University of Business and Technology in Kosovo

UBT Knowledge Center

UBT International Conference 2019 UBT International Conference

Oct 26th, 3:15 PM - 3:30 PM

Cost analysis on Cost of Consumers in Terms of Small markets. (Case study, Vlora)

Elmira Kushta
University Ismail Qemali Vlora, kushtamira@gmail.com

Fatmir Memaj
University of Tirana, fmema@yahoo.com

Follow this and additional works at: https://knowledgecenter.ubt-uni.net/conference

Part of the Business Commons

Recommended Citation

Kushta, Elmira and Memaj, Fatmir, "Cost analysis on Cost of Consumers in Terms of Small markets. (Case study, Vlora)" (2019). *UBT International Conference*. 356.
https://knowledgecenter.ubt-uni.net/conference/2019/events/356

This Event is brought to you for free and open access by the Publication and Journals at UBT Knowledge Center. It has been accepted for inclusion in UBT International Conference by an authorized administrator of UBT Knowledge Center. For more information, please contact knowledge.center@ubt-uni.net.

Cost analysis on cost of consumers in terms of small markets. (Case study, Vlora).

Elmira Kushta[1], Fatmir Memaj,[2] Gladiola Trushaj[3]

[1]Department of Mathematics, Faculty of Technical Sciences, University “Ismail Qemali”, Vlore
[2]Department of Statistics, Faculty of Ekonomy, University of Tirana.
[3]Teacher at Alidemi Vlore high School.

Abstract. This article deals with the observed changes in consumer spending in a region of Albania following the fall in market sales prices. The paper focuses on the statistical study of a local consumer system. In the meantime, we are striving for greater recognition. Respectively in a more generalized and mathematical perspective, the generalized customer or consumer is the user and consumer as the behavioral agent. The purpose of the paper is to identify the best or most appropriate model for the behavioral profile according to the profile determined by the inductive analysis of the variables themselves. For this reason, part of the purpose of the paper is to construct the behavioral profile ie what are the most characteristic variables and the factor group profile. In the region where the study was conducted, it has been found that the price reductions drive costs, which in turn shows that the consumer behaves as a minimalist buyer, in response to high prices rather than budgetary constraints.
This supports the view that consumers in this region are primarily rational consumers It was then found that the most pronounced effect relates to average consumer spending suggesting marketing activity referring to this category of consumers.

Keywords: Probit model,behavior al models,factor analysis.

Introduction

Consumer behavior is an opinion formation process which is difficult to be studied quantitatively and rather complex as evidenced in [1]. In a formal approach they are assumed to act rationality in their decision making by optimizing some utility function, but this last cannot be measured directly. Meanwhile the assumption of rationality does not hold always which is clarified by behavioral theories as discussed in [2], [3] etc. Behavior seems to be too complex to be studied and analyzed by deterministic methods. Even so, researchers and scholars have outdone this difficulty by using statistical tools and probabilities facilitating modeling in this case as presented in many textbooks. Mixed calculation using econometric optimization and network dynamics have been developed as for example in [4] and many other applications. Generally, the consumer’s decision making process in buying is complex but econometrically known and measurable. Aside of general models and regressions, practical calculation have demonstrated their capacities to describe consumer behavior in specific systems as in [5], [6] and many others. In our recent work [7] we applied a logistic regression to identify the consumer profile in a specific area, the factors affecting their behavior and other parameters characterizing the system of consumer attitudes and activities. This study is intended to evaluate a marketing aspect as discounts for example, by specifically considering the nature of

the state of the system, the possible presence of not-apparent factors as latent effects or hidden variables etc.
The study of concrete economic environment unavoidably face challenges to the scholars. Standard questionnaires that aims gathering information from social or economic mediums include different types of variables, non-numerical responses, questionable answers, missing or incomplete records etc. Usually the inquiries might be organized and held in different moment of times. Finally they must be included in modeling say linear multivariate functions

$$Y = A * X + \varepsilon; \qquad (a)$$
$$Y = W * Z + u \equiv W * (\Gamma * X + v) + u \qquad (b) \qquad 1.1$$

or in the logistic type relationship

$$f(z) = \frac{1}{1 + \exp\left(-\alpha + \sum_{i=1}^{n} \beta_i x_i\right)} \qquad 1.2$$

