Predicting food consumer and customer behavior

Ágoston Temesi[1], …, Zoltán Lakner[1]

[1] Department of Agricultural Business and Economics, Institute of Agricultural and Food Economics, Hungarian University of Agriculture and Life Sciences, 1118 Budapest, Hungary

Introduction

Importance of understanding food consumer behavior patterns has gained new emphasis with the appearance of the UN Sustainable Development Goals (sdgs.un.org, 2022), as the achievement of several goals is specifically linked to food consumer and customer behavior. The most visible of these are "zero hunger", "responsible consumption or production" or "good health and well-being", but almost all 17 goals can be linked to our food related behavior. But it is equally important for lifestyle-related diseases and the rise of healthy lifestyles in ageing societies (Chen and House (2021).

However, one of the keys to the broad utility of consumer behavior researches, including food consumer behavior surveys, is to survey not only one country, but preferably several countries and continents. It has become a trend in international research to conduct international/cross-cultural surveys on food consumption. The number of such studies has been considerably increasing over the last 25 years, as illustrated by the fact that more than 10% of the 829 articles were written in the last year alone (webofscience.org, 2022).

With the increasing importance of international, cross-cultural research, the question arises to what extent the food consumption patterns of different countries correspond to each other, whether food consumption behavior understood in one country can be assumed to be similar in another country, while of course we know that there are many potential errors when comparing international research (e.g. Scholderer; 2010).

Aim of research

The aim of our research is to determine how accurately an algorithm trained on a sample of 6 countries, or on a social stratum in one country, can predict expected food consumer and customer behavior in other countries.

Q1: Our first research objective is to investigate the accuracy with which food consumer and customer behavior in a given country can be predicted by an algorithm trained on food consumption related statements in another country. Given that our database contains a representative database of food consumers in 6 countries, we aim to perform this analysis for each country.

Q2: Our second research objective is to identify the attitudinal statements that are most helpful in predicting food consumer and customer behavior.

Q3: Finally, our research aims to investigate whether training the algorithm on a social stratum rather than on a whole sample of a country can lead to more accurate predictions.

Methodology

Data sampling

Between 2017 and 2019, we surveyed consumers in 6 countries: Denmark, Australia and Hungary in 2017 and the United States, the United Kingdom and New Zealand in 2018/2019. In 5 countries we recorded responses online and in one country (Hungary) we recorded responses in person. Care was taken to ensure that respondents were always the food customer in their household. Respondents aged 18 and over were interviewed. All countries achieved representative sampling: the Danish sample followed the gender ratio of the population, the Australian, American and Hungarian samples followed the regional distribution within the country.

The sample composition is presented in Table 1.

Table 1

Socio-demographic composition of the sample

	DK (n=508)	AU (n=505)	HU (n=500)	UK (n=519)	US (n=809)	NZ (n=526)
Age groups						
under 30	19.3%	24.2%	12.8%	12.1%	24.5%	26.4%
31-44 years	22.0%	33.9%	30.4%	24.7%	27.1%	23.2%
45-59 years	32.7%	29.9%	31.8%	30.3%	27.3%	25.1%
60 years old and more	26.0%	12.1%	25.0%	32.9%	21.1%	25.3%
Gender						
male	50.4%	36.2%	25.0%	51.3%	51.9%	55.1%
female	49.6%	63.4%	75.0%	48.6%	48.0%	44.1%
other/no answer		0.2%			0.1%	0.8%
Education						
primary school	6.1%	2.0%	13.2%	0.0%	6.6%	0.4%
secondary school	48.8%	64.0%	66.4%	59.3%	33.9%	54.0%
tertiary education	45.1%	34.1%	20.4%	40.7%	59.6%	45.6%
Number of children						
no children	75.0%	53.5%	65.8%	72.6%	68.6%	62.4%
1 child	12.2%	19.4%	15.8%	13.1%	14.3%	18.6%
2 children	9.1%	16.2%	13.0%	8.5%	11.1%	11.8%
3 children	3.0%	7.9%	3.8%	1.9%	3.5%	3.8%
4 or more children	0.8%	3.0%	1.6%	3.9%	2.5%	3.4%

Data analysis

The master questionnaire was originally written in English and translated into Danish and Hungarian. During the years of data collection, the questionnaire to be surveyed has undergone minor modifications, so only those questions from the originally longer questionnaire that were asked in exactly the same way in each country were kept for data analysis. Thus, the scaling statements, the questions on purchasing and consumption patterns and the demographic variables were brought together and only those statements and questions that were asked in the same way in all countries were kept. Accordingly, the number of uniformly asked scaling statements that are exactly the same in each country and can be analysed accordingly is 84, and the number of statements measuring food purchase and consumption is 30.

In the analysis, first, factor and cluster analysis was performed using SPSS software to analyse the 30 food purchase and consumption statements. The factor analysis was performed using varimax rotation. The analysis resulted in 6 factors. Working further with these, we identified 3 clusters using K-means cluster analysis method.

