BERUFSAKADEMIE SACHSEN
STAATLICHE STUDIENAKADEMIE
RIESA
UNIVERSITY OF COOPERATIVE EDUCATION

Chi²-Unabhängigkeitstest

Ziel des Chi²-Unabhängigkeitstest

Daten: Sie haben z.B. zwei Gruppen (Frauen / Männer) und für alle Befragten die Zustimmung / Ablehnung zu einem Thema.

Anzahl Variablen:

Skalenniveau: nominal (oder höher, wenn Kategorien gebildet werden, aber dann bieten sich meist andere statistische Verfahren an.)

Problemstellung: Sie wollen nun wissen. Stimmen Frauen eher oder weniger zu im Vergleich zu den Männer.

Sinn des Tests: H_0 =Die Häufigkeiten in den untersuchten Gruppen sind gleich.

BERUFSAKADEMIE SACHSEN
STAATLICHE STUDIENAKADEMIE
RIESA
UNIVERSITY OF COOPERATIVE EDUCATION

Varianzanalyse
ANalysis **O**f **VA**riance **(ANOVA)**

ANOVA: Einfaktorielle Varianzanlayse

Problem: Sind die Mittelwerte in drei (oder mehr) Gruppen unterschiedlich groß?

Nullhypothese = Mittelwerte in allen untersuchten Gruppen sind gleich

Skalenniveau:

abhängige Variable: metrisch;

unabhängige Variable: nominal/ordinal mit zwei Ausprägungen (z. B. Geschlecht; Gruppe A/B)

Vorname	Leistungstest	**Gruppe**
Ilsa	22	A
Peter	31	A
Lisa	12	B
Klaus	46	B
Max	34	C
Mara	23	C

Zweifaktorielle Varianzanalyse mit Messwertwiederholung: Überlegung

	Früh	Mittags
Stichprobe 1	1	4
	2	5
Stichprobe 2	4	1
	5	2

Welche Mittelwerte haben Stichprobe 1 und Stichprobe 2 jeweils?
Wenn ich nur Stichprobe 1 mit Stichprobe 2 vergleichen würde, würde ich keinen Unterschied feststellen, da jeweils die Mittelwerte (in diesem Beispiel) identisch sind.

Welche Mittelwerte haben „Früh“ und „Mittags“ jeweils?
Wenn ich nur früh und Mittags vergleichen würde, würde ich keinen Unterschied feststellen, da jeweils die Mittelwerte (in diesem Beispiel) identisch sind.

BERUFSAKADEMIE SACHSEN
STAATLICHE STUDIENAKADEMIE
RIESA
UNIVERSITY OF COOPERATIVE EDUCATION

Wiederholung

Quiz Teil I

Frage	Richtig	Falsch
Bei der einfachen linearen Regression beeinflusst ein oder mehrere Variablen die abhängige Variable.	X	
Das konstante Glied der einfachen linearen Regressionsanalyse entspricht dem Wert des Y-Achsenschnittpunktes der linearen Regressionsgeraden.	X	
Der 1. Schritt in jedem stat. Testverfahren besteht in der Entscheidung, ob die Nullhypothese oder die Alternativhypothese getestet werden soll.		X
Der Korrelationskoeffizient kann nur Wert zwischen 0 und 1 annehmen		X
Der Median wird von Ausreißern beeinflusst.		X
Der Mittelwert ist gegenüber Ausreißern robust.		X

Quiz Teil II

Frage	Richtig	Falsch
Der Modal wert ist der Wert, der genau in der Mitte der geordneten Verteilung liegt.	X	
Der Regressionskoeffizient entspricht der Steigung der linearen Regression	X	
Die Spannweite wird nie von Ausreißern beeinflusst.		X
Die Standardabweichung berechnet sich als positive Wurzel aus der Varianz.	X	
Die Wahrscheinlichkeiten aller möglichen Elementarereignisse eines Zufallsvorgang ergeben zusammenaddiert den Wert 2.		X
Diskrete Variablen mit sehr vielen Ausprägungen gelten auch als quasi stetig.	X	

