

Vorschläge für einen datenorientierten Stochastikunterricht in der Sekundarstufe

Stefan Götz¹ Victoria Döller²

¹Fakultät für Mathematik
Universität Wien
Oskar-Morgenstern-Platz 1
A-1090 Wien
stefan.goetz@univie.ac.at

²Fakultät für Informatik
Universität Wien
Währinger Straße 29
A-1090 Wien
victoria.doeller@univie.ac.at

13. Februar 2020

- 1 Vorbemerkung
- 2 Fünfte bis achte Schulstufe
- 3 Didaktisches Resümee
- 4 Übergang Unter- zur Oberstufe
- 5 Oberstufe

- 1 Credo: Arbeiten mit **realen** Daten, Quellen: **selbst** erheben oder **Internet**
- 2 Dazu: **Excel** oder **GeoGebra**, wobei nur **Standardfunktionen** verwendet werden
- 3 Motiv: **Authentizität**, nicht unbedingt Lebensnähe
- 4 eventuell aufwändige Recherche und Aufbereitung → **keine** ständige Indikation!
- 5 **Keine Scheu** vor Technologieeinsatz und Sackgassen!

2.1 Modellierung stochastischer Situationen		15	
			
Vorgang	ein Schüler springt	Werfen eines Würfels	Wachstum einer Getreideähre
Merkmal	Zensur für die Weite	Augenzahl	Länge
mögliche Ergebnisse	Zensuren 1 bis 6	1, 2, 3, 4, 5, 6	Längen zwischen 5 cm und 15 cm
Bedingungen	z.B. Sprungkraft, Technik, Anlaufweite, Windbedingungen	z.B. Wurftechnik, Unterlage, Würfel	z.B. Bodengüte, Erbanlagen, Wetter

Abb. 2.2 Beispiele zur Prozessbetrachtung aus *Mathematik: Duden 6*, S. 92; mit freundlicher Genehmigung von © Cornelsen Schulverlage GmbH, Berlin. All Rights Reserved

Abbildung: Aus KRÜGER, SILL & SIKORA 2015, S. 15

Zur heuristischen Orientierung (ebd.):

- 1 Welcher Vorgang läuft mit welchen Objekten oder Personen ab?
- 2 Welches Merkmal interessiert mich? Wie kann ich das Merkmal erfassen?
- 3 Welche Ergebnisse sind möglich?
- 4 Welche Bedingungen beeinflussen den Vorgang?

Datenorientierter Zugang

- interdisziplinärer Unterricht
- Hervorhebung realer Situationen und Daten
- aktives, entdeckendes Lernen
- experimenteller Arbeitsstil
- Visualisierung von Konzepten und Begriffen

Datengenerierung:

(vgl. ENGEL 1999, S. 79)

- 1 selbst: Schulweglänge, Spannweite der ausgestreckten Arme, Anzahl der Geschwister, Benzinverbrauch, ...
- 2 aus dem Internet: Schisprungweiten, Bevölkerungsstand Wien, Big Mac-Index, CO₂-Ausstoß, ...



Bevölkerung Wiens – Ausschnitt

Bezirk	Bevölkerungsstand zum Stichtag 1.1.									
	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
Wien	1.610.410	1.632.569	1.652.449	1.661.246	1.671.221	1.680.135	1.689.995	1.702.855	1.717.084	1.741.246
1. Innere Stadt	17.384	17.294	17.285	17.147	16.976	16.623	16.525	16.385	16.290	16.268
2. Leopoldstadt	90.837	91.364	92.170	93.313	93.463	93.822	94.735	95.302	96.582	96.866
3. Landstraße	83.019	83.827	84.208	83.919	83.050	82.979	83.061	83.955	84.574	85.508
4. Wieden	29.394	29.573	30.126	30.189	30.262	30.348	30.331	30.541	30.724	30.989
5. Margareten	51.098	51.882	52.534	52.286	52.434	52.288	52.354	52.753	52.581	53.071
6. Mariahilf	28.891	29.012	29.369	29.463	29.430	29.197	29.323	29.308	29.388	30.117
7. Neubau	29.012	29.161	30.416	29.955	30.054	29.878	30.069	30.062	30.101	30.309
8. Josefstadt	23.309	23.588	23.597	23.613	23.703	23.639	23.435	23.379	23.560	23.930
9. Alsergrund	38.865	39.472	39.264	39.160	39.466	39.167	39.344	39.315	39.431	39.968
10. Favoriten	159.105	163.695	167.238	169.413	171.113	173.227	174.749	176.558	178.358	182.595

Abbildung: <https://www.wien.gv.at/statistik/bevoelkerung/tabellen/bevoelkerung-bez-zr.html>

Zwei klassische Arbeiterbezirke

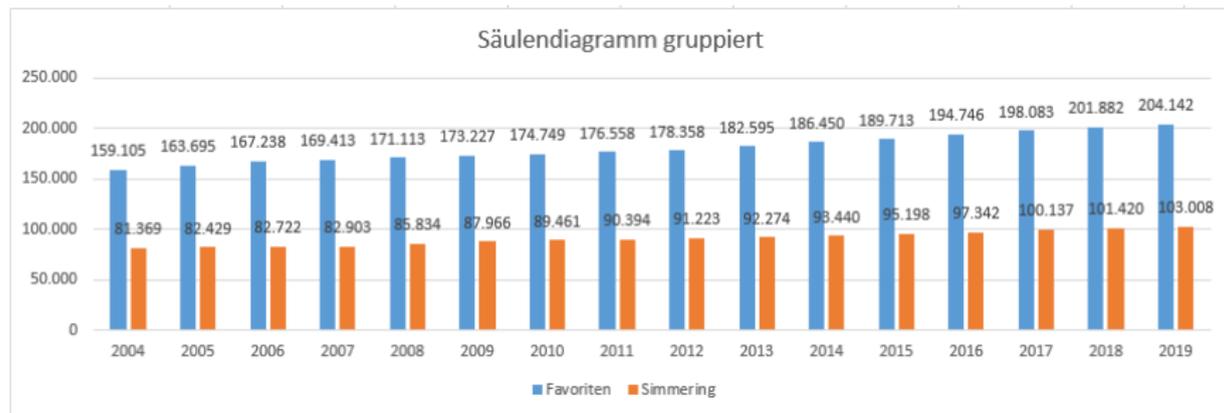
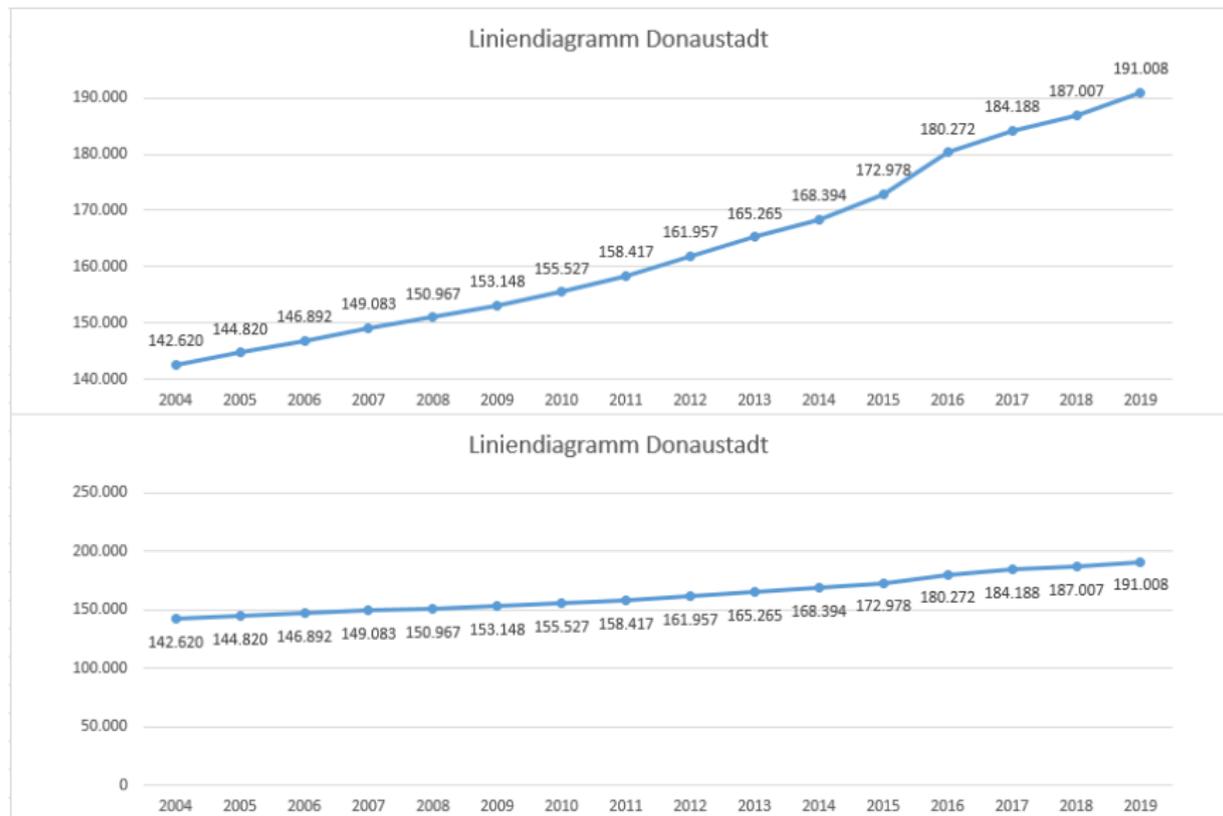


