2022 ■ Vol. 18 ■ no. 2

Teaching Quantitative Methods Vignettes

Focusing on structural features of a scenario: An activity on expected frequencies, deviations, and chi-square contributions

Véronique Chadillon-Farinacci [a] ✉

[a] université de Moncton

Abstract ■ In this paper, I present a way to teach statistics by helping students reuse structural features of a problem to interpret those features. It consists of explaining concepts by presenting corresponding analogical examples following a given story. Expected frequencies, deviations and chi-square contributions are presented in relation to a story where six people found tokens on a geocaching course. The activity prepares students to gradually interpret these three elements of the chi-square statistic test on their own. The activity also allows for a discussion about benefits of the chi-square test of independence, namely its added value compared to proportions that can show spurious results.

Keywords ■ storytelling, expected frequencies, deviations, chi-square contributions, chi-square statistic.

Acting Editor ■ Sébastien Béland (Université de Montréal)

Reviewers ■ Two anonymous reviewers.

✉ veronique.chadillon-farinacci@umoncton.ca

doi 10.20982/tqmp.18.2.v001

Concept to be presented

Several studies encourage the use of real-world examples and humour when teaching statistics (Bizon, 2018; Carver et al., 2016; Cousineau & Harding, 2017; Zeidner, 1991). This approach doesn't exclude integrating independent work on the part of the students. A recent study suggests that making students read worked examples might enhance their performance (Brisbin & Maranhao do Nascimento, 2019). Structured yet casual guidance seems to be a favourable teaching practice.

A given scenario where a problem has to be resolved can be deconstructed into two main components: structural features and surface features (Quilici & Mayer, 1996). Structural features include elements of a problem that allow a solution to be found. As for the surface features, they include elements related to the story itself. However, structural and surface features are integrated differently depending on the type of statistical analysis presented. For example, in front of various scenarios, undergraduates are better at identifying situations resolved by correlation analysis than chi-square independence tests (Quilici & Mayer, 2002).

The proposed activity help students direct their attention to the structural elements of a problem using a parallel scenario. These structural elements are then used to interpret results.

The chi-square test of independence is interesting in this respect: in addition to being considered by some teachers as "essential for good statisticians" (Harth, 2018), it may be one of the first statistical tests that students are introduced to in their academic journey.

We can make it more intuitive for two interrelated reasons: its familiarity and its simple calculation. First, chi-square test of independence is based on frequencies and count data are found in a variety of everyday situations: the number of items purchased in a store, the number of grammar errors in a text, the number of crimes committed by teenagers, etc. Second, the chi-square test of independence is relatively easy to calculate in a spreadsheet or by hand (McHugh, 2013). Thus, teaching this test through a scenario can introduce expected frequencies, deviations, and chi-square contributions in a more familiar way.

The chi-square test of independence tests the relationship between two qualitative variables with nominal scales. Because the mean of categorical data is meaningless, the test compares observations with predictions. In other words, it compares observed frequencies with expected frequencies (Field, Miles, & Field, 2012). This comparison allows contrasting two scenarios and to answer, for example, the following question: which modality of which group has higher or lower frequencies than expected?

The greater the difference between observed and expected frequencies, the higher the chi-square statistic, the less likely the difference is due to chance. Observed frequencies come from the cross-frequency table directly drawn from the data. The expected frequencies are calculated as follows for each cell:

$$E_f = \frac{r_t \times c_t}{T_t}$$

where E_f is th expected frequency, r_t is the row total, c_t is the column total, and T_t is the table total.

In a way, the expected frequencies table is the scenario where all cells contribute to the cross-frequency table according to their weight per group and per category (or modality) in the sample. The difference between observed and expected frequencies are called deviations which are calculated as follows for each cell excluding the totals:

$$D = O_f - E_f$$

where D is the deviation, O_f is the observed frequency, and E_f is the expected frequency.

The larger the deviations, the less likely the hypothesis of independence between the two qualitative variables (Gilles & Maranda, 1994). The chi-square contributions indicate which cells contribute most to the result.

$$C = \frac{D^2}{E_f}$$

where C are the contribution, D is the deviation, and E_f is the expected frequency.

After calculating all the chi-square contributions and summing them up, we refer to a threshold. If the sum of the chi-square contributions is greater than the threshold, the deviation between observed and expected frequencies is statistically significant. Therefore, the probability that the two variables are not independent of each other is low. This threshold as well as the degrees of freedom are not the subject of the proposed activity for the sake of parsimony. The proposed activity focuses on expected frequencies, deviations, and contributions to the chi-square statistic.