where X, are factor variables or predictors and Y are responses variable or indicators whereas u,v,ε are errors and A,W, Γ etj are matrices. Versions 1.1.(a) are the simplest relationship in the models. In the cases 1.1 a regression procedure leads to the calculation of the matrices of coefficients which explain the weight of each variable (i) in responses (j). In the case of 1.1.(b) the problem includes the calculation of the so called latent variable (Z) adding to the coefficients matrices W and Γ. The belongs to structural equation or SEM systems, discussed in [6] , [7] and used largely in sociology [10], econometrics [11], explanatory medicine etc. In all those cases, some necessary statistical assumption should be fulfilled. It happen that in real systems many of them does not hold. Therefore quantitative methods needs for more analysis as seen in the reference [9] and others related to this aspects. In this case, some approximate methods are suggested and elaborated as in [10] or in a more dedicated case in time series in [12]. However, in general it depends on the concrete properties of the data series. In general, preparatory analysis or data elaboration is needed. The second problem is related to the tangible set of the variables included in the models. Again, standard models belong to the standard systems and in real ones there is a considerable difference. But by carefully using simple analytic tools it is possible to avoid the complexity of the model, to control extra errors added during calculation phase and improve overall calculation. In our recent research in the analysis of consumer behavior in district of Vlora, we considered such specifics as an important step [1], [2] etc. The last issue that can affect directly the quality of the modeling is the representable property of the data gathered from measurement related to the sampling process. In practice an appropriate size of the sample might not be accessible [3] or it is difficult to be stated. For numerical continuous and normally distributed random variable , the working formula is

$$n = Z^2_{\frac{\alpha}{2}} \frac{p(1-p)}{M\arg in\_of\_Error} = \frac{\frac{z^2 p(1-p)}{e^2}}{1 + \frac{z^2 p(1-p)}{e^2 N}}$$

where z is the normalized variable $z = \frac{x - \langle x \rangle}{\sigma(\langle x \rangle)}$, Z is the critical value or level α, N is the population size and e is the level of tolerance adapted and p is the sample proportion. But it is difficult to be estimated if we consider categorical variables. In this case we proposed to choose a sample size according to numerical variables and accordingly to use auxiliary statistical tools to identify the error injected in the system by such approach. Next consider that in theoretical approaches one assumes stationary for the system states, homogeneity, formal relationship etc. Detailed analysis on those aspects are provided in many articles-guides and statistical books as [3] or [4]. In this case the problems could be overcomes if we adjusted correctly the sample size or adopt a suitable sampling method. In the case where the above step is not suitable or even impossible, the factorial and descriptive analyses could be used as recommended in standard procedures to manage the sampling error, [5] and general consideration [13].

Verifying statistical meaning to the variables

The data set of our study consists in some general properties of buyers (consumers) that are expected to affect the response behavior which consists in the way (s)he does his expenses. They looks like the shortened Table 1 By first approach we realized the measurement by using standard models and methods over education, family size, gender, employment state etc., that are factor variables whereas quantities of expenses for specific items are response variables. Initially the formula 1.1., 1.2 were adapted and applied using those data as pilot modeling. The results were unsatisfactory. Logistic model 1.2 doesn't worked at all if using response variables as taken from the system. formula 1.2 looks unable to make a relationship between some conditions of the buyer and his expense in goods (i). To check if the model fails or variables were un-appropriate we explored the results of the logistic model 1.2 by removing randomly individuals in the sampling data and next liming the model in different categories of expenses. It resulted that by changing individuals, the outcome of the model does not changes abruptly whereas by referring different expenses (y) the model 1.2 gave contradictory results. Next we observe that the distributions of values of expenses (y) were highly irregular and no shape was unidentified, contrary, the relative values if the expenses showed smooth distribution.

