

Das Simpson'sches Paradoxon

Hier zeige ich, welche Fehlinterpretationen und Paradoxa (=scheinbare Widersprüche) durch eine heterogene statistische Masse entstehen können.

Beispiel: Frauenquote der angenommenen Bewerbungen an einer Universität

Die hypothetische amerikanische Universität X ist so prestigereich, dass sie es sich leisten kann, bei den Erstsemester-Bewerbungen zu selektieren.

Dabei darf sie jedoch auf keinen Fall einseitig Frauen oder Männer bevorzugen, Behinderte oder Schwarze ablehnen etc. Ansonsten würde sie gegen die Grundsätze der “*political correctness*” und der *affirmative action*” verstoßen, welche in den USA sehr ernsthafte Themen sind.

Die **Frauenbeauftragte** der Uni veranlasste eine Untersuchung der Bewerbungsstatistik:

- Es bewarben sich 1000 Frauen und 1000 Männer.
- Abgelehnt wurden 470 Frauen, aber nur 360 Männer.

⇒ Die Ablehnquote ist somit 47% bei den Frauen und 36% bei den Männern.

Die Frauenbeauftragte wirft der Uni X deshalb eine “statistisch signifikante” *Diskriminierung der Frauen* vor. 🤪

Die **Uni** erwiedert, das sei falsch. Im Gegenteil, sie habe sowohl in den naturwissenschaftlichen, als auch in den geisteswissenschaftlichen Fakultäten die Frauen bevorzugt:

- In den Naturwissenschaften wurden 20% der Frauen, aber 30% der Männer abgelehnt.
- In den Geisteswissenschaften waren es 50% Frauen und 60% Männer.

⇒ Welche der beiden Kategorien die Aspiranten auch wählten, *stets* war die Ablehnquote bei den Männern um 10% höher.

Also wurden *in Wirklichkeit die Männer diskriminiert!* 🤪

Die Lösung

Es wurde **nichts über den Anteil der Frauen bei den Bewerbungen für Natur- und für Geisteswissenschaften gesagt!** Aus welchen Gründen auch immer - Frauen bevorzugen im statistischem Mittel die Geisteswissenschaften, Männer die Naturwissenschaften. Hier die Tabelle:

Zahl bzw. Quote	Frauen	Männer
Ablehnquote Naturwiss.	20%	30%
Ablehnquote Geisteswiss.	50%	60%
Bewerbungsanteil Naturwiss.	10%	80%
Bewerbungsanteil Geisteswiss.	90%	20%
Zahl der Bewerber NW	100	800
Zahl der Bewerber GW	900	200
Zahl der Ablehnungen NW	20	240
Zahl der Ablehnungen GW	450	120
Geamtzahl der Ablehnungen	470	360
Ablehnungsquote	47%	36%

Aus *denselben objektiven Zahlen* kann man also folgern

- (i) Die über alle Bewerber gemittelte Ablehnquote bei Frauen ist um 11% höher (letzte Zeile),
- (ii) Man kann nur Naturwissenschaften (NW) oder Geisteswissenschaften (GW) wählen. Was man auch wählt, stets ist die Ablehnquote bei den entsprechenden Fakultäten für Frauen um 10% geringer (ersten beiden Zeilen).

Die Auswahl einer heterogenen statistischen Masse (hier Frauen und Männer mit unterschiedlichen Eigenschaften, sprich Affinität zu Naturwissenschaften) ist die Ursache vieler statistischen Kontroversen und "verdammten Lügen"!

Das beginnt im Kleinen (ein Bäcker reduziert um 18h seine Preis für alle Backprodukte und dennoch steigt nach 18h der Durchschnittspreis pro verkauftem Stück), geht ins hochpolitische (das obige Beispiel kann man analog mit Behinderten, Schwarzen, Latinos etc. machen) und gipfelt in der Frage: "Geht es der Menschheit und der Umwelt immer besser oder immer schlechter?" (vgl. www.lomborg.org und www.anti-lomborg.com).