Speech Prosody 2022
23-26 May 2022, Lisbon, Portugal

Fundamental Frequency Variation in Polarity Questions of Czech

Jan Volín, Michaela Svatošová, Pavel Šturm

Institute of Phonetics, Faculty of Arts, Charles University
Czech Republic
jan.volin@ff.cuni.cz

Abstract

Descriptions of intonation patterns of questions in Czech have so far been impression-based or fragmentary. Our study explores the phonological typology of polarity question melodies proposed in literature and provides quantitative data for the modelling of individual patterns. Although polarity questions are usually found to be more frequent than other interrogative types, their immense variation in form and function still needs to be captured. We use material from a corpus of acted minidialogues produced by 34 Czech speakers. K-means clustering together with auditory analysis and visual inspection of F0 tracks offered an insight into both categorial and continuous variation. Emerging patterns have been specified in terms of their F0 parameters and frequency of occurrence. Although the overall variation in the data is substantial, we suppose that most of it is not random. We expect our results to serve in perception test design, in which various components of meanings other than plain request for information could be tested.

Index Terms: polarity question, yes-no question, Czech, speech melody, intonation

1. Introduction

1.1. Research background

The grammatical approach to speech melodies never fails to differentiate between statements and questions. This dichotomy is useful in many other approaches to prosody description, with an already well-known caveat: not all interrogative structures actually serve to ask for information, and not all grammatical statements are intended to provide information (e.g., [1], [2], [3]). Despite a vast array of communicative functions, however, the actual forms and contexts usually guide the listener sufficiently enough, so that, for instance, a reproach, an expression of disbelief, an invitation to comply or a request for help can be differentiated form a true functional question.

Traditionally, several types of questions are recognized. It seems that *polarity* questions (also *polar*, *yes-no* – we use the term after [4]) are more frequent in speech corpora than various other types [5]. Our unpublished probe into a corpus of TV debates revealed that about 12 hours of spoken texts contained 655 questions produced by 36 speakers, of which more than a half were polarity questions (342 cases, 52.2%), while *question-word* questions made only about one third of the count (205 cases, 31.3%). There were also *alternative* questions (48 cases, 7.3%) and syntactically complex *combined* forms (60 cases, 9.2%). These ratios are similar to the ones found in [3] for Swedish, and reasonably corresponding with [5] for American English, given that the author had a larger corpus and for her study she worked only with a pre-selected sample of items.

Whereas the question-word questions in Czech have been investigated quite thoroughly due to the historical controversies in their descriptions (e.g., [6], [7], [8]), the polarity questions did not attract too much attention. Their descriptions were either impressionistic (e.g., [9], [10]) or based on scarce material (e.g., [11], [12]). Nevertheless, even these early descriptions already indicate that the situation is not trivial, i.e., just a plain rise for polarity questions. Authors in [9], [10] and [11] already speak about two fundamental contours: **rising** vs. **rising-level** (or rise with moderate fall for the latter). Another contour was added by [12]. It is a **high plateau** and it is considered colloquial, and typical of central regions of the Czech lands. A fourth type is sometimes informally discussed but to our best knowledge without a clear published description. It is the **delayed rise** of certain Moravian dialects of Czech. In terms of ToBI convention, but with M for a mid-range target, the four types would be (i) L* H- H%, (ii) L* H- M%, (iii) L+H* M-M%, and (iv) L* L- H%.

It could be noted that the differentiation between various forms of the interrogative and non-conclusive melodies in Czech is not a simple matter either [13]. This might be the consequence of the relatively rich repertoire of contours used in a fairly small geographical area of the Czech Republic. Furthermore, there seem to be disparate preferences for individual patterns in various regions [14]. Although both L* H- H%, and L* H- M% are considered standard, they do not seem to be used interchangeably within the same community of speakers.

1.2. Research questions

Since Czech is one of the languages with no morphosyntactic differences between polarity questions and statements, speech melodies (and a question mark in writing) are the crucial cues for differentiation. The task of our present study is to provide empirical quantitative data concerning the issues mentioned above to allow for further hypothesis testing in the future. Our data will be based on a larger material with clearly described conditions of its collection and research methodology. Specifically, we will answer the questions stipulated below.