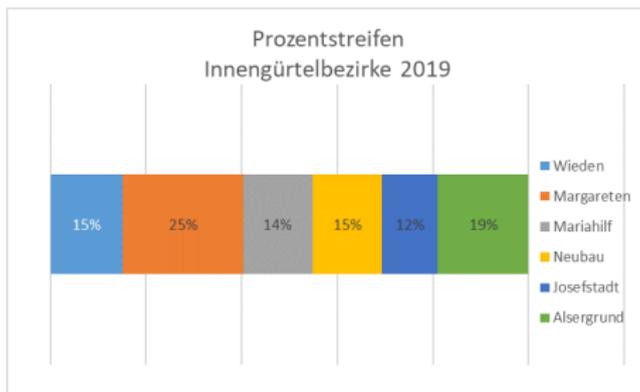
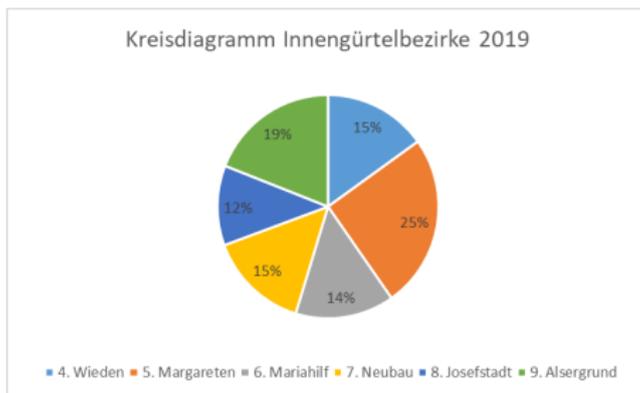
Abbildung: Der zehnte und elfte Wiener Gemeindebezirk

Im Folgenden:

Diagramme mit Excel erzeugt

Eine kleine Manipulation mit einem Flächenbezirk





Big Mac®-Index 2012 für Beispiel eines Balkendiagramms

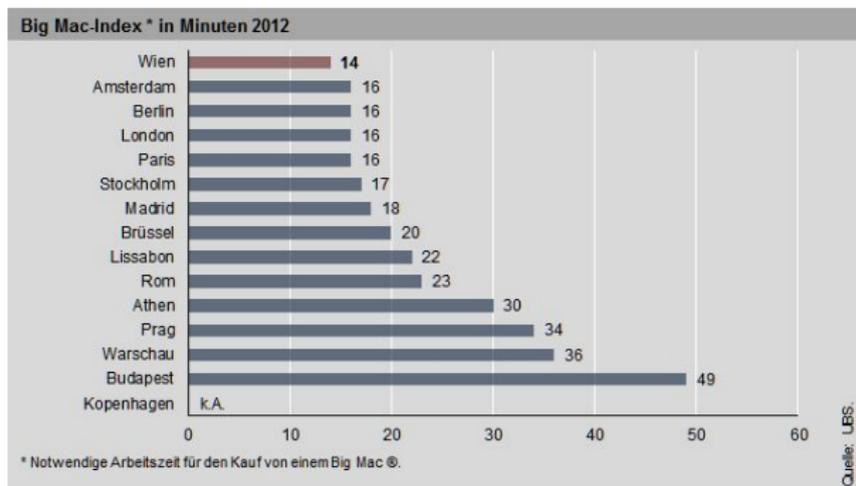


Abbildung:

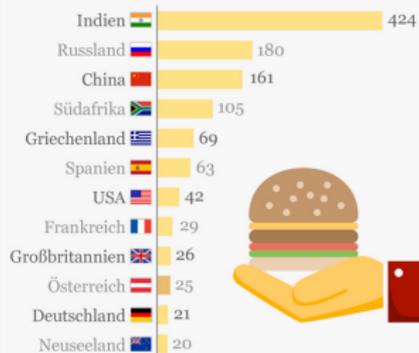
<https://www.wien.gv.at/statistik/staedtevergleiche/international/>

Erklärung (ebd.): Durchschnittlich 14 Minuten (Preis dividiert durch gewichteten Nettostundenlohn aus 14 Berufen) muss in Wien gearbeitet werden, um einen Big Mac® kaufen zu können.

Big Mac®-Index 2017 und 2018

So lang arbeitet man weltweit für einen Big Mac

Minuten, die ein Arbeitnehmer mit Mindestlohnverdienst arbeiten muss, um einen Big Mac bezahlen zu können

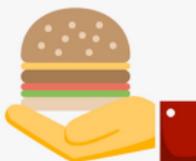


Angenommen wird eine monatliche Arbeitszeit von 173 Stunden,
Stand: Februar 2017

Quelle: wageindicator.org

dSt.at 
@Statista.com

statista 



Big Mac®-Index 2018¹⁾

Wien 17,9 Minuten

Berlin 18,4 Minuten

Rom 23,5 Minuten

Brüssel 26,9 Minuten

Madrid 27,8 Minuten

Warschau 32,8 Minuten

Bukarest 40,2 Minuten

Prag 42,3 Minuten

¹⁾Notwendige Anzahl für den Kauf von einem Big Mac[®].

