

Hinweise des Projektleiters zur Verwendung der Daten des ADFC-Fahrradklima-Tests 2014

Stand: 18.2.2015

Die Bewertungen der Teilnehmer erfolgen auf einer Skala von 1 bis 6. Dabei muss nicht jeder Teilnehmer gedanklich von gleichen Abständen zwischen den Bewertungsstufen ausgehen. Bei der Bildung eines Mittelwertes und der Interpretation als Schulnoten wird von dieser vereinfachenden Annahme ausgegangen.

Um ein Ranking der Städte zu ermöglichen und zu verdeutlichen, sind in den Ergebnistabellen die Werte für die Städte teilweise mit zwei Nachkommastellen Genauigkeit aufgeführt. Obwohl die ausgewiesenen Werte in der Regel die wahrscheinlichsten Bewertungen darstellen, besteht natürlich zwischen zwei Städten mit den Bewertungen von z. B. 3,6 und 3,8 kein relevanter Unterschied. Es muss bei den Städtebewertungen mindestens von einer Schwankung von $\pm 0,05$ Notenstufen ausgegangen werden, in Einzelfällen kann der Wert auch mehr als $\pm 0,2$ Notenstufen schwanken (Verdeutlichung des Sachverhaltes siehe Anhang).

Der in den Tabellen aus Praktikabilitätsgründen ausgewiesene Durchschnittswert für die Bewertungen der Teilnehmer sollte nicht der alleinige Maßstab zur Beurteilung sein. So ergibt beispielsweise der Durchschnitt von 10x Bewertung 2 und 10x Bewertung mit 5 den gleichen Wert wie der Durchschnitt von 10x Bewertung 3 und 10x Bewertung mit 4. Im ersten Fall muss das Ergebnis aber so interpretiert werden, dass es hier zwei Gruppen mit sehr unterschiedlichen Sichten auf die Radverkehrsbedingungen gibt und der Durchschnitt von 3,5 eher ein Zufallsprodukt ist – je nachdem wie stark die Gruppen vertreten sind. Im zweiten Fall kann von gleichartigen Einschätzungen im Bereich zwischen 3 und 4 ausgegangen werden. Deshalb ist in den Einzelstadtübersichten (Städte-PDFs auf der Karte) die genaue Verteilung der Antworten angegeben. Städte mit einer besonders großen Uneinheitlichkeit der Antworten sind in der Gesamttabelle speziell gekennzeichnet.

Die Bewertung der Fragen erfolgt zwischen den Polen von gegensätzlichen Aussagen (semantisches Differenzial). Die Formulierungen der Aussagen sind praxisnah und bildhaft¹, erschweren jedoch die fachlich saubere Auswertung, weil bei unterschiedlichen Fragen nicht immer dieselben Wörter zur Beschreibung des „Endpunktes“ und innerhalb einer Frage nicht immer exakte Gegensatzformulierungen verwendet werden. Beispiele dafür sind:

*Bei uns fahren **alle** Fahrrad – egal, ob alt oder jung ... ist das Fahrrad **eher** Fahrzeug für Kinder oder Sportgerät.*

*Bei uns werden Radwege **regelmäßig** gereinigt ... werden Radwege **selten** gereinigt.*

*Bei uns gibt es **keine** Hindernisse auf Radwegen und Radfahrstreifen ... gibt es **viele** Hindernisse auf Radwegen und Radfahrstreifen (z. B. Drängelgitter, Laternen, Werbeständer).*

*Bei uns kommt Fahrraddiebstahl **selten** vor ... werden Fahrräder **oft** gestohlen.*

*Bei uns werden Radfahrer/innen an Baustellen bequem und sicher vorbeigeführt ... werden Radfahrer/innen an Baustellen **meistens** zum Absteigen und Schieben gezwungen.*

¹ Auch aus Gründen der Vergleichbarkeit mit vorangegangenen Durchgängen des ADFC-Fahrradklima-Tests wurden die Fragen bisher weitestgehend unverändert gehalten.

Damit sind die absoluten Mittelwerte der einzelnen Fragen nicht exakt vergleichbar. Ein guter Ansatz ist hier der Vergleich mit anderen Städten – in der Regel mit dem Durchschnitt der Stadtgrößenklasse. Wenn man zur Bestimmung von Stärken und Schwächen für jede Frage den Abstand der Benotung einer Stadt zum Mittelwert der Stadtgrößenklasse betrachtet, gleichen sich Formulierungseffekte aus, weil die Formulierungen in allen Städten gleich waren.

Überhaupt besteht die große Stärke des ADFC-Fahrradklima-Tests darin, dass die Teilnehmer in einer großen Zahl von Städten zur gleichen Zeit mit der gleichen Methodik befragt werden. Die dadurch mögliche Vergleichbarkeit sollte ausgiebig genutzt werden.

Die Bewertungen eines Einzelteilnehmers ebenso wie die Summe dieser Bewertungen für eine Stadt scheinen von einer bestimmten Grundstimmung getragen zu sein. Die Bewertungen sind hoch korreliert, d. h. wenn die erste Frage (Spaß oder Stress) gut beantwortet wird erhalten auch die anderen Fragen ein „gutes“ Antwortmuster. Stärken und Schwächen einer Stadt spiegeln sich erst in den Nuancen des allgemeinen Bewertungsniveaus wider. Im Ergebnis dieses Effektes können sich beispielsweise Städte mit gleichen Beförderungsbedingungen für die Fahrradmitnahme im Öffentlichen Verkehr in der Bewertung der entsprechenden Fragen deutlich unterscheiden, je nachdem ob sie ein allgemein gutes Fahrradklima aufweisen oder nicht. Hier kann wiederum der Vergleich mit anderen Städten bei der Analyse helfen, indem z.B. der Abstand der Benotung zur Gesamtbewertung aller Fragen der Stadt betrachtet wird.