General Information: factor variables			Expenses: Response variables		Complementary data	
Family type	Education	Other variable	Foods [Lek/monthly]	Other Expenses in the monthly bounded	Relationship with household head	Other properties
2	Professional/College;	..	60000	5000	Head of family	…
3	University	…	50000	30000	…	…

To normalize this, we proposed to use statistical counterpart of some response variables introducing the relative weight of particular expenses.

$$X \rightarrow \frac{X_i}{\sum_{i=1}^{NumberVariables} X_i} \equiv \frac{Specified\_Expense}{Total\_Expenses} \qquad 1.3$$

Secondly we tried to obtain the most stationary variable in y. It resulted that the grouped expenses showed more stable distributions in the bases of analysis [9]. Usually the stationary of the distribution is related to the stationary of the state and statistically we can use each on in the same context. In [10] it has been proposed a very interesting functional form that report directly the stationary of the distribution by the parameter q:

$$p(x)=\frac{1}{Z}[(1+\beta(1-q)(x-\mu)^2]^{\frac{1}{1-q}}=\frac{1}{Z}[(1+\frac{(1-q)}{5-3q}\left(\frac{x-\mu}{\sigma}\right)^2]^{\frac{1}{1-q}} \qquad 1.4$$

q-Gaussians are in stationary state if $1\leq q\leq\frac{5}{3}$ and converge to the Gaussians if q→1 as detailed in [9]. In fact we observe that among many candidates the q-Gaussian and lognormal distribution fits the data better than other distributions tested. In addition and to count for mixed multiplicative properties as usually expected for complex dynamics, we use q-lognormal as detailed theoretically in the reference [4]

$$p(x)=\alpha\frac{1}{x^q}\left[\left(1-\beta(1-q)\left(\left(\frac{x^{1-q}-1}{1-q}\right)^{1-q}-\mu\right)\right)^2\right]^{\frac{1}{1-q}}$$

The distribution should be identified by a fully optimization procedure. The key element in picturing distribution is the optimization of the bin size based on Fredman-Diaconins algorithm. In Figure 1 ., the near to power law distribution for expenses in alcoholic drinks for example tagged it as an un-appropriate variable. Grouping them in four expenses type respectively {basic-vital, necessary, quality-life luxury} expenses resulted in variables with stationary distribution identified in their q partner below 5/3.

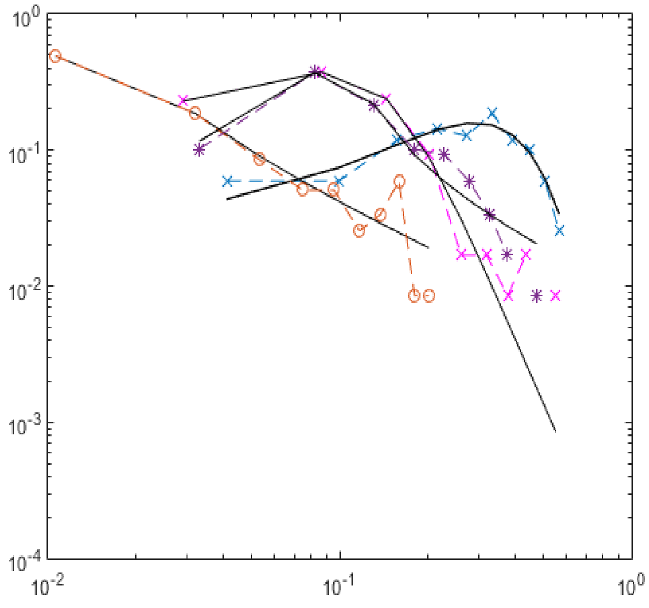

Figure 1: Logarithmic view of distributions for some variables .

Table 1: Variables of the model

Variable		Q-Parameter	Mean	R2	Stationary : Q<5/3

Foods And Non-Alcoholic	P6	2.3193	0.1727	0.9817	FALSE
Alcoholic Ad Cigarettes	P7	2.9968	0.0582	0.9218	FALSE
Clothes	P8	1.9457	0.0884	0.9941	FALSE
Subsistence	P9	1.4498	0.062	0.9992	TRUE
Health	P 10	1.646	0.0069	0.9999	TRUE
Transport	P11	1.9654	0.0169	0.9982	FALSE
Communication	P12	2.1387	0.0932	0.999	FALSE
Entrainments And Child Care	P13	1.7068	0.0207	0.9989	FALSE
Education	P 14	2.3654	0	0.9942	FALSE
Luxury	P 15	1.7264	0.0208	0.9986	FALSE
Services	P 16	2.6702	0.0452	0.9996	FALSE
All Expenses		1.6707	0.1283	0.9987	FALSE

Estimation of the utility and latent variables.