Abbildung: **links:** <https://www.derstandard.at/story/2000059698458/wie-lange-man-weltweit-fuer-einen-big-mac-arbeiten-muss>
rechts: <https://www.wien.gv.at/statistik/staedtevergleiche/>

Im Folgenden: Diagramme mit GeoGebra erzeugt

CO₂-Ausstoß in g/km der schadstoffärmsten PKWs (nach WLTP:
Worldwide harmonized Light vehicles Test Procedure)

niedrig: 96.0,98.0

9 | 9 9 9 9 9

10 | 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 2

hoch: 103.0,103.0,103.0,103.0

Abbildung: TOP 30 Benziner: <https://www.autoverbrauch.at/top/benzin>

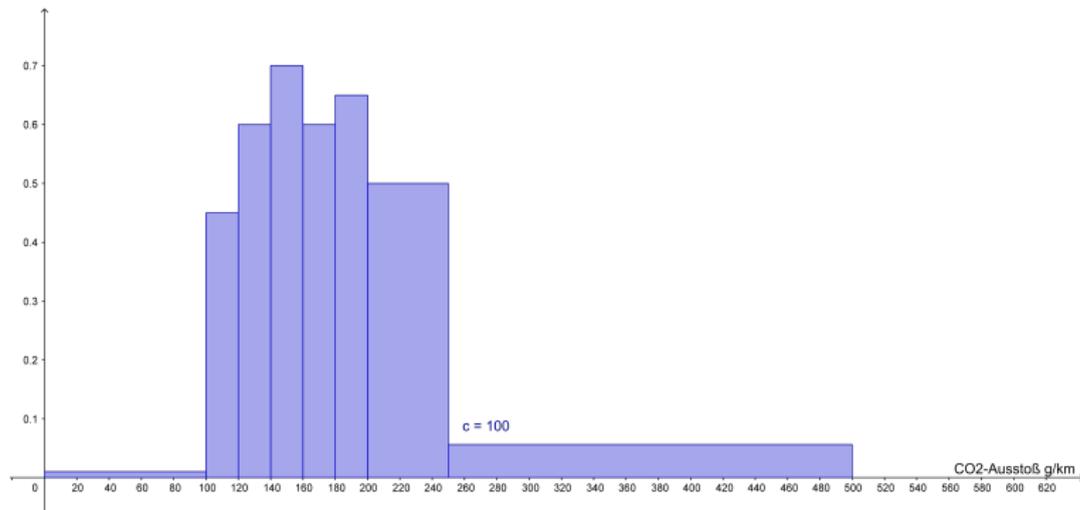
niedrig: 20.0

9 | 8 8 8 9 9 9

10 | 0 0 0 0 0 1 1 1 2 3 3 4 4 4 4 5 5 5 6 6 7 7 7

Abbildung: TOP 30 Diesel: <https://www.autoverbrauch.at/top/diesel>

Histogramm mit ungleich breiten CO₂-Emissionsklassen

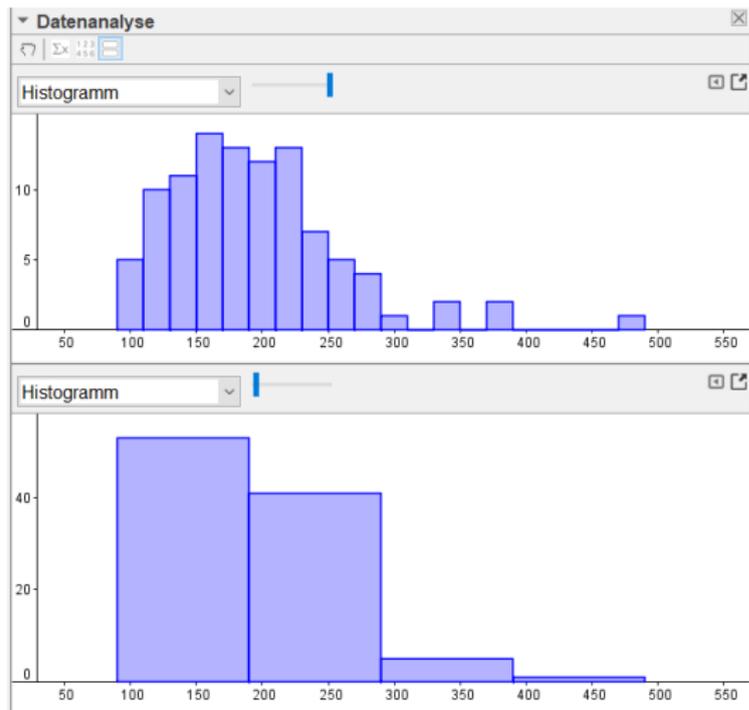


Bundesministerium für Nachhaltigkeit und Tourismus

Statusbericht zu den CO₂-Emissionen neu zugelassener Pkw in Österreich im Jahr 2018, Wien 2019

<https://www.bmnt.gv.at/umwelt/luft-laerm-verkehr/co2-monitoringPKW1.html>

Andere Klasseneinteilung – anderer Eindruck



Absolute Häufigkeiten

Gleichbreite Klassen
oben: 20
unten: vier

Anzahl der geschossenen Tore von FK AUSTRIA WIEN

<http://www.fk-austria.at/de/saison/bundesliga/>

[saison-2019-20/violette-spiele/](#) bzw. [/saison-2018-19/violette-spiele/](#)

in der bisherigen Saison 2019/20

① Urliste: 1, 0, 5, 1, 2, 1, 0, 2, 1, 1, 2, 2, 0, 2, 0, 5, 2, 1

②

Anzahl der Tore	0	1	2	3	4	5
abs. Häufigkeit	4	6	6	0	0	2

③

$$\bar{x}_{18} = \frac{4 \cdot 0 + 6 \cdot 1 + 6 \cdot 2 + 0 \cdot 3 + 0 \cdot 4 + 2 \cdot 5}{18} = 1,5$$

Vergleich zum Vorjahr:

① Urliste: 2, 0, 0, 4, 0, 2, 1, 0, 1, 0, 1, 0, 2, 0, 2, 2, 1, 6

②

Anzahl der Tore	0	1	2	3	4	5	6
abs. Häufigkeit	7	4	5	0	1	0	1

③

$$\bar{x}_{18} = \frac{7 \cdot 0 + 4 \cdot 1 + 5 \cdot 2 + 0 \cdot 3 + 1 \cdot 4 + 0 \cdot 5 + 1 \cdot 6}{18} = 1,3$$

Der Mittelwertabakus (SPIEGEL 1985, S. 16)

Ein handlungsorientierter, anschaulicher Zugang
zur **Schwerpunktseigenschaft** des (gewichteten) arithmetischen Mittels

