

**Санкт-Петербургский филиал федерального государственного
автономного образовательного учреждения высшего образования
"Национальный исследовательский университет
"Высшая школа экономики"**

Факультет Санкт-Петербургская школа экономики и менеджмента

Департамент менеджмента

Рабочая программа дисциплины
Прикладные статистические методы в маркетинге
(читается на английском)

для направления 38.04.02. «Менеджмент»
подготовки магистра
для магистерской программы
«Маркетинговые технологии»

Разработчик(и) программы

Антипов Евгений Александрович, к.э.н., доцент департамента менеджмента

Согласована начальником ОСУП

« ____ » _____ 201 г.

Т. И. Видяева _____ [подпись]

Утверждена Академическим советом образовательной программы

« ____ » _____ 201 г., № протокола _____

Академический руководитель образовательной программы

С.Г.Светуных _____

Санкт-Петербург, 201__



1 Область применения и нормативные ссылки

Настоящая рабочая программа дисциплины устанавливает минимальные требования к знаниям и умениям студента, а также определяет содержание и виды учебных занятий и отчетности.

Программа предназначена для преподавателей, ведущих дисциплину «Прикладные статистические методы в маркетинге», учебных ассистентов и студентов направления подготовки 38.04.02. Менеджмент, обучающихся по образовательной программе «Маркетинговые технологии».

Рабочая программа дисциплины разработана в соответствии с:

- Образовательным стандартом НИУ ВШЭ по направлению подготовки 38.04.02 «Менеджмент»
- Образовательной программой «Маркетинговые технологии», по направлению подготовки 38.04.02 «Менеджмент».
- Объединенным учебным планом университета по образовательной программе 38.04.02 «Менеджмент».

2 Цели освоения дисциплины

Целями освоения дисциплины является

- ознакомление студентов с основными методами прикладной статистики, их сильными и слабыми сторонами, а также распространенными маркетинговыми приложениями
- получение студентами практического опыта работы с пакетом SPSS Statistics для решения маркетинговых задач.

3 Компетенции обучающегося, формируемые в результате освоения дисциплины

Уровни формирования компетенций:

РБ - ресурсная база, в основном теоретические и предметные основы (знания, умения)

СД - способы деятельности, составляющие практическое ядро данной компетенции

МЦ - мотивационно-ценностная составляющая, отражает степень осознания ценности компетенции человеком и готовность ее использовать

В результате освоения дисциплины студент осваивает следующие компетенции:



Код	Уровень	Компетенция	Дескрипторы
СК-1	РБ	Способен рефлексировать (оценивать и перерабатывать) освоенные научные методы и способы деятельности	Распознает соответствия между аналитическими задачами и методами
СК-3	СД	Способен к самостоятельному освоению новых методов исследования, изменению научного и научно-производственного профиля своей деятельности	Владеет технологиями поиска информации о новых методах исследования и их приложения в различных областях знаний
СК-6	РБ	Способен анализировать, верифицировать информацию в ходе профессиональной деятельности, при необходимости восполнять и синтезировать недостающую информацию и работать в условиях неопределенности	Применяет полученные знания для решения задач, связанных с анализом реальных данных
ПК27	СД	Способен выбирать и обосновывать инструментальные средства, современные информационные технологии для обработки информации в соответствии с поставленной задачей в сфере управления, анализировать результаты расчетов и обосновывать управленческие рекомендации	Обосновывает выбор инструментальных средств, интерпретирует полученные результаты, оценивает ограничения проведенного анализа, обосновывает управленческие рекомендации

Формы и методы обучения, способствующие формированию и развитию указанных компетенций:

- Решение комплексных аналитических кейсов

Форма контроля уровня сформированности компетенции:

- Задания текущего и промежуточного контроля, связанные с анализом реальных данных



4 Место дисциплины в структуре образовательной программы

Настоящая дисциплина относится к блоку адаптационных дисциплин и не имеет обязательных пререквизитов, кроме владения английским языком (дисциплина читается на английском).

5 Тематический план учебной дисциплины

ОБЪЕМ ДИСЦИПЛИНЫ - 4 зачетных единицы.

