-Assistent anklicken
 - > Weiter 2 mal anklicken, Titel und Anschriften komplettieren

- Boxplots

Bei einem Boxplot werden mehrere Charakteristika der empirischen Verteilung einer Stichprobe in einer einzigen Grafik zusammengefasst. In der Basisform visualisiert ein Boxplot die Quantile der Ordnung $p = 0,25$ und $p = 0,75$ (unteres und oberes Quartil), den Median sowie die beiden Extremwerte der geordneten Urliste. Die beiden Quartile werden durch einen Kasten (Box) repräsentiert, in dem zusätzlich der Median durch einen Punkt markiert ist. Die Extremwerte werden dann mit der Box durch Striche (Whisker) verbunden.



Anforderungen an graphische Darstellungen

- Sinnvolle Darstellung
- Korrekte Darstellung (z.B. Koordinatenursprung)
- vollständige Anschriften (z.B. Quellenangaben)
- Quellenangaben
- keine „Schummelbilder“!

Die 7 Todsünden von graphischen Darstellungen

- Kommentar fehlt
- Achsen falsch oder gar nicht angeschrieben
- Skalen auf Achsen verzerrt, nicht bei 0 beginnend od. unterbrochen
- Zeitreihen: Tendenziöse Wahl des Anfangsjahres
- Flächen oder Volumendarstellungen im falschen Verhältnis
- Tendenziöse Wahl von Vergleichszahlen
- keine neutrale Farbwahl

Entwicklung der LKB Expert-Fonds

Inserat der Luzerner Kantonalbank in „Luzerner Neue Zeitung“
(Ausgabe vom 27.10.01)