- Are the traditionally recognized patterns accounted for in our material? Is there anything extra or missing?
- What are the frequencies of occurrence of the obtained patterns?
- Do various methods of pattern-seeking produce mutually coherent results?
- What are the typical F0 values in nuclear rises of polarity questions (e.g., what is the typical span of the rise, where is the relative offset of the nuclear rise, etc.)?
- Are idiosyncratic features of individual speakers found consistently in the sample?

 10.21437/SpeechProsody.2022-173

2. Method

2.1. Material

The questions that are examined in the present study come from five-turn dialogues that were scripted to provide clear communicative contexts. The questions always formed the third turn in each dialogue. Speakers (34 philology volunteering students) were participating in pairs and they were recorded in two sessions on two different days. Members of a pair were familiar with each other, usually friends. They were given sufficient time to get acquainted with a dialogue and were asked to act out the situation as naturally as possible with the option of several trials. If either the speakers themselves or the experimenters found the performance unconvincing, the dialogue was re-recorded (cf. [15], [16], [17]). On another day, they swapped the roles and read the same dialogues differently ordered.

The examined material for the present study comes from six dialogues which were separated by at least eight other unrelated dialogues (fillers) during the recording session. This rule was taken to avoid stereotyped prosody for similar linguistic structures. Besides that, each of the sample questions was placed into a unique co-text, i.e., each situation in a single dialogue was different from all the others. Care was taken to produce co-texts that lead to functional information seeking interrogatives. The texts of the target questions together with their English translations are displayed in Table 1.

Table 1: *The target items to be investigated; English translation in italics is done word-by-word, hence outside standard grammar.*

Item	Wordings
1a	Řekneš jim, co si myslíš?
	Tell them what you think? (singular)
1b	Řeknete jim, co si myslíte?
	Tell them what you think? (plural)
2a	Zjistíš, v kolik to pojede?
	Find out at what time it goes? (singular)
2b	Zjistíte, v kolik to pojede?
	Find out at what time it goes? (plural)
3a	Víš, kdy to dostaneš?
	Know when it obtain? (singular)
3b	Víte, kdy to dostanete?
	Know when it obtain? (plural)

The recording took place in a sound-treated studio of the Institute of Phonetics in Prague, with an AKG C4500 B-BC microphone at 32 kHz sampling rate with 16-bit resolution. The speakers were sitting less than 1 meter apart, facing partly each other and partly a screen with the dialogue texts.

2.2. Contexts and structures

Individual target items from the corpus had various contexts and structural properties. These are displayed in Table 2. Our main concern is the number of syllables in the final accent group, i.e., the stretch over which the *nuclear pitch pattern* extends. It can be noted that we did not include one-syllable patterns – Czech is highly inflectional and monosyllabic words in the final position are not particularly common.

Numbers of pre-nuclear syllables also differed and so did the number of words. These facts are not particularly relevant to the research questions of the present study.

Table 2: *Elementary structural properties of the target items: number of words (Wds), and of syllables in the questions (Sylls-Total) and in the nuclei (Sylls-Nuclei)*

Item	n Wds	n Sylls-Total	n Sylls-Nuclei
1a	5	7	2
1b	5	9	3
2a	5	8	3
2b	5	9	3
3a	4	6	3
3b	4	8	4

2.3. Measurements

The F0 tracks were extracted with the autocorrelation method built in Praat [18]. All 204 tracks were subjected to auditory inspection and manually corrected for octave jumps, missed voicing regions, or spurious values. According to [19] the indicative F0 values for melody perception are found in the second third of the syllabic nucleus (in our case always a vowel). We found 5 equidistant point spanning across the second third of the syllabic nucleus and calculated arithmetic mean of the F0 values taken there. In order to approximate human hearing, the values in Hertz (Hz) were converted to semitones (ST). These were also normalized with speakers' mean F0 over several turns to allow for mutual comparison of individuals speaking in different frequency bands. Thus, the value of 0 ST always refers to the speaker's mean pitch (which ranged between 111.5 and 293.9 Hz).

2.4. Analytical procedures

Three procedures were used with the goals related to the research questions listed at the end of Section 1.2.

First, we performed a step-by-step or stepwise analysis based on differences between neighbouring syllables. As the melody based on F0 tracks can be considered falling, rising or level, a threshold separating these categories needs to be specified. Based on [19] we opted for 0.5 ST. All the steps lower than -0.5 ST are reported as falling, those that are higher than 0.5 ST are rising. The smaller steps in between these two values are reported as level in Section 3.1.

