

Working Paper I/2018
Die Person Sum Technique
Ein neues Instrument zur Erhebung quantitativer
heikler Items

Justus Junkermann
Johannes Gutenberg Universität Mainz
Jakob-Welder-Weg 12
55128 Mainz
j.junkermann@uni-mainz.de

7. Februar 2018

Zusammenfassung

Die Person Sum Technique ist ein neues Befragungsinstrument zur Erhebung sensibler, quantitativer Fragen. Sie funktioniert analog zu anderen Instrumenten, wie der Person Count Technique und der Item Sum Technique. Die Befragten werden zufällig in zwei Gruppen aufgeteilt - die Short und die Long List. In der Short List wird ein quantitativer Wert von zwei Bekannten des Befragten erhoben. Der Befragte gibt als Antwort die addierten Werte. In der Long List werden die Werte von zwei Bekannten und des Befragten selbst auf die gleiche Art erhoben. Somit kann aus der Antwort nicht auf den wahren Wert des Befragten geschlossen werden. Durch die Differenz des Durchschnitts der Long und Short List erhält man einen Schätzer für den Durchschnitt der heiklen Frage.

Um dieses neue Instrument zu testen wurde ein Sample ($N = 215$) aus Studenten der Uni Mainz und der Uni Leipzig genutzt. Die Ergebnisse zeigen, dass die PST keine höheren Schätzer liefert als direktes Fragen.

Keywords: *Person Sum Technique, heikle Fragen, ICT, Sensitive Questions*

Inhaltsverzeichnis

1	Die „Person Sum Technique“ (PST)	1
2	Test der „Person Sum Technique“ (PST)	2
3	Diskussion	5

1 Die „Person Sum Technique“ (PST)

Bei der PST handelt es sich um ein neues Instrument zur Erhebung von heiklen Fragen mit quantitativen Antworten. Die Entwicklung erfolgte aufbauend auf bereits existierenden Befragungstechniken für heikle Fragen wie der Item Count Technique (ICT, auch: Unmatched Count Technique oder List Experiment) (Kuklinski et al., 1997; Wolter and Laier, 2014; Tsuchiya and Hirai, 2010; Tsuchiya et al., 2007; Chaudhuri and Christofides, 2007; Comşa and Postelnicu, 2013; Gilens et al., 1998; Hubbard et al., 1989; Kane et al., 2004; Rayburn et al., 2003; Rosenfeld et al., 2015; Walsh and Braithwaite, 2008; Holbrook and Krosnick, 2010; LaBrie and Earleywine, 2000), der Item Sum Technique (IST) (Wolter (2016); Trappmann et al. (2014)) und der Person Count Technique (PCT) (Grant et al. (2012)). Es handelt sich im Speziellen um eine Kombination von IST und PCT. Alle diese Fragetechniken basieren auf der gleichen Grundidee der ICT – Anonymisierung durch Randomisierung der Antworten. Bei der PST bestehen die Non-Key Items jedoch aus quantitativen Antworten über andere Personen.

Das Sample wird für die ICT in zwei Teile geteilt. Eine Gruppe bekommt eine Liste mit nicht heiklen Fragen – die Short List (SL). Der Befragte beantwortet diese Fragen nicht einzeln, sondern zählt die „JA“ Antworten zusammen. Die andere Gruppe bekommt die selbe Liste mit nicht heiklen Fragen plus eine heikle Frage – die Long List (LL). Auch hier werden die „JA“ Antworten zusammengezählt. Zieht man nun den Durchschnitt der SL von dem Durchschnitt der LL ab, erhält man einen Schätzer für die Prävalenz des heiklen Items. Diese Befragungstechnik ermöglicht dem Befragten 100% Anonymität, da nicht einmal der Interviewer auf die wahre Antwort der heiklen Frage schließen kann.