In the analysis of the effect of sales and discounts we proposed to analyze two scenarios: an hidden influences is assumed to be present or the outcome of the behavior is driven form a latent utility. In the first stage we perform factorial analysis to observe any possible reduction in the indicators, analyses as in [15]. We obtain that the number of latent variables could be larger than 2 but 95% of variance can be expressed in the tow first hidden components. Basing on the in the non-stationary of the state and the fact that after sales the state become more relaxed, we hypothesized that another hidden variable affect the purchasing behavior and checked it using MIMIC model. In this case we consider predictors variables in *X=[Gender, Age. Group, Average. Visits Average. Expenses Phone. Contact, Regular. Client], Y=[Average Expenses after sales, Average Visits after Sales]*. Performing full structural model calculation under the assumption of an extra variable we obtain that for this choice, a single hidden variable will explain good the intermediary relationship between Causes X and consequences Y.

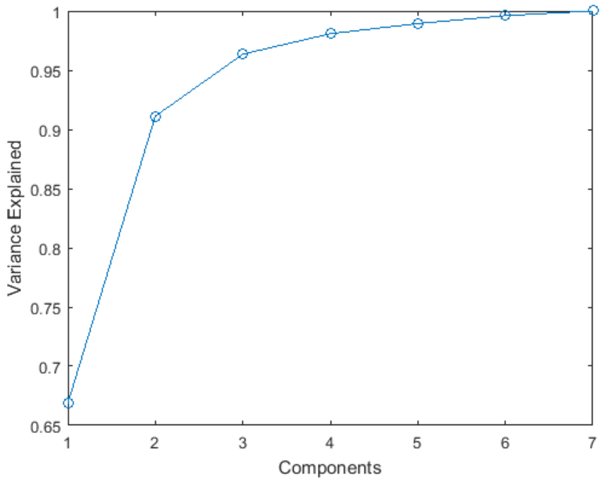

Figure 2: Identification of hidden or latent components

In table 2 we show parameter of linear equation *LV~ Parameters*Factors*

Table 1: Parameters of hidden variables

	Parameter Matrix A				Parmeter Matrix B	
Factors	Two HV		One H.V	Observed		1 HV-Mode
Free parameter	0.0046	0.526	**0.0046**	SpendingAfterSales		-7.9368
Gender	1.0743	1.3994	**1.0743**	VisitsAftersales		141.482
AgeGroup	-0.2782	83.5887	**-0.2782**			
AverageVisits	-1.8217	-3.5924	**-1.8217**			
Average Spending	-0.3789	-0.4879	**-0.3789**			
RegularClient	0.0363	0.1666	**0.0363**			
TelephoneContact	141.482	0.0798	**141.482**			

Interestingly, the factors have different effects on average spending after sales and average visits. As seen in the Table 2, the parameters remain unchanged (up to 3 digits) in modeling with one and two hidden variable and therefore we restrict the model with one single hidden or latent variable. In this case the hidden variable could act as an interconnection between causes and outcomes. The reduction in the number of latent variables is plausible for the model because in this case we can use the utility as the intermediate variable ore stage in the consumer decision making. So far, we assume that the overall decision of the consumer to increase the expenses after discounts have been applied could be interpreted by a continues utility function

$$u_j = \beta_o + \sum_{i=1}^{n} \beta_i x_{i,j} \qquad (1.5)$$

where j is the individual observation and i are variables. The response will be a dichotomous as follows

$$P(Y = y_i | X) = P(a_{i-1} \leq u < a_i | X) \qquad (1.6)$$

and for our binary output there is only one point to be considered say the moment where the continuous probability take the value 0.5.

Firstly we consider the attractiveness of the discounts, so we examine increasing number of market visits after discounts were applied. Here we use chose as dependent variables the positive change in the average number of visits after discount; and for independent factors the gender of buyer, the age group and contacts by calls to announce the offers. By applying probit regression we observe that a good fit is obtained and the marginal errors are normally distributed as seen in the figure (3). The coefficients have been confirmed as different from zero within 90% confidence, whereas the free coefficient seems to not pass the test