Second, an independent perceptual analysis by an expert phonetician (the first author) was carried out. In the past, visual and/or auditory inspection of F0 patterns used to be recognized as a valid method of enquiry even as a sole procedure. Current conventions of empirical research find it insufficient as a single source of results (see also our critical comments in Section 1.1). In certain cases, however, it could serve as a supplementary method. One of such cases is our current study, because our objectives are descriptive, and there are no overt hypotheses being tested. Therefore, the observer should not be biased in favour of any outcome. Independently of the previous analyses, the first author of the study listened to the whole set of 204 items together with visual inspection of an F0 track in Praat [18] and decided whether they fitted any of the described patterns or formed a pattern previously unaccounted for.

Third, K-means cluster analysis was carried out in Statistica Cz 12 by StatSoft [20]. Clustering 'by lines', i.e., of sequences of values was used. The entered lines of values were based on methodology described in Section 2.3. The initiation algorithm was MD = maximum distance. Two-, three- and four-cluster solutions were required consecutively. The resulting clusters were inspected visually and compared with the traditionally recognized patterns and outcomes of the first two analyses.

3. Results

3.1. Nuclear pitch patterns – melodic steps

The first round of analyses focused on melodic steps between neighbouring syllables. The steps were assessed as described above in Section 2.2. Table 3 summarizes the situation.

Table 3: *Number of rising, falling and level melodic steps between neighbouring syllables in interrogative melodemes. Rounded percentages in brackets.*

	Pre to TS	TS to Post1	Post1 to Post2	Post2 to Post3
↗	23 (11)	170 (83)	120 (71)	18 (53)
→	41 (20)	15 (7)	20 (12)	3 (9)
↘	140 (69)	19 (9)	30 (18)	13 (38)
All	204 (100)	204 (100)	170 (100)	34 (100)

The crucial movement is from the tonic syllable (TS = stressed syllable with which the nuclear pitch pattern starts) to the following one (Post1). Such a step was present in all 204 analysed items since there was always at least one syllable after the TS. In 83% of the cases, there was a rise, the rest was more or less equally split between levels and falls (see column *TS to Post1* in Tab. 3). The rising melodic steps prevailed even further on in longer syllabic chains. What should be noted is the movement from the pre-nuclear part (column *Pre to TS* in Table 3). In almost 70% of the cases there was a step down as if in anticipation of the need to create space for the following rise.

An important issue, however, is the overall configuration or pattern, rather than the individual steps. In 2-syllable nuclei, the configuration is quite simple: the rise prevailed with 88% of the cases. However, 3-syllable nuclei (n = 136) offer a possibility of change in the direction. Figure 1 displays two dominant and two marginal contours found in the material.

Contour A accounted for 57.4% of the cases, while contour B for 22.8% of the cases. Contours C and D accounted for 13.9% and 5.9%, respectively. This type of analysis, therefore, suggests that the four patterns reported in literature (Section 1.1) are sufficient to explain the variation in the material. There is one potentially controversial moment though: the moderately falling lines in Figure 1 signal collapsing levels and falls from Table 3. See Section 4 (Discussion) for justification.

The rises between the tonic and the first post-tonic syllables had a mean span of 3.1 ST, ranging from -4.4 ST to 12.8 ST. The tonic syllable (TS) usually lay below the speakers mean pitch, at –2.0 ST on average, regardless of the number of syllables in the nuclear contour.

3.2. Consistency of individual speakers

Out of 34 speakers, 12 produced the same 3-syllable pattern at all four occasions, and 14 produced one exceptional pattern with the rest always identical. Eight speakers were inconsistent.

3.3. Nuclear pitch patterns – perceptual inspection

As explained in Section 2.4, the first author listened to the set of 204 items deciding which of the described patterns they represented. A pattern previously unaccounted for was also an option. This supplementary analysis was unbiased in that it was carried out independently of the previous analyses, i.e., without the knowledge of the other results and without any hypotheses about the occurrence of potential patterns.