Quiz Teil III

Frage	Richtig	Falsch
Ein korrekt durchgeführter stat. Test gestattet eine definitive Aussage über die Korrektheit von Null- und Alternativhypothese.		X
Ein Korrelationskoeffizient von -0,85 deutet auf eine starke lineare Korrelation hin?	x	
Ein Zufallsexperiment ist die beliebig häufige Wiederholung eines Zufallsvorgang unter gleichen Rahmenbedingungen.	X	
Eine zufällig gezogene Stichprobe mit hoher Rücklaufquote ist unabhängig von ihrem Umfang stets repräsentativ.		X
Erwartungswert, Median und Modus einer normalverteilten Variablen alle ungleich groß.		X
Es gibt keine Zufallsvariablen die diskret sind.		X

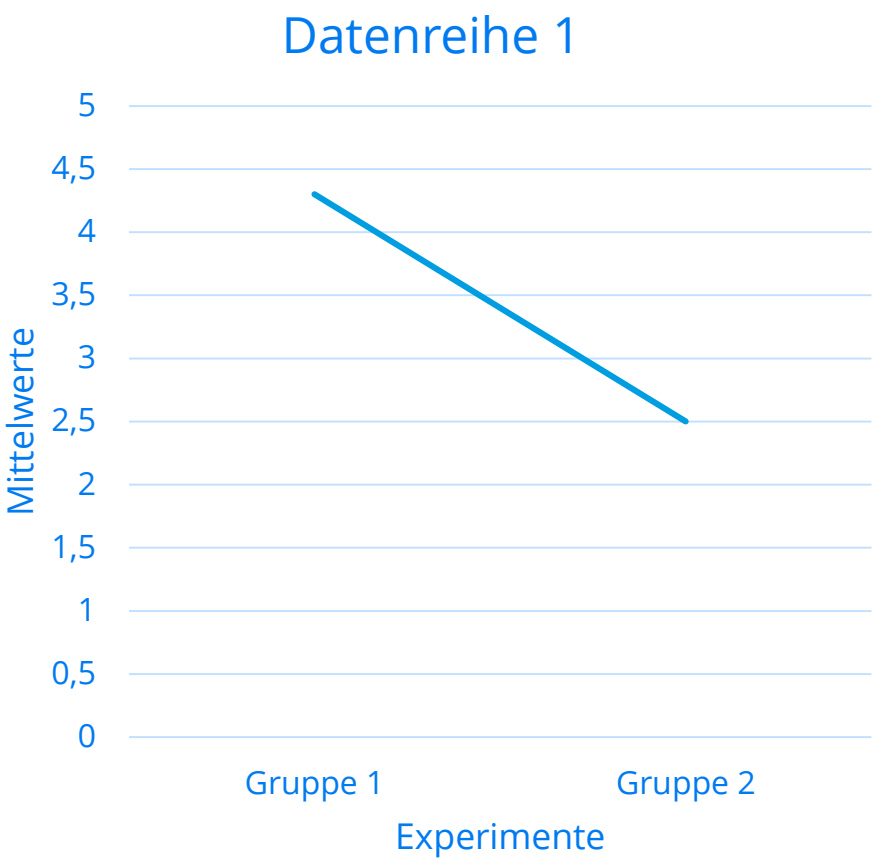
Quiz Teil IV

Frage	Richtig	Falsch
Je mehr Hypothesen man an einem Datensatz testet, desto höher wird die Wahrscheinlichkeit, dass eine davon fehlerhaft als zutreffend angenommen wird.	X	
Kreisdiagramme eignen sich eher für stetige als für diskrete Daten.		X
Nominalskalierte Daten können in eine natürliche Reihenfolge gebracht werden.	X	
Ordinalskalierte Daten können in eine natürliche Reihenfolge gebracht werden.	X	
Stetige Daten sollten vor der Erstellung von Säulendiagrammen klassiert werden.	X	
Streudiagramm zeigen die Verteilung von zwei Variablen.		X

BERUFSAKADEMIE SACHSEN
STAATLICHE STUDIENAKADEMIE
RIESA
UNIVERSITY OF COOPERATIVE EDUCATION

Kleine Fehler finden

Schlechtes Beispiel: Grafik



- Die Abbildung erhebt den Anschein, dass ein Zeitverlauf stattfindet.