Die PCT basiert auf der gleichen Idee. Nur werden hier nicht mehrere Items abgefragt, sondern ein heikles Item über Bekannte des Befragten. Die SL besteht aus zwei (oder mehr) Bekannten des Befragten. Der Befragte zählt wieder zusammen auf wie viele seiner Bekannten die Antwort „JA“ zutrifft. Diesmal bezogen auf seine zwei Bekannten. Die LL besteht aus zwei (oder mehr) Bekannten des Befragten und dem Befragten selbst. Der Befragte zählt wieder zusammen auf wie viele seiner Bekannten und ihm selbst die Antwort „JA“ zutrifft. Aus der Differenz kann wieder ein Schätzer gewonnen werden.

Die IST basiert auf dem gleichen Prinzip (Wolter (2016); Trappmann et al. (2014)). Es werden in der SL ein (oder mehr) quantitative, nicht heikle Fragen gestellt. In der LL wird zu dieser nicht heiklen Frage eine quantitative heikle Frage gestellt. Der

Befragte rechnet beide Antworten zusammen und gibt das Ergebnis der Addition als Antwort an den Interviewer. Aus der Differenz kann wieder ein Schätzer gewonnen werden.

Die Person Sum Technique verbindet nun PCT und IST. Die SL besteht wie bei der PCT aus zwei (oder mehr) Bekannten des Befragten. Es wird in der SL eine quantitative, heikle Frage gestellt. Der Befragte rechnet beide Antworten zusammen und gibt das Ergebnis der Addition als Antwort an den Interviewer. Die LL besteht aus zwei (oder mehr) Bekannten des Befragten und dem Befragten selbst. Auch hier wird eine quantitative, heikle Frage gestellt. Der Befragte rechnet beide Antworten zusammen und gibt das Ergebnis der Addition als Antwort an den Interviewer. Der Schätzer wird wieder über die Differenz von den Durchschnitts von LL und SL gebildet.

Ein Beispiel:

Short List:

Denken Sie dazu bitte an zwei Bekannte von Ihnen, die möglichst unterschiedlich¹ sind. Die folgenden Fragen beziehen sich auf **diese zwei Bekannten**. An wie vielen Tagen machen Ihre zwei Bekannten pro Woche Sport? Bitte addieren Sie die Antworten und tragen die Summe als Antwort ein!

Long List:

Denken Sie dazu bitte an zwei Bekannte von Ihnen, die möglichst unterschiedlich sind. Die folgenden Fragen beziehen sich auf diese **zwei Bekannten und Sie selbst**. An wie vielen Tagen machen Sie und Ihre zwei Bekannten pro Woche Sport? Bitte addieren Sie die Antworten und tragen die Summe als Antwort ein!

2 Test der „Person Sum Technique“ (PST)

Da es sich um ein völlig neues Instrument handelt sollte in einem kleinen Test in Erfahrung gebracht werden ob,

- a) Die Befragten die Fragestellung kognitiv verstehen
- b) Die Befragten diese Technik als anonym, als direktes Fragen einschätzen
- c) Die PST höhere Schätzer von heiklen Items erzielt, als direktes Fragen

¹Die Bekannten sollten möglichst unterschiedlich sein, um Floor und Ceiling Effekte zu vermeiden. Sind alle Werte Null oder besonders hoch, kann auf die wahre Antwort des Befragten geschlossen werden.

Für diesen ersten kleinen Test der PST wurde eine Befragung (CASI) von Studenten der Soziologie der Universität Mainz und Leipzig durchgeführt mit insgesamt 215 Teilnehmern durchgeführt. Die Befragung erfolgte in Statistik-Software Kursen. Die Studenten sitzen dabei an PCs. Die Studenten wurden aufgefordert einen Link, der auf den Fragebogen führte, in den Browser einzugeben und selbigen dann zu beantworten.

Der Fragebogen enthielt ein experimentelles Design. Die Studenten wurden zufällig auf 3 Gruppen zugeteilt. Diese waren „Direct Questioning“ (DQ), die Short List (SL) und die Long List (LL). In den LL und SL Gruppen wurden die Anweisungen sehr detailliert mit einem Beispiel erläutert.