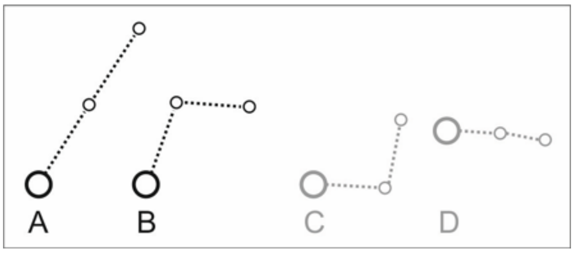

Figure 1: *Schematic diagram of four melodic patterns resulting from the step analysis (see Section 3.1). Grey colour is used for less frequent patterns (see text).*

Two-syllable nuclei (n = 34) offered a very similar outcome to the previous analysis: 29 cases were plain rises, and only 3 falls and 2 levels. The occurrence of rises is in line with traditional descriptions, whereas the falls and levels have to be perceptually tested in future for acceptability scores.

Three-syllable nuclei resulted in a more diversified picture, which we summarize in Table 4. Please note that R in Table 4 parallels with Contour A in Fig. 1. Likewise, RL ≈ Contour B, DR ≈ Contour C, and L ≈ Contour D. Four cases fell into the category Other (O). These were perceptually unusual – almost bizarre, and did not fit into previously established classes.

Plain gradual rises (R) were again most frequent, even if their prevalence was smaller in comparison with the outcome of the previous analysis. The rise-level contours came second in the frequency of occurrence, but were more common than suggested in Section 3.1. Delayed rises occurred in more or less the same ratio, while the level contours were again more common than established by the preceding analysis.

Table 4: *Occurrences of rising (R), rise-levelling (RL), level (L), delayed rising (DR) and other (O) melodic patterns in three-syllable interrogative melodemes.*

	R	RL	DR	L	O
n	53	41	13	25	4
%	39.0	30.1	9.6	18.4	2.9

3.4. Nuclear pitch patterns – K-means clustering

K-means clustering offered various solutions depending on the number of clusters required. The method is generally well accepted as an exploratory tool in larger data sets. However, certain limitations need to be mentioned, especially in intonology. The method is blind to human perceptual mechanism as it is based just on residual variance in the data. Therefore, the identified clusters are not necessarily perceptually relevant.

Be that as it may, two-syllable item clustering completely ignored a few exceptional cases and produced only rising patterns that differed either in the pitch range span or level. This result does not address any of the research questions for the present study, therefore, it will not be presented in detail.

Clustering of 3-syllable items, on the other hand, already contributed to our pattern mapping. The outcome is displayed in Figure 2. There are two rising contours (A and C), the rise-level (B) and the level contour (D). The difference established above between gradual and delayed rises is thus dissolved in the data, probably due to the low number of the latter. Interestingly, this difference was revealed in the clustering of four-syllable items (see Figure 3).

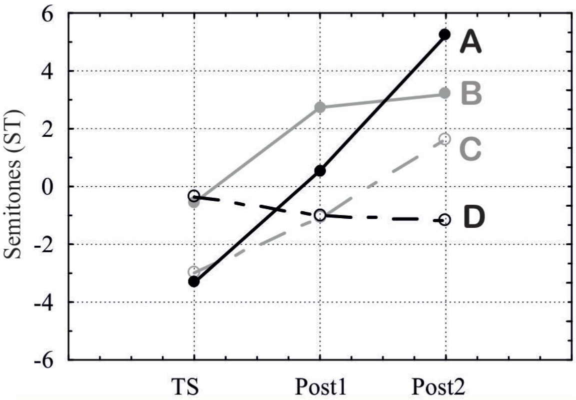

Figure 2: *Centroids of 3-syllable melodic patterns based on K-means method (see Section 3.3). Grey colour is used only for better visual differentiation.*

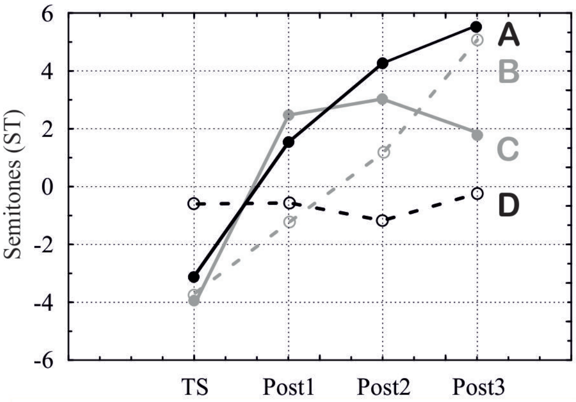

Figure 3: *Centroids of 4-syllable melodic patterns based on K-means method (see Section 3.3). Grey colour is used only for better visual differentiation.*