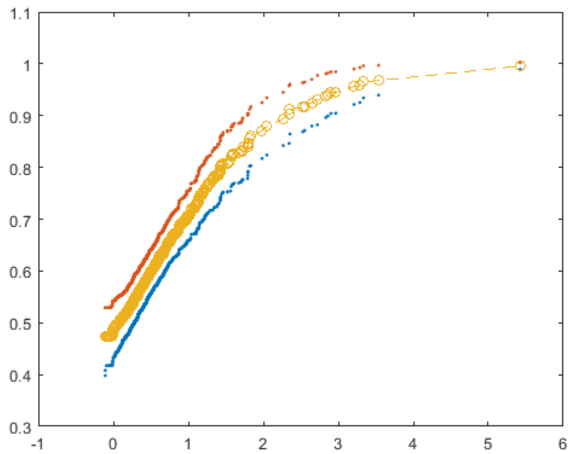

Figure 3: Probit regression for Increasing Expenditures after discounts

Therefore the utility of the attractiveness is obtained by probit regression as follows

$$Y^* = \{0.0698\} \ -0.1367 * \text{GenderBuyer} + 0.0962 * \text{Age Group} + \ 0.0002 * \text{PhoneContact} + \varepsilon \tag{1.7}$$

From relation (1.5) we observe that the gender of buyers (F=1,M=2) is mostly decisive in the increasing number of visits in the market after sales, and usually male buyers are not more frequent in the market after discounts have been applied. The phone contact has a slight effect on it. The age group has comparable role to the gender of consumer. In Figure 4 is seen that the probability for more visits in the market is high for almost all the values of utility function (1.7) and only few values are less than 0.5. In this sense, for nearly all consumers' specifics, the marketing strategy (prices discounts) has been found attractive for peoples that respond by increasing the number visits in the market. Thus is the intermediate change on the consumer behavior. In the second stage, the final behavior is considered.

Now the response variable is the change in expenditure measured by the natural function "is greater than" e.g., in absence of the marketing stimulus. It is possible that the buyer, under budget constraint, would respond to the discount spending the same quantity of money and therefore just buying some more goods, so this variable is meaningful and not trivially known. Here the independent variables include even average expenditures before sales and registered cards consumer. The first is expected to give information about which consumer category has increased the expenses, and the second could inform the role of being a formal consumer. Performing probit regression we obtain the utility function

$$y^* = -0.916 + 0.093 * C.Gender \; -(0.0133 * C.AgeGroup)$$
$$+ 0.0199 * AverageVisit + 0.854 \, AverageExpenses$$
$$- 0.136 * CardHolder + (0.0013 Call) + 0.0396 PeriodOpen \quad (1.8)$$

In (1.8), the statistical significance is acceptable for all variables except Tel,Call and Age.Group (the p.Value is high, ~0.3) so we put it in parentheses. Notice that their coefficients are small and so this does not affect the estimation of the utility so we kept them in the equation (1.8). Now we make use of the binary outcome expression using continues probability. The switching value for utility is

$$P(ExpencesAfter > ExpencesBeforeX) = P(u \geq 0.08) \quad (1.9)$$

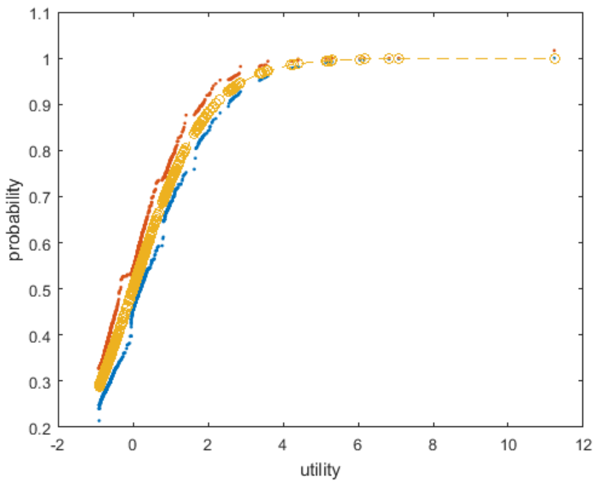

Figure 4: Full model probit regression

By using equation (1.7) we can realized conditions that (1.8) is fulfilled and therefore it is possible to forecast what happen with purchasing behavior for an individual with values

$$Consumer\{i\} \equiv n_1 \, n_2 \ldots . n_7 \quad (1.10)$$

This is done by just putting values (1.9) in equation (1.8). We observe that rational or cognitive issues weight more in the utility value as seen from the coefficients for female byer that usually behave as major house holdings buyers, common average expenses that indicate the level of budget in the purpose, the time of effectiveness for sales. The expected psychic parameters as telephone call have less effect in shifting the utility. By randomly selecting consumer (predictor) values according to the population considered, we see that the value of (1.8) is usually reached, therefore the conclusion herein seem to be a global tendency in the area studied. Remember that those findings are general characteristics because the distribution in this state has been acknowledged as being stationary.