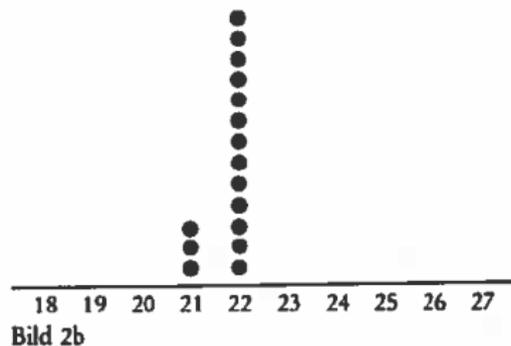
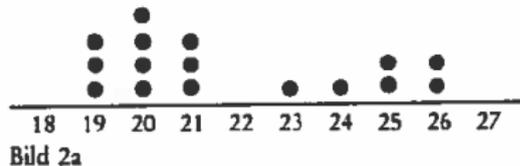
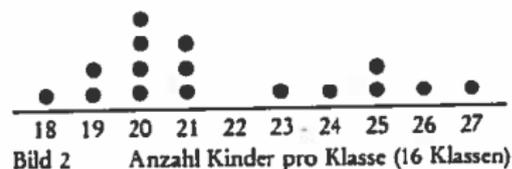


Abbildung: Die Ausgangslage und eine Iteration

Abbildung: Der Endzustand

Der **Mittelwert** übersteigt die unter der linken Säule stehende Zahl um soviel Sechzehntel, wie die Höhe der rechten Säule angibt.

Zeitlicher Temperaturvergleich in Wien

<https://www.data.gv.at/katalog/dataset/69a06550-1ede-4f50-9c36-e7fb5cf6e7e8>

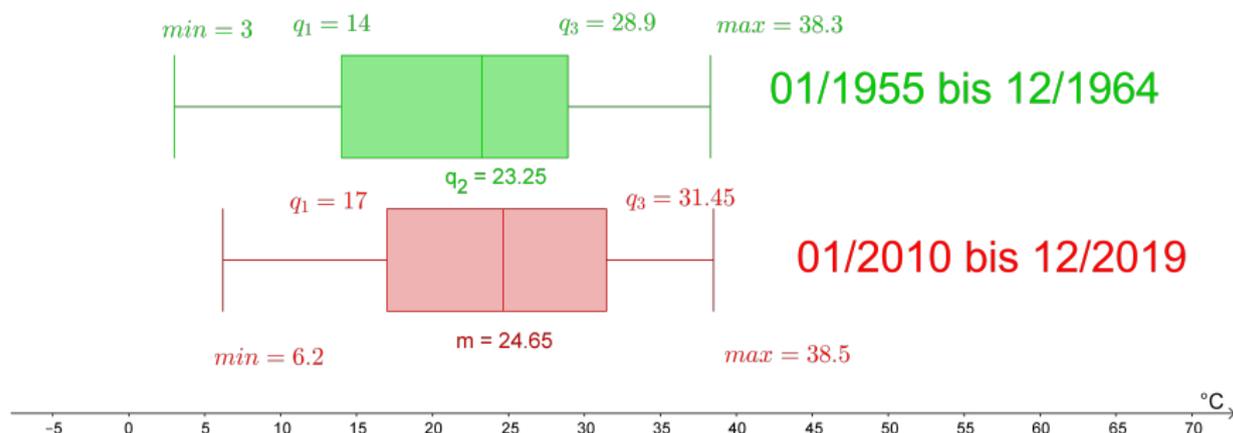


Abbildung: Tagesmaxima über jeweils 120 Monate

Zum Vergleich: $\bar{x}_{1955-1964} = 21.75^{\circ}\text{C}$ und $\bar{x}_{2010-2019} = 24.02^{\circ}\text{C}$

- 1 Erkennen der **Notwendigkeit** von **Daten**: aus Überzeugungen werden empirische Daten (“recognition of the need for data”).
- 2 **Flexible Repräsentation** von Daten (“transnumeration”): Jede neue Problemstellung wirft die Frage nach einer sinnvollen Analyse- methode der Daten auf.
- 3 Einsicht in die **Variabilität** statistischer Daten (“consideration of variation”):

Umfrage
Neue Stichprobenziehung } → andere Ergebnisse, obwohl sich
Daten } Rahmen nicht ändert

- 4 Verbinden von **Kontext** und Statistik (“integrating the statistical and contextual”)

(EICHLER & VOGEL 2009, S. XII f.)

Didaktisches Leitbild: „Datendetektiv(in)“

vgl. GÖTZ & SÜSS-STEPANCIK 2012 und LEAVY & HOURIGAN 2016

Zwei Ausprägungen

- 1 „Während die ‚Wanderer‘ ihren (Analyse-),Weg‘ kennen, einen Analyseplan vor Augen haben und zielorientiert auf die Daten zusteuern,
- 2 gehen die ‚Wunderer‘ eher unvoreingenommen an die Daten heran und lassen sich von den Auffälligkeiten und Mustern in den Daten inspirieren und lenken und bauen darauf ihre Theorien auf.

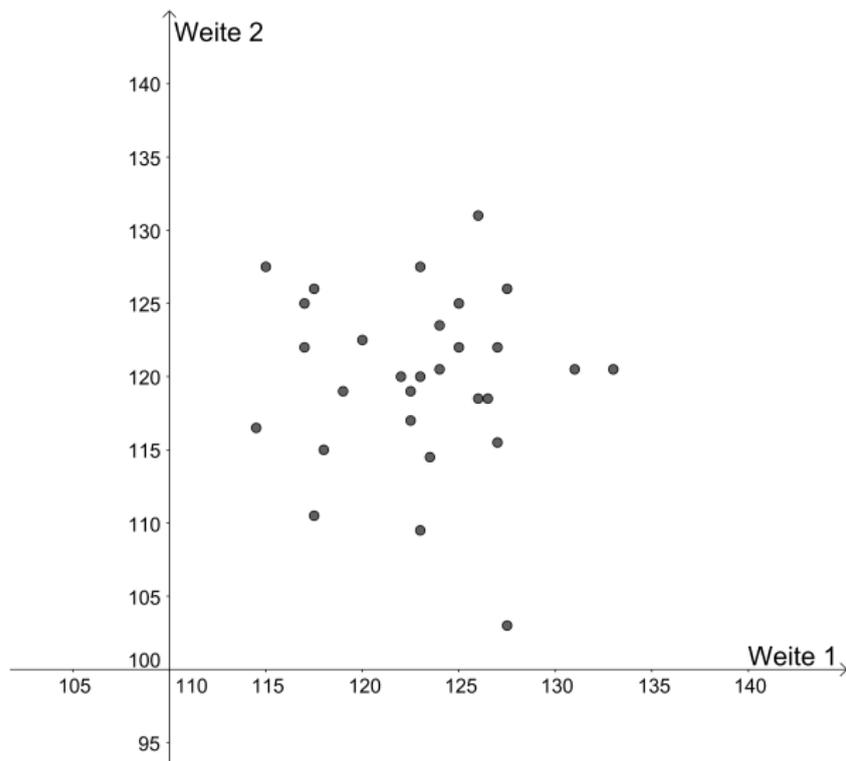
Idealerweise mischen sich die beiden Vorgehensweisen.“ (BIEHLER & FRISCHEMEIER 2015, S. 8)



Vierschanzentournee Innsbruck Berg Isel 2020

Weiten (in m) der 30 besten Springer:

<https://www.vierschanzentournee.com/de/ergebnisse/einzelergebnisse#>



$$r = -0.04$$

Der Grund?