Schlechte Beispiele: Tabelle

	Unternehmen A	Unternehmen B	N
Frauen	700	500	1200
	31,111%	22,222%	53,333%
Männer	400	650	1050
	17,777%	28,888%	46,666%
n	1100	1150	2250

Welche Kritikpunkte sehen Sie?:

- Einfache Zahlen nichts rechtsbündig, eine Prozentzahl weicht ab vom Layout der Prozentzahlen.
- Alle Rahmen dargestellt: Besser auf fast alle Linien verzichten, außer auf die unbedingt notwendigen.
- Prozentzahlen auf die 3. Nachkommastelle ist übertrieben. Sind sie überhaupt notwendig?
- Die Angabe fehlt, worauf sich die Prozentangaben beziehen
- Überschrift und Tabellen Beschreibung fehlt
- Das n für die Stichprobengröße ist einmal kleine und einmal groß geschrieben

Schlechte Beispiele: Manipulation des Mittelwerts

Mittelwert Notendurchschnitt von zwei Klassen. Welche Schlussfolgerung könnten Sie ziehen? Ist diese sinnvoll?

Klasse 1
1
1
1
1
1
5
Mittelwert: 1,6
Median: 1,5

Klasse 2
1
1
1
2
2
3
Mittelwert: 1,6
Median: 1,5

Schlechte Beispiele: Prozente

In einer Analyse eines Fragebogens steht die Aussage: „Der Anteil der Männer (40 %) war in der Stichprobe geringer als der von Frauen (60 %).“ Wie ensteht die Aussage bei einer Gesamtzahl von 5 Befragten? Warum ist es nicht sinnvoll, diese Aussage zu machen?

Rohdaten:

Nummer	Geschlecht
1	W
2	M
3	W
4	W
5	M

Stichproben wählen

Befragung zum Thema Lärmbelästigung von Anwohnern.
Stichprobe: Alle Anwohner, die Mitglieder des Vereins „Grüne Stadt“ sind.
Was ist nicht gut an dieser Stichprobe?

Problem: Die Meinung der Mitglieder ist verzerrt.

Bezugspunkte ändern

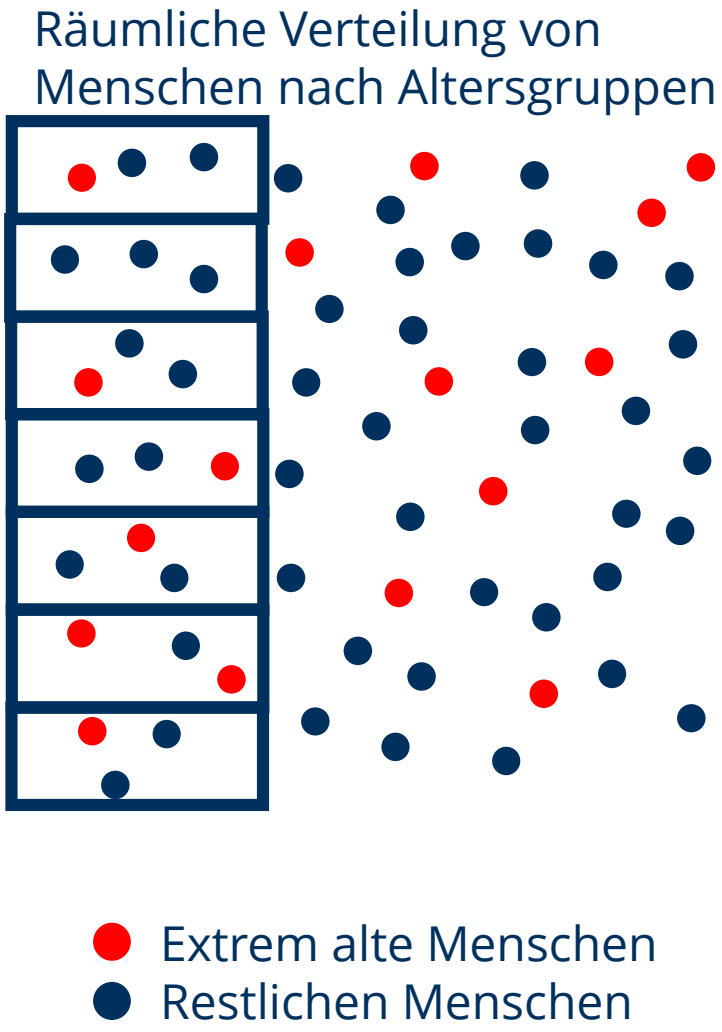
Ein Sportverein hat die Gruppen Junioren und Senioren. Leute mittleren Altern dürfen frei wählen. Was passiert, wenn der 35-jährige wechselt?