Conclusions

By combination of common analytic tools we have analyzed the behavior of the consumer in a market environment which can be classified as small, heterogeneous and dynamic system. In such particular situation usually common individual techniques can produce fuzzy results and inappropriate findings. Practically herein we concluded that econometric modeling would result a successful analysis for small and heterogonous market mediums if assisted by some statistical tools. Before modeling or applying a known econometric model it is very helpful to analyze the state of the variables and choosing among them most stationary ones. Next, the factorial analysis improve significantly the process of reduction of the dimension of the system by offering a data oriented method to obtain principal factors or ‘basic variables” in the system rather appointing them subjectively or using standard modeling. Herein we used statistical interpretation of the information gathered from direct questionnaires to produce meaningful units for variables, and thereafter we used in those last series in the models. Finally, using simple statistical fit from a minimal or pilot model, we arrived in a deduction which match better to the problem under examination. As result, by combination such simple techniques we were able to identify the most characteristic properties of the consumer behavior or merely the profile of the averaged consumer. It resulted that the income types, the employment, family size, education level age of the householder are the most important variables that define the consumer profile in this case. In econometric approach, the marketing should be based on those most characteristic elements when dealing on the attraction of the consumers. Therefore, analyses of the market, measurements of quantities and statistical study for this system should be better performed on the after-discounts states. Particularly we conclude that the consumer reaction to the discounts was characterized by the increase of spending itself, not only the volumes of items purchased. We identified the load of each factor in the increase of expenditures and acknowledge the utility form in this case.

References

1. Elmira Kushta, Teze doktorature
2. Elmira Kushta, Dode Prenga, Fatmir Memaj. Analysis of consumer behavior in a small size market unit: case study for Vlora District, Albania. IJSRM,2018
3. Mugo Fridah W.,Sampling in research.
4. Jorge Faber and Lilian Martins Fonseca. How sample size influences research outcomes. Dental Press J Orthod. 2014 Jul-Aug; 19(4): 27–29.
5. Ronald Jay Polland. Essentials of survey research and analysis. https://www.psycholosphere.com
6. Scott, David W. Multivariate Density Estimation and Visualization Papers Humboldt-Universitet Berlin, Center for Applied Statistics and Economics (CASE), no. 2004,16.
7. Kalr Jorskog, Arthur Goldbwerg. Estimation of a model with multiple indicator and multiple causes of ingle latent variable. Journal of the American statistical association. Volume 70, issue 351(Sep. 1975)
8. Gene V Glass Percy D. Peckham, James R. Sanders Consequences of failure to meet assumptions underlying the fixed effects analyses of variance and covariance . Review of education research vol. 42, no. 3. 1972
9. Sabir Umarov, Constantino Tsallis, Murray Gell-Mann, Stanly Steinberg. Generalization of symmetric -stable Lévy distributions for q>1. Journal of mathematical physics 51, 033502 2010
10. M.A. Robinson. Quantitative research principles and methods for human-focused research in engineering design. 'Research methods' publications. May 2016. DOI: 10.1007/978-3-319-33781-4_3
11. Jisana T. K. Consumer behaviour models: an overview. Volume 1, Issue 5 (May, 2014)

12. Kwiatkowski, D.& al. (1992)..Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? . Journal of Econometrics 54 (1992) 159-178
13. Kadri G Yilmaz , Sedat Belbag. Prediction of Consumer Behavior Regarding Purchasing Remanufactured Products: A Logistics Regression Model. International Journal of Business and Social Research Volume 06, Issue 02, 2016
14. Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the çase of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. Philosophical Magazine, 50(5), 157-175.
15. Jeff Bray. Consumer Behaviour Theory: Approaches and Models . http://eprints.bournemouth.ac.uk