Unsinn in den Medien – Vom allzu sorglosen Umgang mit Daten:

Konfidenzintervalle

BREXIT, AFD UND JETZT DIE US-WAHL

Darum waren die Prognosen schon wieder falsch!

...

► „Das komplizierte amerikanische Wahlsystem trägt dazu bei, dass Umfragen größere Ungenauigkeiten aufweisen", erklärt INSA-Chef Hermann Binkert. Es beginnt damit, dass im Vergleich zu Deutschland (rund 62 Millionen Wähler) in Amerika 220 Millionen Wahlberechtigte an die Urnen gerufen werden.

Bei 1000 oder 2000 Befragten ist die Fehlerquote deshalb natürlich höher.

(gefunden auf http://www.bild.de/politik/ausland/us-wahlen/trump-sieg-warum-hat-das-keiner-kommen-sehen-48673374.bild.html (Zugriff am 10.11.2016))

Kommentar: Dass Wahlumfragen manchmal das tatsächliche Bevölkerungsbild am Ende nicht widerspiegeln, kann an mehreren Faktoren liegen: Es können sich die Verhältnisse seit den Umfragen noch verändert haben (Skandal kurz vor der Wahl), es können von einem Teil der Befragten nicht die wahren Antworten gegeben worden sein („soziale Wünschbarkeit") oder die Antwortverweigerer haben anders gewählt als die Antwortenden (Nonresponse-Bias). Die Statistik stimmt, aber die Daten sind schlecht – arme Meinungsforscher!
Aber kann es – wie BILD behauptet – ebenso an der größeren Schwankungsbreite ε („Fehlerquote") der Umfragen im Vergleich zu z.B. Deutschland liegen? Rechnen wir für einen tatsächlichen Anteil von 48 % und einer Stichprobe vom Umfang n = 1000 nach:

$$\varepsilon_{USA} = u_{1-\alpha/2} \cdot \sqrt{\frac{\pi \cdot (1-\pi)}{n} \cdot \frac{N-n}{N-1}} = 1{,}96 \cdot \sqrt{\frac{0{,}48 \cdot (1-0{,}48)}{1000} \cdot \frac{220\text{Mio}-1000}{220\text{Mio}-1}} = 0{,}03096545$$

$$\varepsilon_{GER} = u_{1-\alpha/2} \cdot \sqrt{\frac{\pi \cdot (1-\pi)}{n} \cdot \frac{N-n}{N-1}} = 1{,}96 \cdot \sqrt{\frac{0{,}48 \cdot (1-0{,}48)}{1000} \cdot \frac{62\text{Mio}-1000}{62\text{Mio}-1}} = 0{,}03096527$$

Natürlich nicht! Einfache Zufallsstichproben vom Umfang 1000 aus der deutschen bzw. aus der amerikanischen Wahlbevölkerung besitzen unter den gegebenen Umständen jeweils eine Schwankungsbreite von $\pm$ 3,1 %. Denn wenn nur 1000 Personen zufällig ausgewählt werden, dann spielt es in Hinblick auf die Genauigkeit keine Rolle mehr, ob dies aus einer Population von 62 oder 220 Millionen erfolgt. Der zweite Bruch unter der Wurzel ist einfach in beiden Fällen nahezu 1. Die Schwankungsbreiten unterscheiden sich erst an der 7. Nachkommastelle!

(Für den Kommentar verantwortlich: Andreas Quatember, IFAS)