Activity

Typically, metaphors present concepts so that a source models a target. For example, correspondence metaphor involves relating entities or attributes with their counterparts in the target. This strategy can be used to teach statistics (Tay, 2022). We propose the use of a correspondence metaphor to highlight the structural features of a scenario by equating these features in the story.

Therefore, the proposed activity follows Cousineau and Harding's (2017) recommendation: we present calculations as a step in a story rather than as an intimidating equation. Accordingly, the familiar scenario is presented first. In a second step, we present a survey situation.

Firstly, six people must find tokens on a geocaching course. First common names are suggested: Alice, Bob, Chuck, Dan, Eve, and Franck. We are interested in the number of tokens found by everyone.

These six people have different characteristics. Some are amateur athletes: it is the case for Alice, Chuck, and Eve. Others are professional athletes: Bob, Dan, and Franck compete professionally.

In addition, these people used different means of transport to find their tokens: Alice and Bob used their car that day, Chuck and Dan cycled, and Eve and Franck walked on the course. The people and their performance are illustrated in Figure 1.

The setting of the survey context is presented next: 337 sports amateurs and athletes are asked what their usual mode of transportation is. In the cross-frequency table found in Figure 1 to the right. The columns represent the two groups, i.e., sports enthusiasts and athletes. The rows represent three means of transport: car, bike and walk.

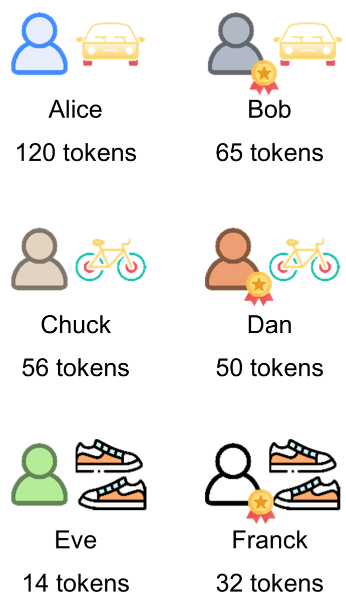
In both scenarios, the idea is to compare an actual performance to an expected performance. The structural elements of the story include the columns and rows of the table, i.e., the two types of people and their means of transport to carry out the course, but also the "performance" whether expressed as found tokens or people answering a survey. The six people correspond to the number of cells of the cross-frequency table (excluding the marginal totals).

Whether in the geocaching course scenario or the survey and cross-frequency table, to make a comparison, we ask how the groups (non-athletes and athletes) relate to the modalities (means of transport). In other words, did Alice, considering that she is not a professional athlete and that she used her car on the course, performed as expected? And what is meant by "performed as expected"?

If the tokens were distributed according to table columns and rows, we would draw two observations:

1. The amateur athletes would have 56.38% of the token of their column, because they have 190 tokens

Figure 1 ■ Six people, amateur and professional athletes, found tokens on a geocaching course using different means of transportation Left: pictorial representation; right: Cross-frequency table of the usual means of transportation of amateur sports enthusiasts and athletes ($n = 337$). The images were designed using the resources of *Flaticon.com*.



	Amateurs	Athletes	Total
Car	120	65	185
Bike	56	50	106
Walk	14	32	46
Total	**190**	**147**	**337**

out of the total of 337 tokens found on the geocaching course ($190/337 \times 100\% = 56.38\%$). The professionals would have the balance of tokens for their respective column. Indeed, they found a total of 147 tokens ($147/337 \times 100\% = 43.62\%$).

2. The tokens would be distributed according to the mode of transportation used in the geocaching course. For example, both Alice and Bob used their car. Together, they collected a little over half of the tokens found ($185/337 \times 100\% = 54.90\%$). Thus, this portion of the tokens should be divided between them so that Alice gets 56.38% of the tokens and Bob gets the balance to reflect the type of athlete they are – amateur or professional.

Tables 1 and 2 illustrate the proportions required to obtain a distribution by groups and modalities according to the situation presented.

Table 1 ■ Percentages of an equitable distribution by type of athlete ($n = 337$)

	Amateurs	Athletes	Total
Car	54.90%	54.90%	
Bike	31.45%	31.45%	
Walk	13.65%	13.65%	
Total	100.00%	100.00%	

To produce the expected performance for each person, we gather frequencies of a proportional distribution among groups and categories following the proportions presented above. This prediction represents the number of tokens expected per person. The prediction is theoretical: there are decimals even though a token can't be split.