<https://www.tt.com/artikel/16502598/norweger-lindvik-siegt-auch-auf-dem-bergisel-kraft-wird-vierter>

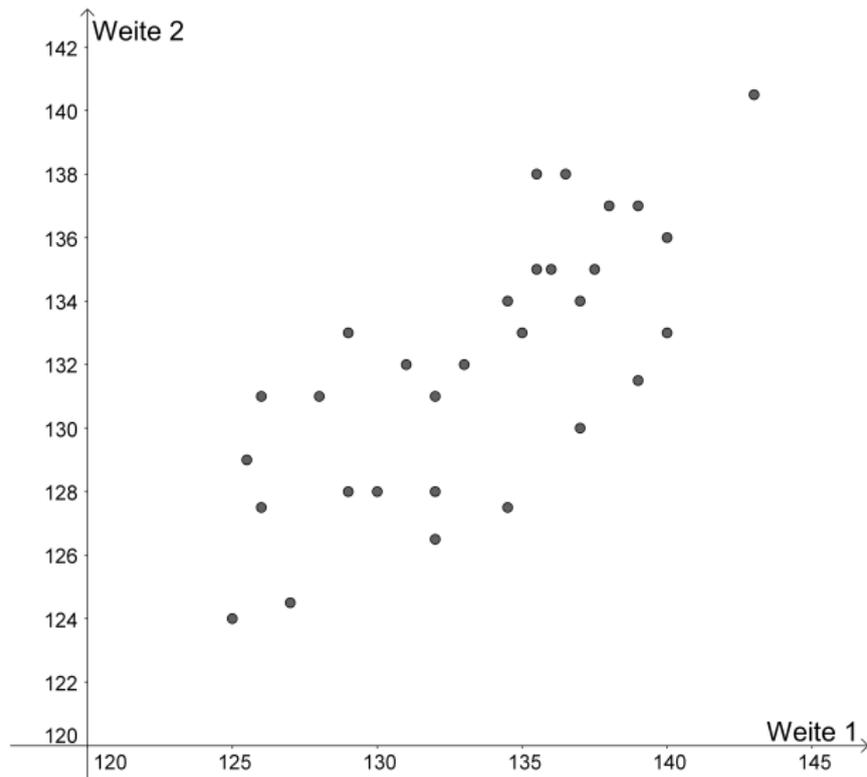


Abbildung: „Innsbruck – Ein turbulentes Bergiselspringen hat die Gesamtwertung in der Vierschanzentournee durcheinandergewirbelt. Bei **wechselnden Windbedingungen** siegte erneut die norwegische Unbekümmertheit namens Marius Lindvik.“ (Hervorhebung S. G.)

Vierschanzentournee Bischofshofen 2020

Weiten (in m) der 30 besten Springer:

<https://www.vierschanzentournee.com/de/ergebnisse/einzelergebnisse#>



$$r = 0.75$$

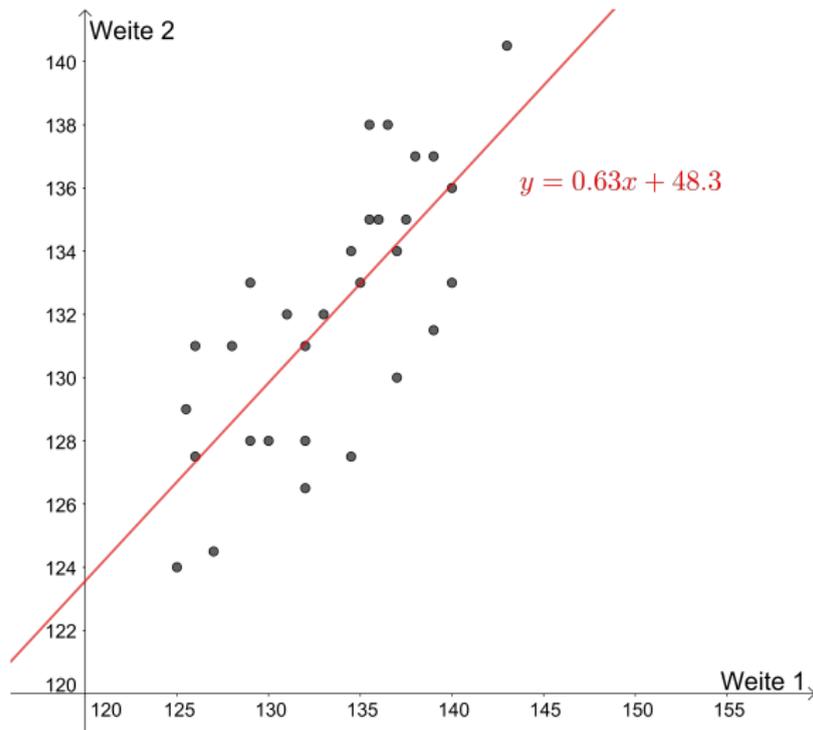


Abbildung: Lineare Regressionsgerade 1

Eine Alternative zur Methode der kleinsten Quadrate

Median-Median-Gerade

- 1 Clustern von Weite 1 in drei gleich große Teile der Größe nach geordnet: $i = 1, 2, 3$

- 2 Medianpunkte in den drei Clustern:

	$z_1^{(i)}$		$z_2^{(i)}$
M_1	(127.5		128.5)
M_2	(134.5		132.5)
M_3	(138.5		135.5)

- 3 Steigung a aus M_1 und M_3 :

$$a = \frac{135.5 - 128.5}{138.5 - 127.5} = 0,\overline{63} = \frac{7}{11}$$

- ④ Schnittpunkt mit y -Achse der drei Geraden mit Steigung a durch M_1 , M_2 und M_3 :

$$128.5 = \frac{7}{11} \cdot 127.5 + b_1 \longrightarrow b_1 = 47.36$$

$$132.5 = \frac{7}{11} \cdot 134.5 + b_2 \longrightarrow b_2 = 46.91$$

$$135.5 = \frac{7}{11} \cdot 138.5 + b_3 \longrightarrow b_3 = 47.36 = b_1$$

⑤ $\bar{b} = \frac{2 \cdot b_1 + b_2}{3} = 47.\overline{21}$

$$y = \frac{7}{11} \cdot x + 47.\overline{21}$$

(EICHLER & VOGEL 2009, S. 95)

