MISLEADING STATISTICS

Author: Julie Vlachá

Supervisor: Mgr. Jana Krajčová, Ph.D., M.A.

CZECH TECHNICAL UNIVERSITY IN PRAGUE

MASARYK INSTITUTE OF ADVANCED STUDIES

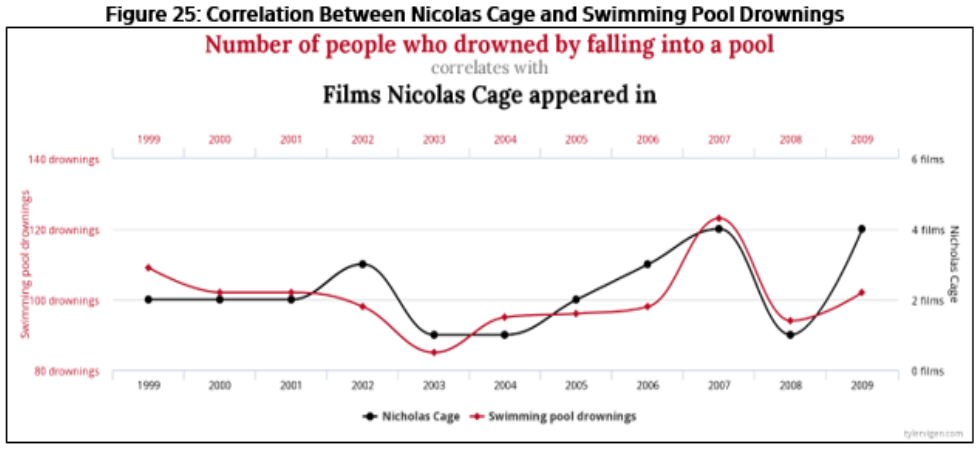
Abstract

Misleading statistics has become a serious issue in today's age of media. General public reads and misinterprets information published on the Internet which can lead to further misunderstandings. Understanding and distinguishing the difference between low-quality media articles and actual research publications is of key importance nowadays. This thesis summarizes the basic concepts of statistics which general public is often making mistakes in. Further, it demonstrates several real-life cases which media often uses to mislead or shock the reader in order to warn against the outcomes and opinions which may appear due to the misleading information.

Abstrakt

Zavádějící statistika se stala závažným problémem dnešních médií. Široká veřejnost mnohdy čte a špatně interpretuje informace publikované na internetu, což může vyústit v nedorozumění. Pochopení a rozpoznávání nekvalitních zdrojů od akademického a vědeckého výzkumu je v dnešní době nezbytné. Tato práce shrnuje základní statistické koncepty, ve které široká veřejnost mnohdy chybuje. Dále poskytnu názorné příklady, ve kterých média zavádějí čtenáře, popřípadě se snaží výroky šokovat. Touto prací chci varovat před následky, které mohou nastat při nesprávném používání a interpretování statistiky.

Figure 25: Correlation Between Nicolas Cage and Swimming Pool Drownings



Source: Tyler Vigen

Table 4: Daily production and 4-day moving average

A	A_{ma}	B	B_{ma}
5	-	9	-
8	-	7	-
6	-	10	-
9	7	8	8.5
7	7.5	11	9
10	8	9	9.5
8	8.5	12	10
11	9	10	10.5

The correlation coefficients for each set of variables was as follows:

$$for\ A, B: r = -0.089087081$$

$$for\ A_{ma}, B_{ma}: r = 1$$

The low coefficient in case of variables A and B means that for product B, the leftover capacity from lower production of product A will be used (or vice versa) as there is no correlation. If production is described with moving averages (for an evident trend), the correlation coefficient shows a strong correlation.

Conclusion

Although, data-driven policies and decision-making have many positive impacts on socio-political and economical lives, statistics is, more often than not, heavily misunderstood and sometimes even overused by people who lack proper understanding of statistics. Currently, there are numerous online platforms that, intentionally or unintentionally, spread misleading information. At times, these information/claims are supported by misinterpreted or misrepresented statistics; while other times, the online platforms simply do not provide information on how the data was collected, processed, analysed, and interpreted.

This thesis is mainly aimed at the general public which often does not realize the dangers of misleading or false information. Media often creates shock or even panic due to the spread of misleading information. The tendency of the Internet (and media generally) to spread misinformation faster than promoting accurate and reliable sources demands for each consumer of media to consider learning more about the way the media works and critically read media published articles. Understanding the basics of statistics could possibly help the readers to distinguish between the good and the bad sources. Furthermore, statistics can help the reader expand on their critical thinking skills as basic mathematical arguments can be utilized in every day arguments based on logic.

Works Cited

Akkerboom, H., & Dehue, H. (1997). The Dutch Model of Data Collection Development of Official Surveys. *International Journal of Public Opinion Research*, 126-145.

Berk, R. A. (1983). An introduction to sample selection bias in sociological data. *American sociological review*, 386-398.

Christensen, L., Johnson, R., & Turner, L. (2014). *Research Methods, Design, And Analysis*. Boston: Pearson.

Fisher, R. J. (1993). Social desirability bias and the validity of indirect questioning. *Journal of consumer research*, 303-315.

Gardenier, J., & Resnik, D. (2002). The misuse of statistics: concepts, tools, and a research agenda. *Accountability in Research: Policies and Quality Assurance*, 65-74.

Geenstone, M., & Gayer, T. (2007). Quasi-Experimental and Experimental Approaches to Environmental Economics.

Grinberg, N., Joseph, K., Friedland, L., Swire-Thompson, B., & Lazer, D. (2019). Fake news on Twitter during the 2016 US presidential election. *Science*, 374-378.

Levitt, S. D., & List, J. A. (2008). Field experiments in economics: The past, the present and the future. *European Economic Review*, 1-18.

Madsen, K. M., Hviid, A., Vestergaard, M., Schendel, D., Wohlfahrt, J., Thorsen, P., . . . Melbye, M. (2002). A population-based study of measles, mumps, and rubella vaccination and autism. *New England Journal of Medicine*, 1477-1482.

Moura, H. (2011). Sharing Bites on Global Screens: The emergence of Snack Culture. In D. Y. Jin, *Global Media Convergence and Cultural Transformation: Emerging Social Patterns and Characteristics* (p. ch. 3). Hershey.

Mrozek-Budzyn, D., Kieltyka, A., & Majewska, R. (2010). Lack of association between measles-mumps-rubella vaccination and autism in children: a case-control study. *The Pediatric infectious disease journal*, 397-400

And others...