3.5. Cluster type evaluation

The differences between the cluster types in Figure 2 are visually apparent, but it would be useful to know whether this can be confirmed statistically, too. CLUSTER TYPE (3 levels: Rising, Rise-Level, Level) was therefore included as a fixed effect in an LME analysis predicting the span of the first or second F0 step in a three-syllable nuclear contour (SPEAKER and ITEM were treated as random effects). The original F0 values are used here, adding the category label from the k-means clustering as a predictor.

In a model that evaluated the span of the first step (subtracting Post1 from TS values), CLUSTER TYPE reached significance ($\chi^2(2) = 8.9$, $p = 0.012$). After Bonferroni correction, only the Level type differed significantly from the other types in paired comparisons (-3.06 ST, $p = 0.029$ for Level vs. Rise-Level comparison; -2.63 ST, $p = 0.011$ for Level vs. Rising comparison), while Rise-Level and Rising clusters were not separated (0.43 ST, $p = 1.0$). Another model evaluated the span of the second step (subtracting Post2 from Post1 values), and confirmed CLUSTER TYPE to be a significant predictor as well ($\chi^2(2) = 12.6$, $p = 0.002$). This time, pairwise comparisons showed no significant difference between Level and Rise-Level cluster types (1.06 ST, $p = 0.437$). However, the Rising cluster differed significantly from the Level and Rise-Level types (-2.48 ST, $p = 0.046$ and -3.53 ST, $p < 0.001$).

4. Discussion

All three methods of inquiry actually lead to mutually compatible results. The stepwise analysis in Section 3.1 accounted for the patterns reported in older literature, although the less frequent level and falling steps had to be collapsed into one category. This arbitrary act might seem opportunistic, but given the research questions in Section 1.2 together with the limited space for this study, it actually seems logical. Hopefully, it will inspire future perceptual testing to see whether the differences *in the given positions* between level and moderate fall are meaningful (even if in terms of affective or conative components of the meaning).

Interestingly, the stepwise analysis (Section 3.1) differed from the impressionistic analysis (Section 3.2) in the ratio of rises vs. rise-levels, i.e., the two most frequent contours in the material. One possible explanation is that mechanistic threshold of 0.5 ST for differentiation of rises from levels is in perceptual evaluation too 'local'. Perhaps, the human assessment of the melodic turns is influenced by the preceding melodic context and some of the rises are perceived as levels if the rise from the preceding syllable is much steeper. In other words, the turn from sharp to moderate rise in three-syllable items might be perceived as a functional rise-level, or L* H- M% pattern. This, again, deserves future attention of researchers.

Slightly over a third of the speakers seemed to be consistent in their choice of patterns, with another third nearly consistent (Section 3.2). This outcome might attract attention of forensic experts, and we believe such an analysis could be further refined if the contours were expressed as curves rather than just lines between syllable nuclei.

Finally, we can perhaps reach outside the scope of our study and report that when listening to our recordings, we found neither 'neutral' nor 'emotional' renderings. Instead, all our speakers displayed various affective stances, depending on how they evaluated and mentally reconstructed the context of the dialogues. It should be stressed that evaluation and context reconstruction is seldom conscious – the speakers usually just 'feel' that the state of affairs is such and such. It is most probably outside their capacity to review all their life experience and check which part of it is relevant for the situation at hand. These affective and conative messages are of utmost importance in our everyday lives. We hope that our research will help to prepare perception experiments that could elucidate these communicative cornerstones, whether in Czech or in other languages.

5. Conclusions

As to the research questions in Section 1.2, three melodic patterns reported in literature were confirmed and even the fourth one that escaped attention of past prominent authors was found in non-negligible counts. Three different methods of analysis were used, and their outcomes were found mutually corroborative. In addition, we were able to provide occurrence ratios of individual types based on a sample of clearly specified material of unprecedented extent. Moreover, the quantitative specification of contours, which was missing in Czech phonetics, was produced and can be further exploited.

6. Acknowledgements

This work was funded by the Czech Science Foundation (GAČR), standard project GA21-14758S, for all three authors.