	Junioren	Senioren
	5	
	10	40
	15	45
	20	50
	25	55
	30	65
	35	70
Mittelwert **vor** Wechsel:	20	57,1

	Junioren	Senioren
	5	35
	10	40
	15	45
	20	50
	25	55
	30	65
		70
Mittelwert **nach** Wechsel:	17,5	54,4

Stichprobengröße: Ein kleines Dorf mit sehr alten Leuten

- Es gibt kleine Dörfer in dem sehr viele Menschen sehr alt sind.
- Liegt das am gesunden Lebensstil oder auch an der Statistik?
- Annahme: Die Extrem alte Menschen und sind zufällig verteilt auf einem Gebiet. Das Gesamtgebiet wird nun in Städte und Dörfer unterteilt.



Fehlschluss: Große Stichproben sind immer gut

Je größer die Stichprobe, desto eher werden Sie Unterschiede finden. Auch wenn die Unterschiede noch so klein sind (für Zusammenhänge gilt das auch).

Gruppe 1	Gruppe 2
1	2
2	3

p-Wert des t-Tests: 0,293

Gruppe 1	Gruppe 2
1	2
2	3
1	2
2	3

p-Wert des t-Tests: 0,050

Trends zu gewagt

Körpergröße von 18 jährigen Männern in Metern. Welcher Trend lässt sich anhand der Daten formulieren. Wäre das sinnvoll?