Um a) zu untersuchen wurde mit einer 5er Likert-Skala gefragt, ob die Fragestellung verstanden wurde. Die Befragten haben die Fragestellung kognitiv verstanden, wie in Tabelle 1 gut ersichtlich ist. Das Verständniss der LL und SL Fragen unterscheidet sich nicht signifikant vom Verständnis der direkten Fragen.

Tabelle 1: Verständnis der Fragestellung

Anweisungen verstanden	DQ	SL	LL	Total
1 trifft gar nicht zu	1 (1.23%)	0(0%)	0(0%)	1(0.48%)
...2	0(0%)	1(1.54%)	0(0%)	1(0.48%)
...3	2(2.47%)	3(4.62%)	1(1.61%)	6(2.88%)
...4	9(11.11%)	13(20%)	13(20.97%)	35(16.83%)
5 trifft voll zu	69(85.19%)	48(73.85%)	48(77.42%)	165(79.33%)
Total	81(100%)	65(100%)	62(100%)	208(100%)
$\chi^2(8) = 8.0573$				$p = 0.428$

Für Punkt b) wurde ebenfalls mit einer Likert-Skala erhoben, ob die Befragten das Gefühl hatten, dass die Anonymität ihrer Angaben gewährleistet ist. Die Ergebnisse in Tabelle 2 zeigen jedoch, dass die Befragten sich nicht signifikant anonym fühlten.

Tabelle 2: Anonymitäts empfinden

Anonymität gewährleistet	DQ	SL	LL	Total
1 trifft gar nicht zu	3(3.75%)	3(4.62%)	3(4.84%)	9(4.35%)
...2	4(5%)	2(3.08%)	5(8.06%)	11(5.31%)
...3	8(10%)	4(6.15%)	6(9.68%)	18(8.7%)
...4	30(37.5%)	17(26.15%)	18(29.03%)	65(31.4%)
5 trifft voll zu	35(43.75%)	39(60%)	30(48.39%)	104(50.24%)
Total	80(100%)	65(100%)	62(100%)	207(100%)
$\chi^2(8) = 5.9025$				$p = 0.658$

Jeder Gruppe wurde in dem entsprechendem Modus gefragt, wie oft sie Sport pro Woche machen, an wie vielen Tagen pro Woche sie Alkohol trinken, wie viele Zigaretten sie pro Tag rauchen, wie viel Geld sie pro Monat für Alkohol ausgeben und schließlich wie viel Geld sie pro Monat für illegale Drogen ausgeben. In Tabelle 3 sind die Ergebnisse der direkten Fragen sowie die Differenz Schätzer von LL zu SL aufgeführt.

Tabelle 3: Ergebnisse

	$\overline{DQ}(SD)$	Schätzer $\hat{\mu}$	$\overline{LL}(SD)$	$\overline{SL}(SD)$
N	82		67	66
Sport pro Wiche	1.51 (1.25)	1.90	5.34 (2.37)	3.44 (2.23)
Alkohol pro Woche	1.89 (1.34)	0.47	4.19 (3.72)	3.72 (2.90)
Zigaretten pro Tag	1.85 (3.45)	-2.05	5.24 (11.15)	7.29 (9.41)
€/Monat Alkohol	22.84 (20.04)	3.58	51.51 (47.44)	47.93 (46.12)
€/Monat Drogen	7.53 (19.16)	-9.74	22.50 (59.77)	32.24 (85.01)

Es ist gut zu erkennen, dass die SL oft höhere Werte als die LL hat und die Technik somit negative Schätzer produziert. In keinem Fall ist der Schätzer größer als der Durchschnitt der direkten Fragen. Die PST hat die Erwartung c) also eindeutig nicht erfüllt.