These three clusters became the dependent variables in our next analysis, which was conducted using the Random Forest tool (Breiman, 2001) using the RandomForest machine learning package of the R software. 84 scaling statements describing food consumption behaviour were used to predict which cluster the respondent would be placed in. This was first tested on the full sample, then the program was trained on a sample from one country and its accuracy tested on the sample from another country, and finally comparisons were made for different social groups.

Results

The clusters of food consumer and customer behavior are illustrated in Table 2. Considering that the food consumption and purchase statements used to characterise the clusters are ordinal scales, the mode values have been chosen to represent the clusters.

Table 2.

Clusters based on food consumption and purchasing behavior

	Mode (n=3367)		
	Cluster "A" **simplicity seekers**	Cluster "B" **demanding**	Cluster "C" **practical housewife**
I eat bread / I drink milk	7	5	7
I eat fish	5	5	4
I eat lollies, desserts and/or cakes	5	4	5
I drink wine	5	5	1
I drink beer and/or cider	**5**	1	1
I spend more than one hour in the kitchen weekdays preparing main dish	5	5	**7**
I eat pulses (lentils, peas, beans etc.) / I eat dinner in a restaurant	4	4	1
I shop at the fruit and vegetable shop	1	5	5
I shop at a convenience store / I shop at the food or farmer's market / I shop at the bakery / I shop at the butchers / I eat breakfast on the go / I eat lunch at work / I eat snacks on the go	1	**5**	1
I shop online / I shop at a farm gate/shop / I shop at the fishmonger	1	**4**	1
I spend more than one hour baking	1	4	4
I eat fruit / I eat vegetables	7	7	7
I shop at the supermarket / I eat salad / I eat red meat / I spend more than one hour in the kitchen weekends preparing main dish	5	5	5
I eat creamy or buttery sauces / I eat pizza	4	4	4
I shop at a cheese shop	1	1	1
Cluster size (%)	40.4%	26.2%	33.4%
n	1360	883	1124

*Scales from 1 to 7, where a higher value indicates a higher frequency (1 - never, 2 - less than every 6 months, 3 - 1–5 times every 6 months, 4 - 1–3 times every month, 5 - 1–2 times per week, 6 - 3–4 times per week, 7 - every day or almost every day)

The clusters are characterized by demographic variables as follows (all characteristics are significant differences at $p<0.01$). Most members of cluster "A" are Danish (25.0%), least members are Hungarian (5.4%). This cluster includes the largest number of single-adult households (34.4%) and the largest number of households without children (75.7%). This cluster includes older respondents: most respondents aged 45-59 (32.6%) and 60+ (32.4) were in this cluster. This cluster has been named "simplicity-seeking" based on their food consumption, food purchasing and socio-demographic characteristics.

Most members of cluster "B" are American (35.4%), least members are Danish (6.3%). Cluster B has the highest number of respondents with tertiary education (52.1%). The cluster is characterised by younger respondents. It includes most respondents under 30 (34.9%) and 30-44 years old (32.7%). In addition, most households with 2 (16.0%), 3 (5.7%) and 4 or more children (3.6%) are also in this cluster. This cluster has been termed "demanding" and includes working (large) family respondents by socio-demographic characteristics.

Cluster "C" was composed mostly of Hungarian respondents (27.4%) and least of Danish respondents (10.1%). This cluster has the highest proportion of women (58.2%) (and the lowest number of men) respondents and the highest proportion of people with secondary education (64.1%). The third cluster is therefore composed mainly of "practical housewives" who are often involved in cooking and baking.

Predictions

In the first analyses, we investigated the accuracy with which the randomForest algorithm trained in each country can predict the cluster to which a respondent belongs in another country, using the 84 scaling statement (see Table 3).

Table 3.

Accuracy of prediction on base of total population

		Test countries						
		DK	**AU**	**HU**	**UK**	**US**	**NZ**	**All other countries**
Training ("pattern") countries	**DK**	1	0.414	0.202	0.578	0.533	0.443	0.446
	AU	0.516	1	0.524	0.504	0.513	0.523	0.515
	HU	0.278	0.527	1	0.299	0.447	0.498	0.417
	UK	0.675	0.390	0.178	1	0.481	0.426	0.438
	US	0.703	0.511	0.310	0.588	1	0.500	0.522
	NZ	0.612	0.580	0.552	0.434	0.536	1	0.555

Statements with the largest pooled MeanDecreaseAccuracy:

- I keep a food diary to track my calories.
- I use apps and/or wearable technology to monitor my dietary intake.
- I often take photos of food when dining out to share with friends on social media
- I like to take photos of food cooked at home to share with friends on social media.
- I use recipe apps to generate shopping lists.

* y axis: country training the algorithm (independent variable); x axis: country/all other countries validating the algorithm (dependent variable)

Given that we have distinguished three clusters based on food consumer and customer behavior, our results can be considered relatively accurate at a value of at least 0.5 (compared to a random value of 0.33), which is true for 17 out of 25 cases. The Danish sample was the most lagged in terms of predictability, which can be predicted with reasonable accuracy by the behavior of British, New Zealand and especially American shoppers. The Hungarian sample was the least accurate in predicting the behavior of respondents from other countries in the sample.