Vierschanzentournee Bischofshofen 2020

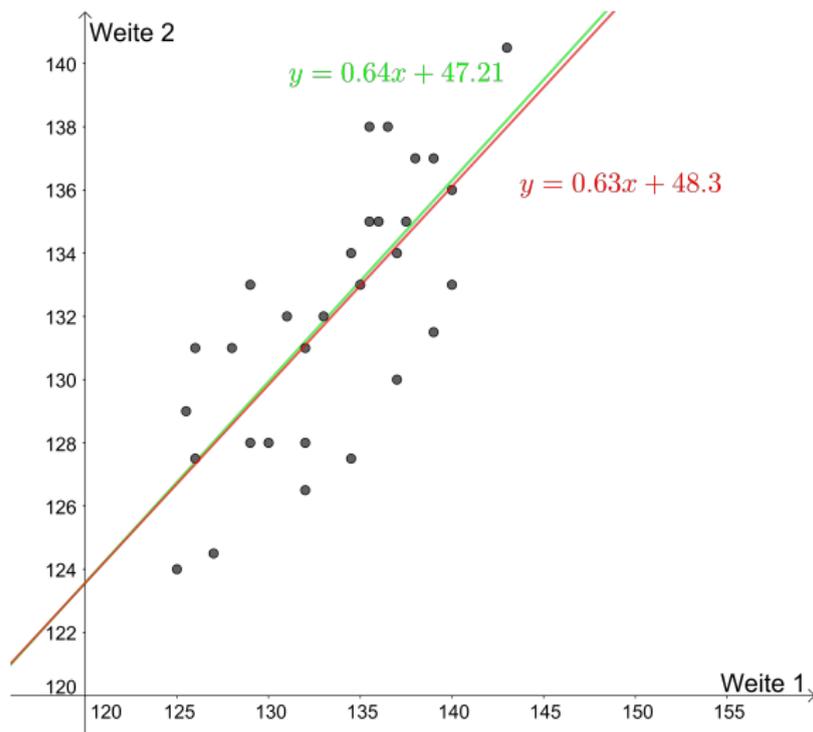


Abbildung: Zwei Regressionsgeraden

Ein einfacher Korrelationskoeffizient zum Zählen

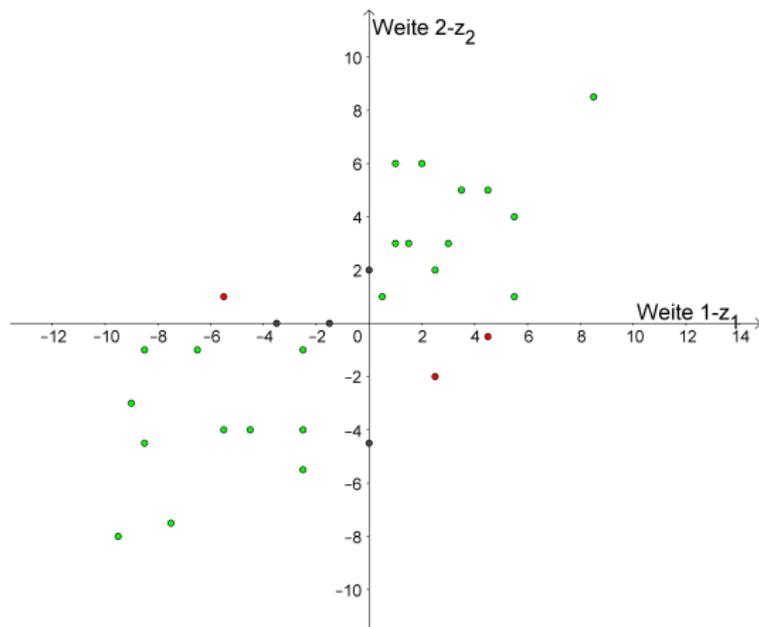


Abbildung: Abstände zu den Medianen z_1 bzw. z_2

$$r_z = \frac{n^+ - n^-}{n}$$

$$r_z = \frac{25 - 5}{30} = \frac{2}{3} = 0.\dot{6}$$

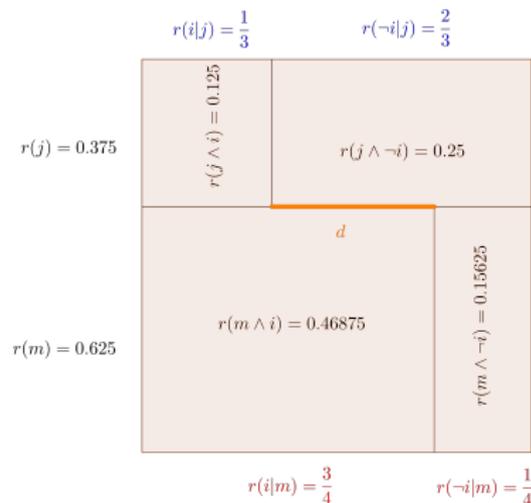
n^+ Anzahl der Punkte im ersten bzw. dritten Quadranten

n^- Anzahl der Punkte im zweiten bzw. vierten Quadranten

(EICHLER & VOGEL 2009, S. 101)

Auf dem Weg zur bedingten Wahrscheinlichkeit

	spielt Instrument	spielt kein Instrument	Σ
Junge	0.125 (4)	0.25 (8)	0.375 (12)
Mädchen	0.46875 (15)	0.15625 (5)	0.625 (20)
Σ	0.59375 (19)	0.40625 (13)	1 (32)



Assoziationsmaß

$$d = r(j|m) - r(i|j) = \frac{3}{4} - \frac{1}{3} > 0 :$$

Mädchen spielen häufiger ein Instrument.

Graphische Vierfeldertafel und Einheitsquadrat (EICHLER & VOGEL 2009, Abschnitt 3.1)

Ein erster Schritt zur Unabhängigkeit

	$r(i j)$	$r(\neg i j)$
$r(j)$	$r(j \wedge i)$	$r(j \wedge \neg i)$
$r(m)$	$r(m \wedge i)$	$r(m \wedge \neg i)$
	$r(i m)$	$r(\neg i m)$

Wir sehen $d = 0$ und

$$r(i|j) = r(i|m) ,$$

das ist äquivalent zu

$$r(i \wedge j) = r(i) \cdot r(j)$$

und zu

$$r(i|j) = r(i) .$$

Anschaulicher Zugang zur Unabhängigkeit (von Ereignissen)!

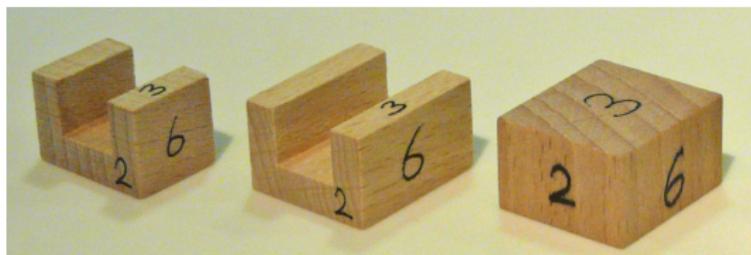


Abbildung: U-Würfel und Quaderwürfel

Quaderwürfel

1	2	3	4	5	6
0.05	0.1	0.35	0.35	0.1	0.05

U-Würfel

1	2	3	4	5	6
0.12	0.08	0.21	0.39	0.08	0.12

(vgl. EICHLER & VOGEL 2009, S. 166, und RIEMER 1988)

Lernen aus Erfahrung — welcher Würfel ist es?

vgl. EICHLER & VOGEL 2009, Abschnitt 6.3

- ① A priori (**vor** der Datengewinnung): **subjektives** Vorwissen

$$P(L) = P(Q) = P(U) = \frac{1}{3}$$

- ② „2“ wird geworfen: A posteriori (**nach** der Datenerhebung)

$$\begin{aligned} P(L|2) &= \frac{P(2|L) \cdot P(L)}{P(2|L) \cdot P(L) + P(2|Q) \cdot P(Q) + P(2|U) \cdot P(U)} = \\ &= \frac{\frac{1}{6}}{\frac{1}{6} + 0.1 + 0.08} = 0.48 \dots \end{aligned}$$

- ③ Analog:

$$P(L|2) = 0.48 \quad P(Q|2) = 0.29 \quad P(U|2) = 0.23$$

Das ist die neue A-priori-Einschätzung!