Тематический план:

Topic	Total hours	Lecture hours	Computer practice hours	Self-study hours
Data management and pre-processing	12	1	2	9
Descriptive analysis and Custom tables	18	1	4	10
Hypothesis testing : parametric and non-parametric methods	17	1	4	10
Correlation analysis	12	1	4	9
Ordinary least squares regression	12	1	4	9
Logistic regression	12	1	2	9
Classification and regression trees	12	1	2	9
Cluster Analysis	12	1	2	9
Factor analysis	12	1	2	9
Correspondence analysis	12	1	2	9
Multidimensional scaling	12	1	2	8
Conjoint analysis	12	1	2	8
Total	152	12	32	108



6 Формы контроля знаний студентов

Тип контро- ля	Форма контроля	Год 1		Параметры
		1	2	
Текущий	Контрольная работа №1	5		Тесты из 10 вопросов с выбором варианта от- ветов
	Контрольная работа №2		5	
	Домашнее задание №1	4		
	Домашнее задание №2		3	Кейсы по анализу маркетинговых данных
	Домашнее задание №3		4	
Итоговый	Экзамен (Exam)		*	Кейс по анализу маркетинговых данных

7 Критерии оценки знаний, навыков

Оценки по всем формам текущего контроля, а также за экзамен выставляются по 10-ти балльной шкале. Оценка максимально формализована и основана на системе баллов, выставляемых за каждый выполненный пункт задания. Все задания имеют 10 пунктов. Выполненный пункт – 1 балл. Невыполненный или частично выполненный пункт – 0 баллов.

8 Содержание дисциплины

Topic	Total hours	Subtopics (number of hours)	Literature
Data management and pre-processing	12	Measurement scales (3). Variable labels (3). Value labels (3). Variable creation and transformation (3).	1,2
Descriptive analysis and Custom tables	18	Summary statistics for nominal and ordinal variables (6). Frequency distributions (6). Custom tables (6).	1,2
Hypothesis testing : parametric and nonparametric methods	17	T-tests (10). Mann-Whitney test (4). Wilcoxon sign-rank test (3)	1,2,3
Correlation analysis	12	Parametric correlations (6). Nonparametric correlations (6).	1,2,3
Ordinary least squares regression	12	Simple linear regression (4). Multiple linear regression (4). Hypothesis testing (4).	1,6



Logistic regression	12	Binary logistic regression (6). Multinomial logistic regression (6).	3,6,7
Classification and regression trees	12	CHAID, CART, QUEST algorithms (hours distributed evenly)	4,5
Cluster Analysis	12	Hierarchical clustering (6). K-means clustering (6).	2,3
Factor analysis	12	Principal component analysis (6), Maximum likelihood estimation (6)	2,4
Correspondence analysis	12	Simple correspondence analysis (6). Multiple correspondence analysis (6).	3
Multidimensional scaling	12	Multidimensional scaling (12)	4
Conjoint analysis	12	Rating-based conjoint analysis (6). Choice-based conjoint analysis (6).	4,8

9 Образовательные технологии

Лекции и практические занятия проходят в компьютерном классе. Вначале преподаватель дает вводный лекционный материал по теме и демонстрирует приемы работы в пакете IBM SPSS Statistics. После этого студентам дается 80 минут на самостоятельное выполнение заданий аналитического кейса.

Методические рекомендации студентам:

- Перед выполнением контрольных и домашних работ знакомьтесь с конспектами семинаров, выкладываемыми в LMS
- Используйте справочную систему пакета SPSS при выполнении заданий
- Для поиска академических статей используйте Google Scholar в сочетании с электронными ресурсами НИУ ВШЭ

10 Оценочные средства для текущего контроля и аттестации студента

Комплексные аналитические кейсы являются основным средством текущего контроля и аттестации студента. Все они имеют следующую структуру:

- Данные в формате csv или xls
- Описание данных
- Введение в бизнес-проблему



- Измеримая или четко сформулированная цель (например, предсказать значения зависимой переменной с максимальной точностью классификации или получить действенную сегментацию и описать полученные сегменты)

Примеры открытых и закрытых вопросов, используемых в тестах.