The second projection was only carried out among those with tertiary education. Our results are presented in Table 4.

Table 4.

Accuracy of prediction among those with tertiary education

		Test countries					
		DK	**AU**	**HU**	**UK**	**US**	**NZ**
Training ("pattern") countries	**DK**	1	0.483	0.275	0.578	0.571	0.488
	AU	0.576	1	0.392	0.597	0.500	0.521
	HU	0.205	0.390	1	0.270	0.311	0.417
	UK	0.707	0.453	0.245	1	0.554	0.442
	US	0.769	0.535	0.304	0.611	1	0.558
	NZ	0.664	0.512	0.373	0.512	0.554	1

Among those with tertiary education, our results show only a few cases of notable change compared to the overall sample. The Danish sample is even more accurately predicted by all countries except the algorithm taught on the Hungarian sample. The algorithm taught on the US sample can more accurately predict the UK sample. Hungarian consumers with higher education, on the other hand, seem to be far from the food consumption patterns of the other five countries, neither can explain these patterns of behavior nor can they explain the pattern of behavior of other countries.

For our third analysis, we have chosen a trend that is currently shaping food consumer behaviour: responsible food consumer and customer behavior. In our previous article (Brunsø et al., 2021) the "food responsibility" scale has been created. In our analysis, we did not include respondents who tended to disagree with these statements (those who answered 1-3 on the 7 point scales – 5 scales). Our results are illustrated in Table 5.

Table 5

Accuracy of prediction among respondents with food responsibility

		Test countries					
		DK	**AU**	**HU**	**UK**	**US**	**NZ**
Training ("pattern") countries	**DK**	1	0.438	0.167	0.561	0.573	0.475
	AU	0.528	1	0.464	0.409	0.516	0.504
	HU	0.252	0.504	1	0.325	0.498	0.493
	UK	0.674	0.426	0.217	1	0.523	0.461
	US	0.739	0.500	0.181	0.553	1	0.475
	NZ	0.647	0.621	0.420	0.515	0.580	1

Table 5 shows that the responses of British, American and New Zealand respondents who are responsible in their food consumer and customer behavior are even more accurate in predicting the behavior of similar Danish consumers.

In Table 3, it is noticeable that scales that measured food consumer and customer behavior related to technology and social media were the most useful in helping the algorithm to work. In our final analysis, we therefore excluded respondents who tended to disagree with these statements (those who gave responses from 1 to 3 on the 7 point scales – 5 scales). Our results are illustrated in Table 6.

Table 6

Accuracy of prediction among technology and social media users in food consumer and customer behavior

		Test countries					
		DK	**AU**	**HU**	**UK**	**US**	**NZ**
Training ("pattern") countries	**DK**	1	0.726	0.632	0.581	0.727	0.875
	AU	0.840	1	0.789	0.516	0.715	0.670
	HU	0.840	0.762	1	0.484	0.703	0.670
	UK	0.520	0.512	0.316	1	0.517	0.591
	US	0.800	0.762	0.737	0.516	1	0.670
	NZ	0.880	0.726	0.684	0.548	0.698	1

Our results show that those who are avid users of technology and social media show the greatest similarities across the 6 countries. In almost all cases, a training algorithm on this respondent group in one country made a relatively accurate prediction in another country.

Summary

The aim of our research was to examine the comparability of food consumption patterns. Using the Random Forest method, we conducted pairwise comparisons on a representative sample of food customers from 6 countries. To carry out the prediction, factor and cluster analysis were performed and three clusters ("simplicity seekers", "demanding" and "practical housewives") were distinguished. In the analysis, the algorithm was applied to food shopper respondents in one country and the accuracy of the algorithm was tested in the other countries. Our results also further clarify the differences and similarities in food consumer behaviour across the 6 countries. Our results show that the Danish food consumption pattern is the most predictable. In particular, the responses of British, American and New Zealand respondents are the most suitable. It can also be seen from our results that geographical proximity does not necessarily give a country sample a good predictive ability. Finally, we have shown the homogenising effect of technology and social media use on food consumption and purchasing behaviour.

References

Brunsø, K., Birch, D., Memery, J., Temesi, Á., Lakner, Z., Lang, M., Dean, D. & Grunert, K. G. (2021). Core dimensions of food-related lifestyle: A new instrument for measuring food involvement, innovativeness and responsibility. *Food Quality and Preference*, 91, 104192.
sdgs.un.org (2022): Sustainable Development Goals of UN. https://sdgs.un.org/ (accessed: 12.04.2022)
Chen, L. A., & House, L. (2021). Food lifestyle patterns among contemporary food shoppers. *International Journal of Consumer Studies*.
webofscience.org (2022): https://www.webofscience.com/ (accessed: 12.04.2022) boolean: TS=(("cross-cultural" OR "cross cultural" OR "international" OR "across nations") AND ("food consum*"))
Scholderer, J. (2010). Data handling in cross-cultural studies: Measurement invariance. In Consumer-driven innovation in food and personal care products (pp. 470-487). *Woodhead Publishing.*
Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.