1970: 1,80

1990: 1,85

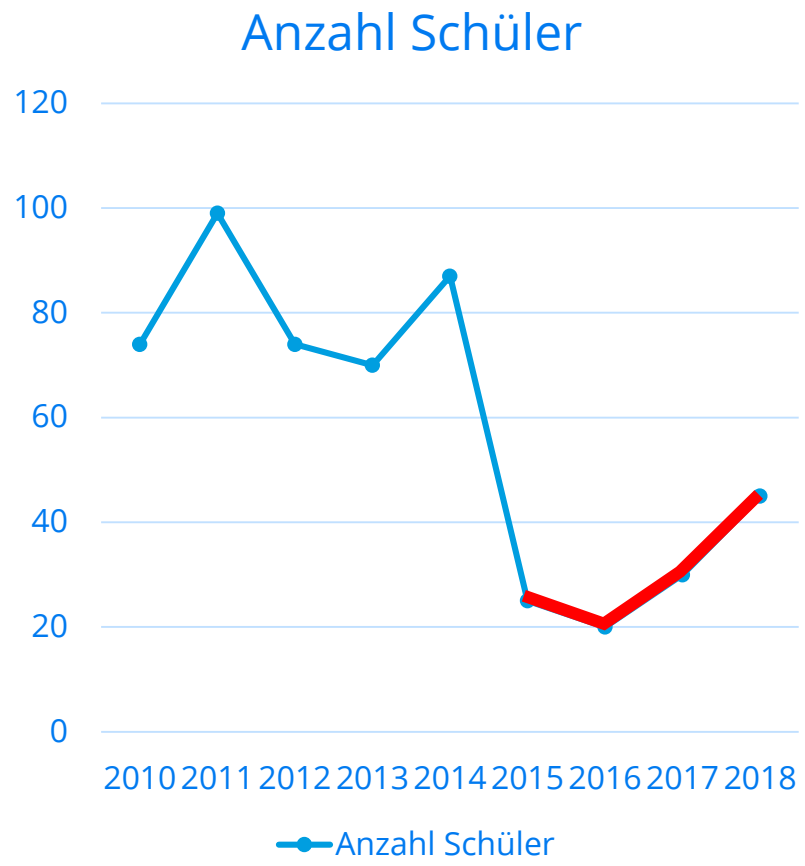
2010: 1,90

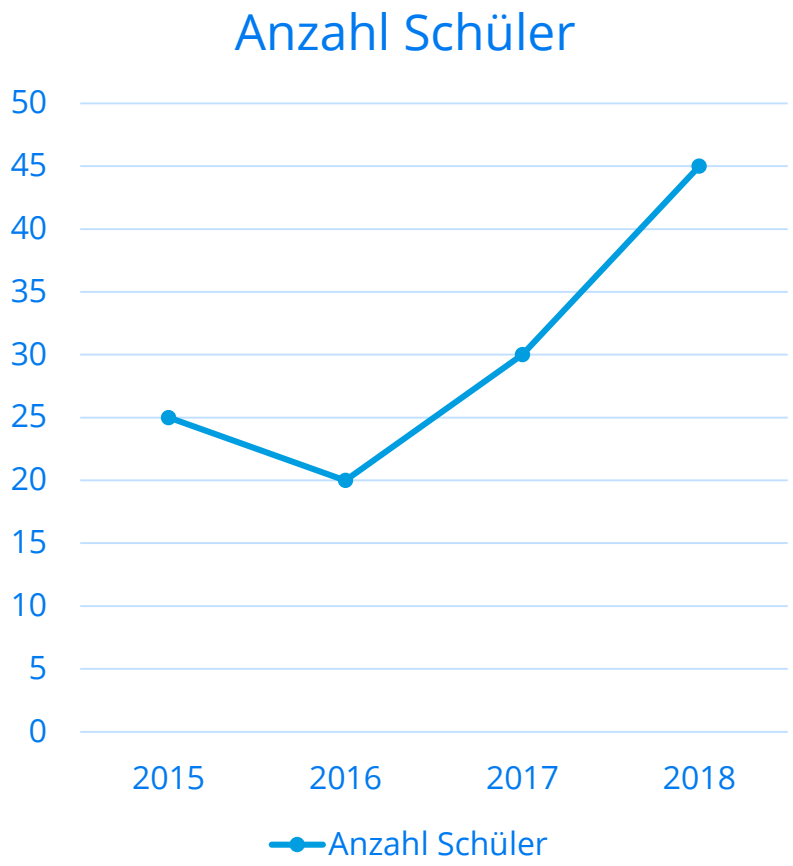
2030: 2,00 ?

Es wäre möglich, den Trend zu formulieren, dass Männer aller 20 Jahre um 5 Zentimeter größer sind. Aber wissenschaftlich begründet ist das nicht.

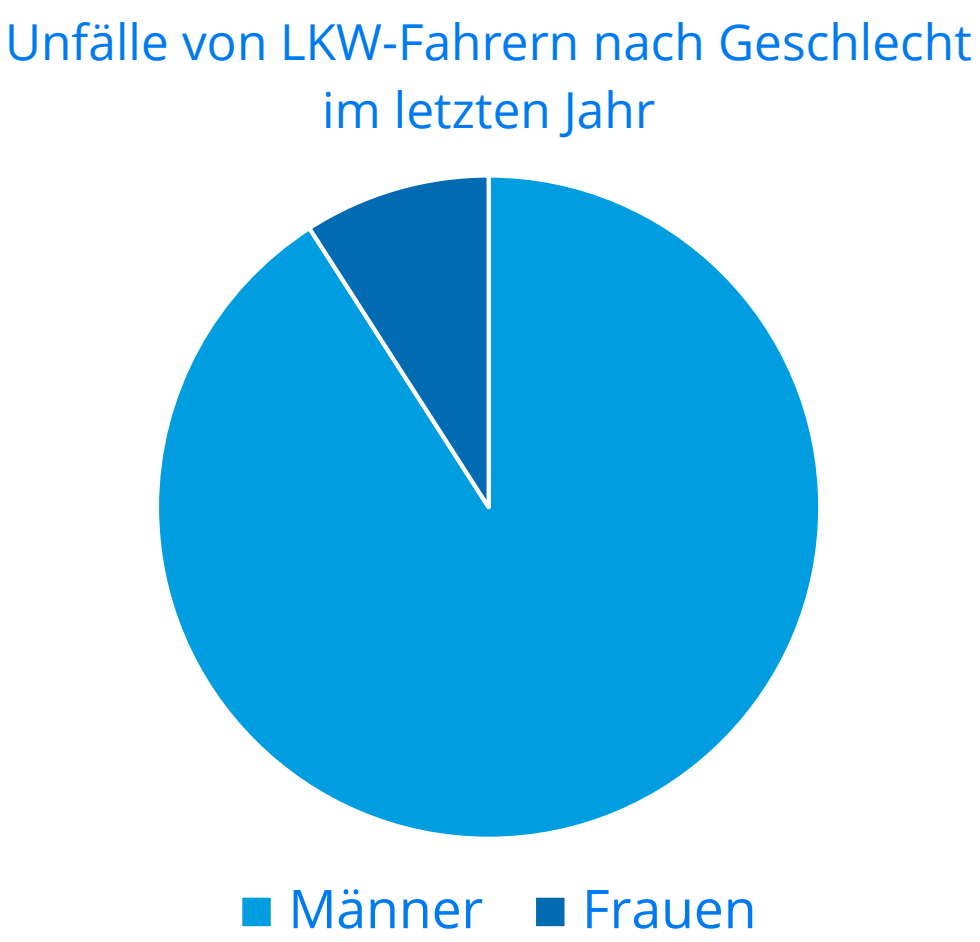
Quelle: Statista. Lügen mit Statistiken https://de.statista.com/statistik/lexikon/definition/8/luegen_mit_statistiken/ Abgerufen: 10.12.2018