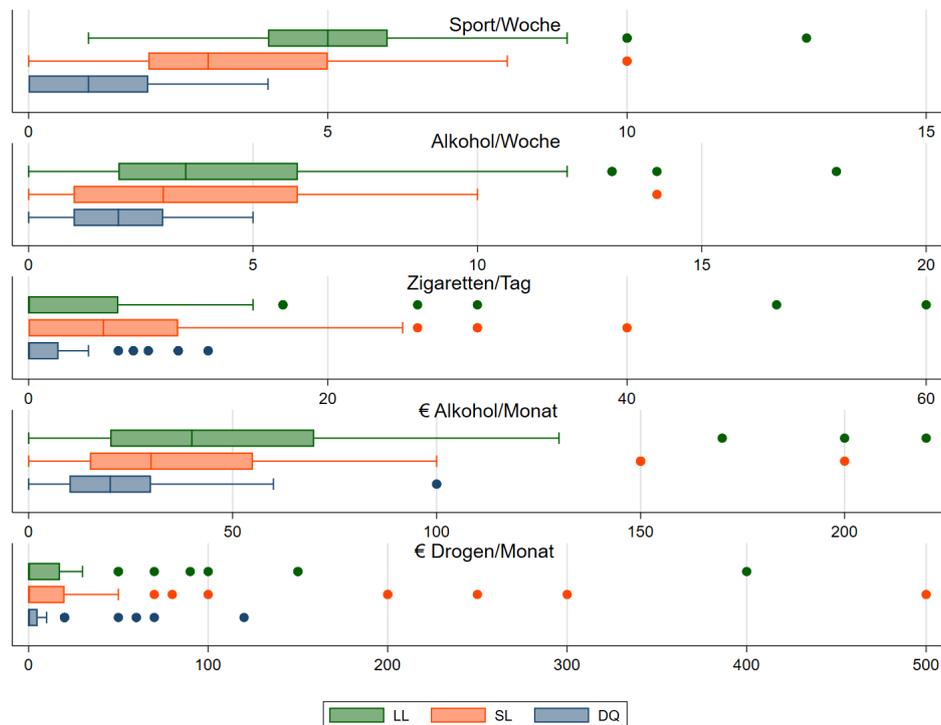


Abbildung 1: Boxplots der heiklen Items für Long List, Short List und Direct Questioning

3 Diskussion

Das Scheitern der PST an einer verbesserten Schätzung sensitiver Items kann mehrere Gründe haben. Die Standardfehler sind deutlich zu groß, teilweise sogar mehr als doppelt so groß, wie die eigentlichen durchschnittlichen Werte. Die Items zu €/Monat Drogen und Zigaretten sind ‘zero inflated’, da die Prävalenz dieses Verhaltens allgemein gering ist, wodurch die Schätzer verzerrt werden. Zusätzlich sind in den jeweiligen Gruppen sowie insgesamt zu wenig Befragte vorhanden. Für vergleichbare Techniken (PCT, ICT, IST) werden normalerweise mindestens Stichproben von mindestens ~500 Befragten gezogen. Ausreißerwerte haben hier einen großen Einfluss auf die Mittelwerte. Die Probleme mit den Verteilungen der Werte werden in den Boxplots in Graph 1 besonders ersichtlich.

Eine andere Erklärung wäre ein Einfluss der Selbsteinschätzung auf die Fremdeinschätzung. In der Short List werden ausschließlich Werte von anderen Personen durch den Befragten geschätzt. In der Long List hingegen schätzt der Befragte jedoch aus eigener Erfahrung und kann auf Erinnerungen zurückgreifen, die eine genauere Schätzung seines eigenen Wertes ermöglichen. Es wäre durchaus möglich, dass der Befragte die Werte seiner Bekannten an seinem eigenen Wert orientiert. Dies würde eine systematische Verzerrung zwischen Long List und Short List bedeuten und die Technik damit unbrauchbar machen. Dies hätte auch weitere Implikationen auf die Validität der Person Count Technique. Die Gründe für das Scheitern der PST in diesem ersten Test können jedoch nicht eindeutig bestimmt werden.