7. References

[1] O. Müllerová, “Otázka a odpověď v dialogu”, *Slovo a slovesnost*, vol. 43, no. 3, pp. 200–212, 1982.

[2] M. Selting, “Prosody as an activity-type distinctive cue in conversation: the case of so-called 'astonished' questions in repair initiation”, in *Prosody in Conversation: Interactional Studies,* E. Couper-Kuhlen and M. Selting, Eds. Cambridge University Press, 1996, pp. 231–270.

[3] J. Edlund, D. House, and S. Strömbergsson, “Question types and some prosodic correlates in 600 questions in the Spontal database of Swedish dialogues”, in *Proc. of Speech Prosody 2012*, Shanghai, vol. 6, pp. 737–740, Tongji University Press, 2012.

[4] K. J. Kohler, *Communicative functions and linguistic forms in speech interaction*, Cambridge: Cambridge University Press, 2018.

[5] T. Stivers, “An overview of the question–response system in American English conversation”, *Journal of Pragmatics*, vol. 42, pp. 2772–2781, 2010.

[6] P. Šturm, and J. Volín, “Perceptual evaluation of early versus late F0 peaks in the intonation structure of Czech question-word questions,” in *Proc. of Interspeech 2019*, pp. 1976–1980, Graz, ISCA. Sept. 2019.

[7] J. Volín, and P. Šturm, “Fundamental frequency tracks of question-word questions in natural and synthetic speech,” *Akustické listy,* vol. 25, no 3-4, pp. 3–9, 2019.

[8] J. Volín, and P. Šturm, “Melodie české doplňovací otázky,” [Melody of the Czech question-word question], *Naše řeč*, vol. 103, no 4, pp. 337–351, 2020.

[9] S. Petřík, “*O hudební stránce středočeské věty*,” [The melodic aspect of the sentence in Central Bohemian Czech]. Praha, FF UK, 1938.

[10] F. Daneš, “*Intonace a věta ve spisovné češtině*,” [Intonation and sentence in standard Czech]. Praha, ČSAV, 1957.

[11] M. Romportl, “*K tónovému průběhu v mluvené češtině*,” [Towards tone contours in spoken Czech]. Praha, KCSN, 1951.

[12] M. Romportl, “Melodie (tónový průběh) otázky zjišťovací v hovorové střední češtině”, *Listy filologické/Folia Philologica*, vol. 75, no. 6, pp. 268–274, 1951.

[13] J. Janíková, “Phonologically relevant F0-patterns and their limits in yes-no question and progredient utterance in Czech,” In Z. Palková and H.-W. Wodarz (Eds.): *Papers in Phonetics and Speech Processing. Forum Phoneticum* 70, Hector Frankfurt am Main, pp. 51–67, 2000.

[14] J. Janíková, “Územní preference v percepci zjišťovací otázky,” in: Z. Hladká, and P. Karlík (Eds.) *Čeština - univerzália a specifika 4.* Praha: LN 2002. pp. 313–315, 2002.

[15] O. Niebuhr, and A. Michaud, “Speech Data Acquisition – The Underestimated Challenge.” *Kieler Arbeiten in Linguistik und Phonetik* 3, pp. 1–42, 2015.

[16] O. Niebuhr, B. Peters, R. Landgraf, and G. Schmidt, “The Kiel Corpora of Speech & Emotion - A Summary,” *Proc. of 41st Conf. of the German Acoust. Soc.*, Nuremberg, DEGA, pp. 1–4, 2015.

[17] K. J. Kohler, B. Peters, and M. Scheffers “*The Kiel Corpus of Spoken German – Read and Spontaneous Speech. New Edition, revised & enlarged. Documentation,*” Kiel: Christian-Albrechts-Univ., http://www.isfas.uni-kiel.de/de/linguistik/forschung/kiel-corpus, 2017.

[18] P. Boersma, and D. Weenink, “Praat: doing phonetics by computer,” [Computer program] version 6.0.47, retrieved Feb. 8, 2019, http://www.praat.org/, 2019.

[19] D. J. Hermes, “Stylization of pitch contours”, in S. Sudhoff et al. (Eds.) *Methods in Empirical Prosody Research*, Berlin, Walter de Gruyter, 2006, pp. 29–61.

[20] StatSoft, Inc. “Statistica Cz,” [Data analysis software system] version 12, www.statsoft.com, 2013.