Trends zurechtbiegen





Bezugsrahmen verheimlichen



Ihr Unternehmen analysiert die Unfallstatistik bei den angestellten LKW-Fahrenden nach Geschlecht. Jemand zeiht den Schluss: es müssen mehr Frauen eingestellt werden, weil die weniger Unfälle machen.

Der Bezugsrahmen fehlt.

Z.B.

Von den 10 LKW-Fahrerinnen hatte eine einen Unfall.

Von den 100 LKW-Fahrerin hatten 10 einen Unfall.

Prozente können schön klingen

Partei X 2010: 2 Frauen
Partei X 2015: 4 Frauen

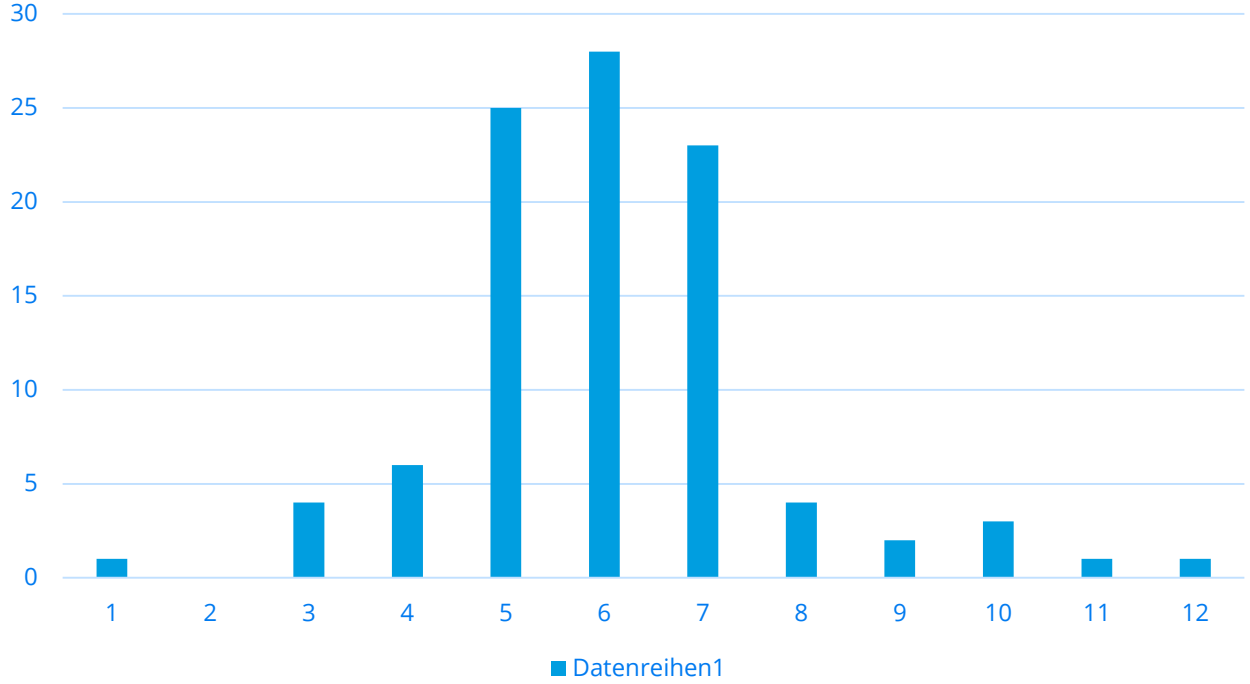
Verdopplung des Frauenanteils in einer Partei X