Als Empfehlung für weitere Tests kann jedoch festgehalten werden, dass bei der Auswahl von Items darauf zu achten ist, dass die Standardfehler möglichst gering sind, die Verteilung nicht ‘zero inflated’ ist, die Verteilung nicht zu viele Ausreißerwerte nach oben aufweisen sollte und die Antworten den Befragten kognitiv zugänglich sein sollten. Die Schätzer, welche sich in dieser Erhebung noch am besten, wenn auch nicht gut, bewährt haben, waren solche, die sich auf eine Verhaltenshäufigkeit pro Woche bezogen haben. Sie haben ein natürliches Minimum und Maximum, wodurch Ausreißer verhindert werden und die Standardfehler nicht zu groß werden können. Gleichzeitig sind den Befragten die Antworten oft kognitiv zugänglicher, als bei längeren Zeiträumen oder allgemeineren Fragen.

Literatur

Chaudhuri, A. and T. C. Christofides

2007. Item count technique in estimating the proportion of people with a sensitive feature. *Journal of Statistical Planning and Inference*, 137(2):589–593.

Comşa, M. and C. Postelnicu

2013. Measuring social desirability effects on self-reported turnout using the item count technique. *International Journal of Public Opinion Research*, 25(2):153–172.

Gilens, M., P. M. Sniderman, and J. H. Kuklinski

1998. Affirmative action and the politics of realignment. *British Journal of Political Science*, 28(1):159–183.

Grant, T., R. Moon, and S. Gleason

2012. *Asking Many, Many Sensitive Questions: "Person-Count" Method for Social Desirability Bias*. Conference Presentation: http://www.mapor.org/confdocs/absandpaps/2012/2012_slides/2C4_Grant_slides.pdf.

Holbrook, A. L. and J. A. Krosnick

2010. Social desirability bias in voter turnout reports: Tests using the item count technique. *Public Opinion Quarterly*, 74(1):37–67.

Hubbard, L., R. A. Casper, and J. T. Lessler

1989. Respondent reactions to item count lists and randomized response. *Proceedings of the American Statistical Association (Section on Survey Research Methods)*, Pp. 544–548.

Kane, J. G., S. C. Craig, and K. D. Wald

2004. Religion and presidential politics in florida: A list experiment. *Social Science Quarterly*, 85(2):281–293.

Kuklinski, J. H., P. M. Sniderman, K. Knight, T. Piazza, P. E. Tetlock, G. R. Lawrence, and B. Mellers

1997. Racial prejudice and attitudes toward affirmative action. *American Journal of Political Science*, 41(2):402–419.

LaBrie, J. W. and M. Earleywine

2000. Sexual risk behaviors and alcohol: Higher base rates revealed using the unmatched-count technique. *Journal of Sex Research*, 37(4):321–326.

- Rayburn, N. R., M. Earleywine, and G. C. Davison
2003. Base rates of hate crime victimization among college students. *Journal of Interpersonal Violence*, 18(10):1209–1221.
- Rosenfeld, B., K. Imai, and J. N. Shapiro
2015. An empirical validation study of popular survey methodologies for sensitive questions. *American Journal of Political Science*, 60(3):783–802.
- Trappmann, M., I. Krumpal, A. Kirchner, and B. Jann
2014. Item sum – a new technique for asking quantitative sensitive questions. *Journal of Survey Statistics and Methodology*, 2(1):58–77.
- Tsuchiya, T. and Y. Hirai
2010. Elaborate item count questioning: Why do people underreport in item count responses? *Survey Research Methods*, 4(3):139–149.
- Tsuchiya, T., Y. Hirai, and S. Ono
2007. A study of the properties of the item count technique. *Public Opinion Quarterly*, 71(2):253–272.
- Walsh, J. A. and J. Braithwaite
2008. Self-reported alcohol consumption and sexual behavior in males and females: Using the unmatched-count technique to examine reporting practices of socially sensitive subjects in a sample of university students. *Journal of Alcohol and Drug Education*, 52(2):49–72.
- Wolter, F.
2016. *Using the Item Sum Technique to Avoid Nonresponse to the Income Question and Underreporting of Self-Reported Alcohol Consumption*. Working Paper: University of Mainz.
- Wolter, F. and B. Laier
2014. The effectiveness of the item count technique in eliciting valid answers to sensitive questions. an evaluation in the context of self-reported delinquency. *Survey Research Methods*, 8(3):153–168.