Während in den „großen“ Stadtgrößenklassen (>200.000 EW, 100.000-200.000 EW) alle deutschen Städte vertreten sind und in der Klasse 50.000-100.000 nur sieben von 107 Städten fehlen, dürfen in der „kleinsten“ Stadtgrößenklasse (<50.000 EW) Städte mit relativ schlechten Noten nicht automatisch als „Schlusslichter“ betrachtet werden. In der Gruppe der kleinsten Städte sind zwangsläufig nur die fahrradaktiveren Städte dabei. Allein schon die Tatsache, dass die Mindestteilnehmerzahl überschritten wurde, zeigt im Vergleich zu den wahrscheinlich wirklichen Schlusslichtern ein Grundinteresse an Radverkehr auf.

Allgemeingültige Mittelwerte einer Stadtgrößenklasse lassen sich deshalb auch nur für die Städte größer 50.000 Einwohner angeben und untereinander vergleichen. Dennoch lässt sich z. B. aus den „Top Ten“ der Stadtgrößenklassen ableiten, dass gerade in den kleinen Städten ein besonders hohes Potenzial für Fahrradfreundlichkeit vorhanden ist.

Es wäre schön, wenn auch das Ergebnis der Anstrengungen zur Fahrradförderung auf Kreis- und Landesebene analysiert werden könnten. Dem sind allerdings enge Grenzen gesetzt, weil regional sehr unterschiedliche Bedingungen herrschen, z. B. hinsichtlich Topographie oder der Zusammensetzung aus unterschiedlich großen Städten mit mehr oder weniger fahrradfreundlichen Stadtstrukturen. Zudem wird der Durchschnitt von Regionen dadurch beeinflusst, ob nur die typischen Fahrradstädte in die Wertung gekommen sind oder ob lokal auch eine gute Teilnehmerwerbung in anderen Städten erfolgte. Tendenzen lassen sich ableiten, indem man die Bewertungen der Städte in der Region, für die Werte aus mehreren Fahrradklimatest-Durchgängen vorliegen, im Zeitverlauf betrachtet.

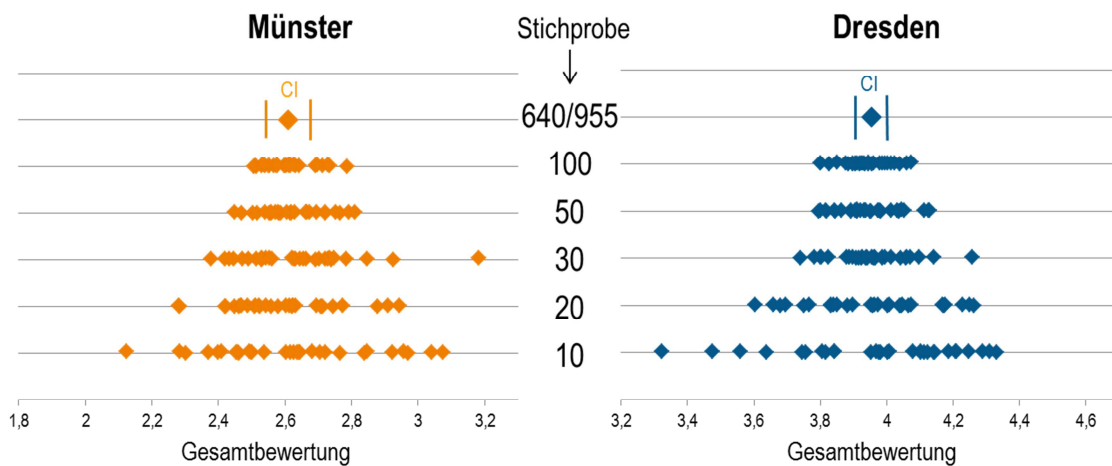
Während die ausgewiesenen Werte der Fahrradklimatest-Durchgänge bis zum Jahr 2003 in einigen Fällen nur den Durchschnitt aus wenigen Einzelbewertungen darstellen, können die Werte ab dem Jahr 2005 als relativ verlässlich betrachtet werden.

Anhang: Betrachtungen zur statistischen Sicherheit (Zufallsfehler)

Zur Betrachtung des Zufallsfehlers aus dem zufälligen Eintreffen von mehr oder weniger guten Bewertungen sollen die Interviews der Städte Münster und Dresden aus dem Jahr 2012 verwendet werden.

Frage: Wie wäre die Gesamtbewertung ausgefallen, wenn aus Dresden (955 Teilnehmer, $\bar{x} = 3,96$) und Münster (640 Teilnehmer, $\bar{x} = 2,61$) nur eine zufällige Stichprobe von 100/50/30/20/10 Teilnehmern eingegangen wäre?

Die Abbildung zeigt die sich dann ergebende Gesamtbewertung. Für jede Stichprobengröße wurden 30 „Ziehungen“ durchgeführt und die sich ergebende Gesamtbewertung als orangene bzw. blaue Rauten eingetragen. In Abhängigkeit von der Stichprobengröße schwanken die Werte mehr oder weniger stark, bei 50 Teilnehmern ca. $\pm 0,2$ Notenstufen. Im oberen Teil ist mit senkrechten Strichen der tatsächliche Fehlerbereich (Konfidenzintervall) der maximalen Stichprobe von 640 bzw. 995 Interviews angegeben². Er liegt etwa im Bereich von $\pm 0,05$ Notenstufen.



² 95%ige Wahrscheinlichkeit, numerisch ermittelt (Bootstrapping, 1000 Stichproben)