- 4 „6“ wird geworfen:

$$P(L|6) = 0.65 \quad P(Q|6) = 0.12 \quad P(U|6) = 0.23$$

- 5 **Zusammengesetzte Informationen:**

$$P(L|2 \rightarrow 6) = \frac{P(2 \rightarrow 6|L) \cdot P(L)}{P(2 \rightarrow 6)} = \frac{P(2|L) \cdot P(6|L) \cdot P(L)}{P(2 \rightarrow 6)}$$

mit

$$P(2 \rightarrow 6) = P(2|L) \cdot P(6|L) \cdot P(L) + P(2|Q) \cdot P(6|Q) \cdot P(Q) + \\ + P(2|U) \cdot P(6|U) \cdot P(U)$$

und schließlich

$$P(L|2 \rightarrow 6) = \frac{P(6|L) \cdot \frac{P(2|L) \cdot P(L)}{P(2)}}{\frac{P(2 \rightarrow 6)}{P(2)}} = \frac{P(6|L) \cdot P(L|2)}{P(2 \rightarrow 6|2)}$$

Subjektiver Wahrscheinlichkeitsbegriff

führt zu **intersubjektiver Übereinstimmung** mit $P(Q) = 0.5$ und $P(L) = 0.25 = P(U)$ a priori:

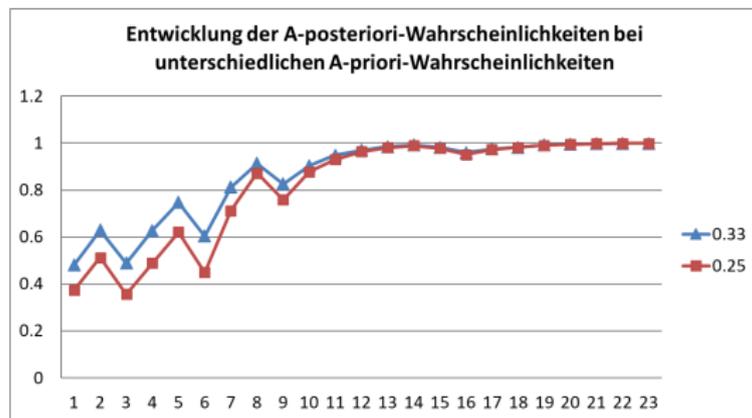


Abbildung: Entwicklung der Einschätzung für LAPLACE-Würfel

für die Wurffolge (vgl. EICHLER & VOGEL 2009, S. 202)

5-5-3-5-5-3-6-6-4-1-2-6-2-2-4-4-6-1-5-5-2-5-5

Die Landtagswahl 2019 in der Steiermark

Konfidenzintervall für den Anteil der ÖVP-Wähler(innen)



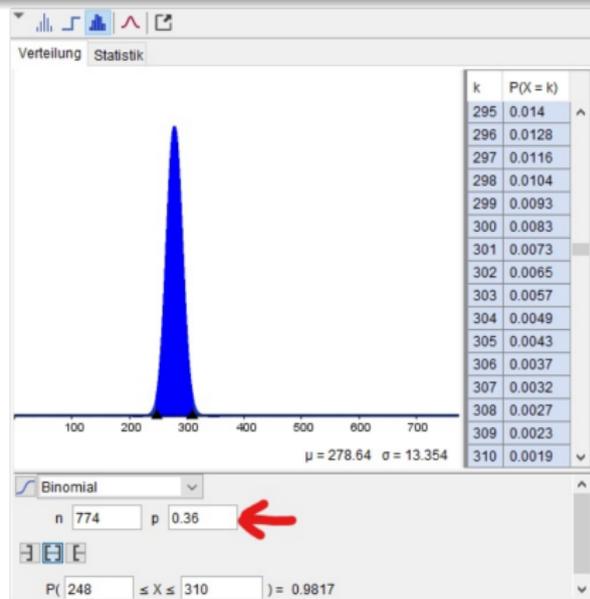
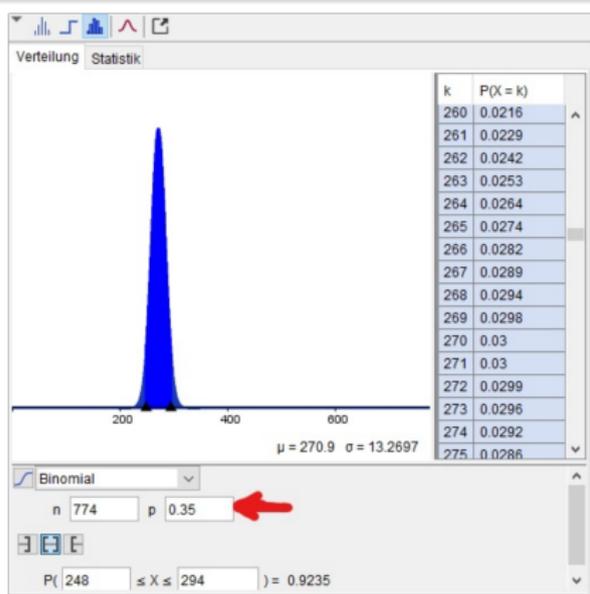
Abbildung: <https://neuwahl.com/wahlumfragen/>

Relevante Daten: 16.11.2019

$n = 774$ und $r_{774} = 0.32$, also $h_{774} = 0.32 \cdot 774 \approx 248$

Herantasten aufgrund der Definition

All jene p mit $0 \leq p \leq 1$, deren γ -Streubereiche den beobachteten Wert \hat{p} enthalten (überdecken), bilden ein bestimmtes Intervall $[p_1; p_2]$. Man nennt es das zum Konfidenzniveau γ gehörige γ -Konfidenzintervall [...] für das unbekannte p (GÖTZ & REICHEL 2013, S. 161).



- ① → Stochastisches Denken durch Drehen an der p -Schraube:

$$p_1 = 0.29 \quad \text{und} \quad p_2 = 0.35$$

- ② Mit Approximation durch die Normalverteilung:

$$0.32 \pm 1.96 \cdot \sqrt{\frac{0.32 \cdot 0.68}{774}} \rightarrow [0.2871; 0.3529]$$

- ③ GeoGebra: „GAUSS-Schätzer eines Anteils“ →

$$[0.2875; 0.3533]$$

- ④ Angegebene Schwankungsbreite $\pm 3.5\%$ verglichen mit

$$\frac{0.3533 - 0.2875}{2} = 3.3\% !$$

Steiermark Gesamt - endgültiges Ergebnis 287 von 287 Gemeinden wurden ausgezählt		
Partei	LT2019	
	Stimmen	
Gesamt	606.528	
Ungültig	4.469	
Gültig	602.059	
- SPÖ	138.572	23,02%
- ÖVP	217.036	36,05%
- FPÖ	105.294	17,49%
- GRÜNE	72.749	12,08%
- KPÖ	36.062	5,99%
- NEOS	32.346	5,37%
- FRANK		
- PIRAT		

Abbildung: https://egov.stmk.gv.at/wahlen/LT2019/LT2019_60000.html

Dazu: 99%-Konfidenzintervall

[0.2772; 0.3636]

Die Nationalratswahl in Österreich 29.09.2019 und danach

Österreich, Endgültiges Endergebnis (inkl. Wahlkarten - Gesamt)