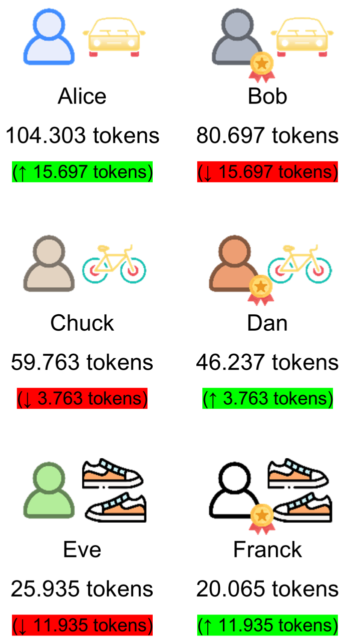
How close (or far!) is the prediction to the actual performance? Some performed better, some performed worse than expected. To find out, we subtract the predicted and the actual performance. The remainder is called a deviation. When a person finds more token than predicted, the deviation is positive. In that case, the concerned group in the cross-frequency table is said to be overrepresented for the corresponding qualitative category.

Figure 2 illustrates predicted performance for each person and the token difference with actual performance. For example, Alice found 120 tokens when 104.3 were pre-

Table 2 ■ Percentages of equitable distribution by means of transportation ($n = 337$)

	Amateurs	Athletes	Total
Car	56.38%	43.62%	100.00%
Bike	56.38%	43.62%	100.00%
Walk	56.38%	43.62%	100.00%
Total			

Figure 2 ■ Predicted performance and difference with actual performance. Left: pictorial representation; top right: Theoretical frequencies of the usual means of transportation of amateur sports enthusiasts' athletes ($n = 337$); bottom right: Deviations of observed and expected frequencies of the usual means of transportation of amateur sports enthusiasts athletes ($n = 337$)



	Amateurs	Athletes	Total
Car	104.303	80.697	185
Bike	59.763	46.237	106
Walk	25.935	20.065	46
Total	**190**	**147**	**337**

	Amateurs	Athletes	Total
Car	15.697	-15.697	
Bike	-3.763	3.763	
Walk	-11.935	11.935	
Total			

dicted. She overperformed, the deviation indicates a difference of +15.7 tokens. To the upper right, tables present theoretical frequencies of the usual means of transportation of amateur sports enthusiasts' athletes and deviations. Amateur athletes are overrepresented in the car category as indcated in the table on the lower right of Figure 2.

As another example, Eve found 14 tokens when she was predicted 25.93: she underperformed by -3.76 tokens. In her case, the deviation is negative, thus indicating that, in the survey, amateur athletes are underrepresented in the walking category.

Knowing which table cell or person over- or underperformed, the number of predicted tokens must be put into perspective: if this number is small, the difference of each additional token found by a given person will have more impact than if the number is large. Figure 3 below illustrates this conceptually. For each person, the number of tokens found is presented with a comparison to their prediction in percentages. On the right, the contributions to the chi-square statistic are presented on the right side of Figure 3. The cells with the highest contributions to the chi-square statistic correspond to the individuals that deviated most, in proportion to the number of tokens found, from their predicted performance.

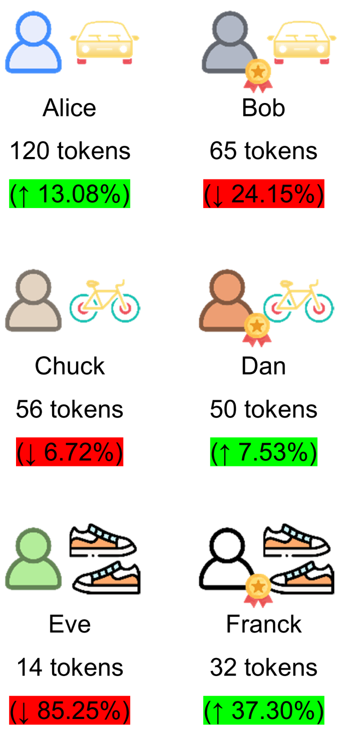
In other words, differences between types of athlete and means of transportation appear to be more pronounced for those who walk. For example, on one hand, Franck, the professional athlete found 32 tokens when 20.065 were predicted. He overperformed by 7.10 tokens or 37.30%. On the other hand, Eve, the sports enthusiast found 14 tokens when 25.935 were expected. She underperformed by 85.25%!