Partei Y 2010: 100 Frauen
Partei Y 2015: 120 Frauen

Anstieg um nur 20 Prozent in Partei Y

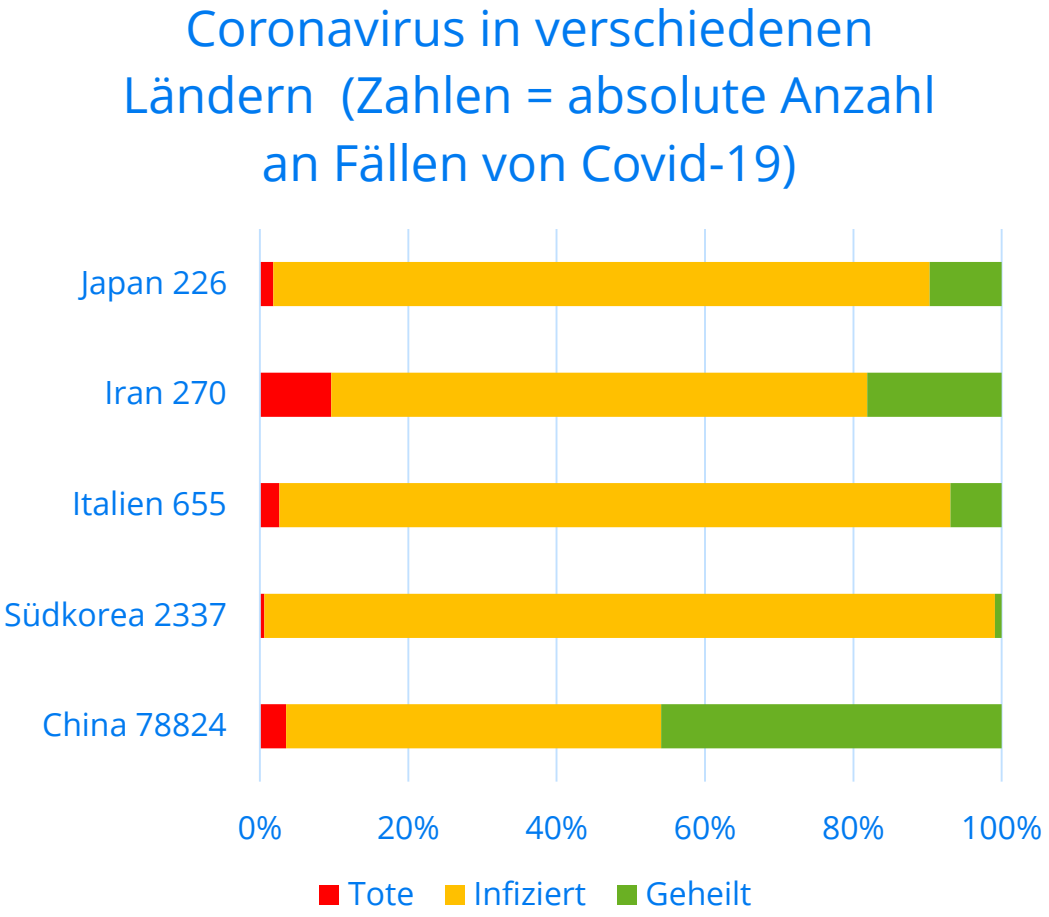
„Normalerweise passiert das nicht“

Abgebildet sind die Lieferzeiten Ihres Möbelwarengeschäfts. Ein Kunde beschwert sich, er hat eine Lieferzeit von 12 Tagen und meint, das ist doppelt so lang wie der Durchschnitt. Das ist doch nicht normal.



Wenn Normal bedeutet, es liegt in der Normalverteilung dann können auch doppelt so lange und halb so lange bedeuten … aber eben nur selten.

Zeitpunkt der Statistik



Datenquelle de.Statista.com
28.02.2020

Probleme mit dieser Statistik

- Wer wird in welchem Land als krank erfasst
- In China war die Krankheit zu dem Zeitpunkt viel länger → daher waren dort mehr geheilte und
- Gesundheitssysteme variieren zwischen den Ländern