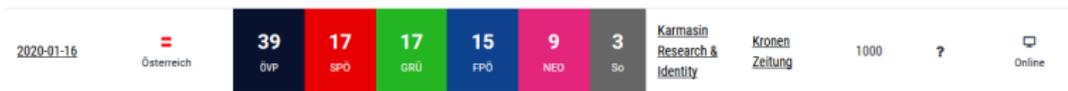
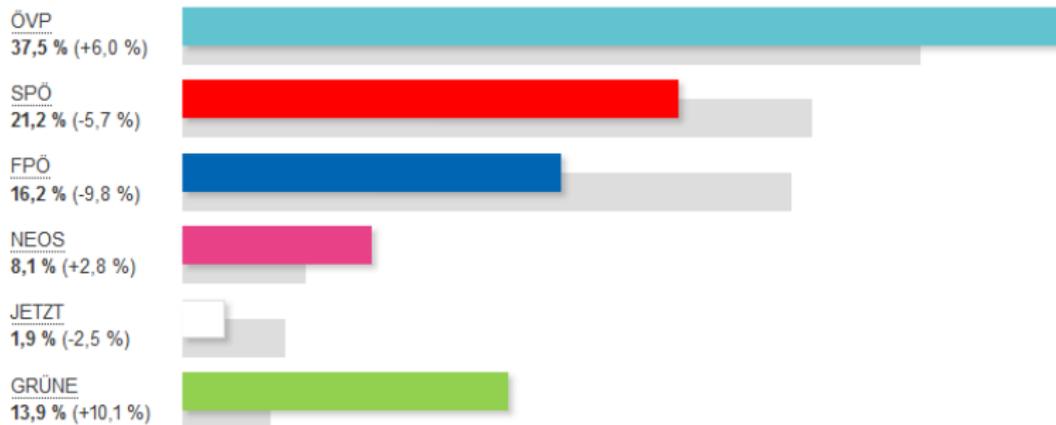


Abbildung: <https://wahl19.bmi.gv.at/> bzw.
<https://neuwal.com/wahlumfragen/> (16.01.2020)

Ist die SPÖ signifikant schwächer geworden?

- ① Zufallsvariable X ... Anzahl der SPÖ-Wähler(innen) unter 1000 zufällig Befragten $\rightarrow X \sim \text{BV}(1000; 0.212)$ unter

$$H_0 : p \geq p_0 := 0.212$$

$$H_1 : p < p_0$$

- ② Ablehnungsbereich für H_0 : $\{0, 1, 2, \dots, k - 1, k\} =: K$

$$P(X \leq k)_{H_0} \leq 0.05$$

$$\sum_{i=0}^k \binom{1000}{i} \cdot 0.212^i \cdot 0.788^{1000-i} \leq 0.05 \quad \text{und} \quad k \text{ max}$$

der **kritische Wert**

Ja! $k = 190$ und $0.17 \cdot 1000 = 170 \in K$

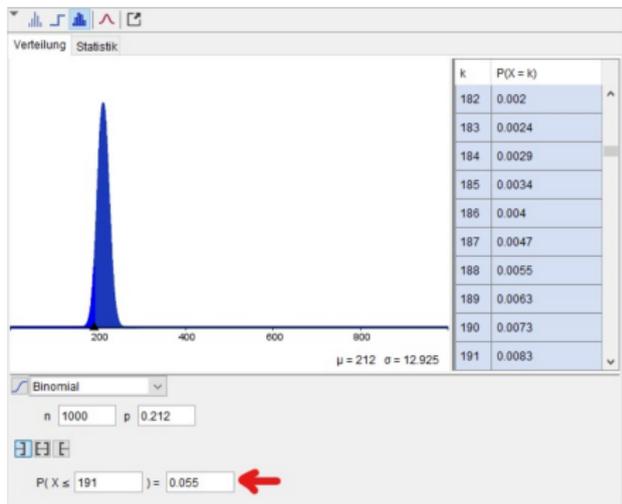
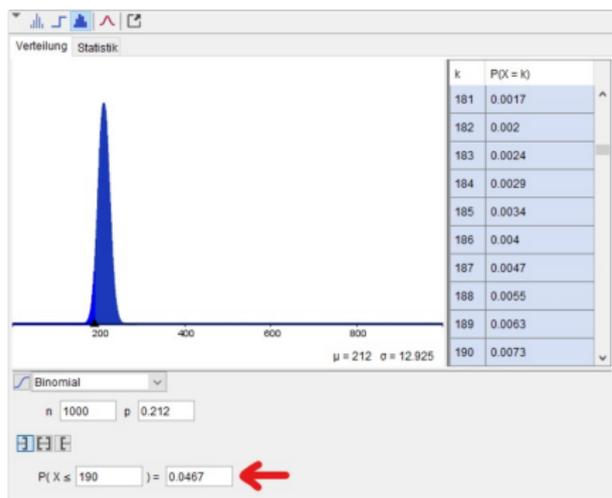


Abbildung: Stochastisches Denken durch Drehen an der k -Schraube!

sogar hochsignifikant:

$k = 181$ für $\alpha = 0.99$ und $170 \in \{0, 1, \dots, 180, 181\}$

- Biehler, R. und Frischemeier, D.: „Verdienen Männer mehr als Frauen?“ — Reale Daten im Stochastikunterricht mit der Software TinkerPlots erforschen. In: *Stochastik in der Schule* **35** (2015), Heft 1, S. 7–17.
- Eichler, A. und Vogel, M.: *Leitidee Daten und Zufall. Von konkreten Beispielen zur Didaktik der Stochastik*. Vieweg+Teubner, Wiesbaden 2009.
- Engel, J.: Computer und Erziehung zur Datenkompetenz. In: G. Kadunz, G. Ossimitz, W. Peschek, E. Schneider und B. Winkelmann (Hrsg.): *Mathematische Bildung und neue Technologien*. Vorträge beim 8. Internationalen Symposium zur Didaktik der Mathematik. Universität Klagenfurt, 28.9.–2.10.1998. B. G. Teubner, Stuttgart · Leipzig 1999, S. 77–84.

- Götz, S. und Reichel, H.-C. (Hrsg.): *Mathematik 8* von R. Müller und G. Hanisch. öbv, Wien 2013.
- Götz, S. und Süß-Stepancik, E.: Daten, Daten, Daten — was sie uns verraten. Die Leitidee „Daten und Zufall“ in der AHS Unter- und Oberstufe. In: *Schriftenreihe zur Didaktik der Mathematik der ÖMG* 2012 (Heft 45), S. 29–42.
- Krüger, K., Sill, H.-D. und Sikora, C.: *Didaktik der Stochastik in der Sekundarstufe I*. Springer Spektrum, Berlin Heidelberg 2015.
- Leavy, A. und Hourigan, M.: Tatort und geheimnisvolle Spieler! Mit Leitfragen die Entwicklung von Statistical Literacy unterstützen. In: *Stochastik in der Schule* **36** (2016), Heft 3, S. 2–9.
- Riemer, W.: RIEMER-WÜRFEL. Klett, Stuttgart 1988.
- Spiegel, H.: Der Mittelwertabakus. In: *mathematiklehren* 8, Februar 1985, S. 16–18.