In the survey, in comparison with other means of transportation, the over-representation of athletes in the walking category contributes most to the difference between sports enthusiasts and athletes in terms of the usual means of transportation. Together, they contribute nearly two thirds of the sum of the contributions to the chi-square statistic ($12.591/18.550 \times 100\% = 67.86\%$). By relying solely on proportions, we wouldn't have been able to see this difference since they represent only 46 people on 337.

Strategy to Assess the Activity and Conclusion

Once the setting-driven story is completed, the activity is repeated in a slightly different version. During this second iteration, students must reconstruct a similar story in small groups. This second situation concerns a research object related to their academic discipline, for example sociology

Figure 3 ■ Predicted performance and difference with actual performance. Left: pictorial representation; right: contributions to chi-square statistic of the usual means of transportation of amateur sports enthusiasts and athletes ($n = 337$)



	Amateurs	Athletes	Total
Car	2.362	3.053	5.416
Bike	0.237	0.306	0.543
Walk	5.492	7.099	12.591
Total	**8.091**	**10.458**	**18.550**

or criminology. For example, it could be the type of crime for a sample of incarcerated adults and youth: violent crime, property crime, drug crime, etc. In this case, these offenders could also participate in a geocaching course for the story scenario. The question would then be whether being an adult or youth offender is associated with over- or under-representation by crime types. Students are encouraged to identify and use the structural elements to formulate the research question, but also the interpretation of the results. This second setting is partly guided.

Conclusively, a third situation empowers the students to interpret expected frequencies, deviations, and contributions by themselves. This one is not guided. The students must work on their own in small groups and present their story to the whole class.

References

Bizon, W. (2018). Tell them it's easy: Framing incentives in learning basic statistical problems. *Journal of Behavioral and Experimental Economics*, *76*, 76–81. doi:10.1016/j.socec.2018.08.001

Brisbin, A., & Maranhao do Nascimento, E. (2019). Reading versus doing: Methods of teaching problem-solving in introductory statistics. *Journal of Statistics Education*, *27*(3), 154–170. doi:10.1080/10691898.2019.1637801

Carver, R., Everson, M., Gabrosek, J., Horton, N., Lock, R., Mocko, M., . . . Witmer, J. (2016). *Guidelines for assessment and instruction in statistics education (gaise) college report*. Alexandria: American Statistical Association.

Cousineau, D., & Harding, B. (2017). Pourquoi les statistiques sont-elles difficiles à enseigner et à comprendre? *Quelques réflexions. Revue de Psychoéducation*, *46*(2), 397–419. doi:10.7202/1042257ar

Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using r*. Thousand Oaks: Sage.

Gilles, A., & Maranda, P. (1994). *Éléments de méthodologie et d'analyse statistique pour les sciences sociales*. New York: McGraw-Hill.

Harth, H. (2018). *An analysis of the teaching of introductory statistics at university in "context" [thesis* (Doctoral dissertation, Loughborough University). unpublished thesis. doi:10.26174/thesis.lboro.8246996.v1

McHugh, M. L. (2013). The chi-square test of independence. *Biochemia Medica*, *23*(2), 143–149. doi:10.11613/bm.2013.018

Quilici, J. L., & Mayer, R. E. (1996). Role of examples in how students learn to categorize statistics word problems. *Journal of Educational Psychology*, *88*(1), 144–149. doi:10.1037/0022-0663.88.1.144

Quilici, J. L., & Mayer, R. E. (2002). Teaching students to recognize structural similarities between statistics word problems. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *16*(3), 325–342. doi:10.1002/acp.796

Tay, D. (2022). Metaphor types as strategies for teaching regression to novice learners. *Journal of Statistics and Data Science Education*, *30*(1), 3–14. doi:10.1080/26939169.2021.2024777

Zeidner, M. (1991). Statistics and mathematics anxiety in social science students: Some interesting parallels. *British Journal of Educational Psychology*, *61*(3), 319–328. doi:10.1111/j.2044-8279.1991.tb00989.x

Citation

Chadillon-Farinacci, V. (2022). Focusing on structural features of a scenario: An activity on expected frequencies, deviations, and chi-square contributions. *The Quantitative Methods for Psychology*, *18*(2), v1–v6. doi:10.20982/tqmp.18.2.v001

Received: 07/04/202 ∼ Accepted: 15/07/202

Extended activity metadata

Concept illustrated	expected frequencies, deviations, and contributions (chi-square independence test)	*Type of activity*	In-class
Prerequisite	Descriptive statistics	*Types of data*	Storytelling
Co-requisite	NA	*Computation by*	NA
Suitable class size	30 undergraduate students	*Duration*	1 - 2 hours