1. Explain the purpose of Market basket analysis. Explain what is meant by the support, the confidence and the lift of a market rule. Explain briefly what is Simpson’s Paradox in this context (9 marks)

Market Basket Analysis is a method for association rule induction which aims at finding regularities in the shopping behavior of customers (Ma & Dennis 2004). The purpose of MBA is to determine which products are frequently purchased together and to improve the effectiveness of marketing and sales tactics primarily by stimulating cross- and up-selling (for example, by placing expensive beer next to diapers, if it was found that the probability of buying beer is higher, when diapers are bought).

Let A and B stand for two different sets of items (products). Then an association rule can be written in the form: $A \rightarrow B$ (if itemset A is bought, then itemset B is purchased, too). For example, A – diapers, B – beer (in this case each itemset consists of just one item). Let’s define support, confidence and the lift of a rule:

- **Support**=(Number of transactions with both A and B)/(Total number of transactions)= $P(A \text{ and } B)$
For example, if diapers and beer are bought together in 20% of all the purchases, then Support=0.2
- **Confidence**=(Number of transactions with both A and B)/(Number of purchases with A)= $P(B|A)$
For example, if beer is purchased in 60% of transactions when diapers are bought, then the confidence of the rule $A \rightarrow B$ is 0.6, i.e. the probability of buying a beer given that diapers are purchased is 0.6.
- **Lift**= $P(B|A)/P(B)$
For example, if, on average, the probability of purchasing beer is 10% (i.e. among all transactions, beer is bought in 10% of cases), but the probability increases up to 60% when diapers are bought, then the lift of the rule Diapers \rightarrow Beer is $0.6/0.1=6$. Lift>1 indicates positive dependence between itemsets A and B, while Lift<1 – negative.

Simpson's paradox is a paradox, in which an association but disappears or reverses when non-aggregated data are combined (Ma & Dennis 2004). In the case of market basket analysis, a possible reason for such a problem may be that the analysis is conducted using category-level data instead of sku-level data. For example, the analysis may show that the lift of the rule “Fruit \rightarrow Dairy product” is around 1, indicating that the rule is not very interesting. It is even possible that the lift of the rule is below 1, because there are many vegetarians among those who buy fruits (vegetarians do not eat dairy products or purchase only some of them). However, if we used non-aggregated data, we could find that the rule “Strawberry \rightarrow Whipped cream” stands out with a lift of, say, 3, reminding us that it may be useful to place an expensive brand of whipped cream next to strawberry so that our buyers do not forget to buy whipped cream that goes well with strawberry, but not so often used with apples or oranges (i.e. other fruits). Therefore, the choice of the level of aggregation is very important.



2. Explain the different stages of the SEMMA process and explain how these are used in a nonmarketing application (8 marks)

I will explain SEMMA process using the credit scoring example (Azevedo 2008):

- **Sample.** The time period covered by the sample should be reasonably large to provide a large number of observations (for most data mining techniques – over 1000 cases), but at the same time it should preferably include only recent data for the analysis not to be biased by structural breaks in the financial behavior of bank customers. The sample size also depends on the computing power available to the credit organization (big data technologies allow to work with billions of cases). Random sampling is preferable. The sample with known outcomes (1-default, 0 – no default) should be split into the training and the testing parts. The former is used for model building, while the latter – for model validation.
- **Explore.** At this stage descriptive analysis and visualizations are done. We can look at whether all the variables are in the expected range, detect outliers and suspicious observations and look at pairwise relationship to test hypotheses and explore non-trivial relationships primarily using simple methods (correlation analysis/crosstabs/scatter diagrams). Boxplots, histograms and other visualization tools are useful at this stage. Examples of relationships studied at this stage include analysis of default by gender, race, employment status, etc.
- **Modify.** Variable transformations may be necessary, such as taking logarithms of some of the variables, create squared variables (such as age^2 to account for non-monotonous relationships between age and the probability of default), reduce the number of a categorical variable’s levels (merge rare races with more common once), recode ordinal variables into binary, etc.
- **Model.** In the Model phase we apply classification algorithms to the training data. The dependent variable is whether the person defaulted or not, while the predictors are age, occupation, employment status, income and other variables. Classification methods that can be used are classification trees, Random Forest, logistic regression, discriminant analysis, etc.
- **Assess.** The model should be assessed first using the testing sample and then, later, the model’s performance on new observations should be tracked. The assessment is typically based on classification accuracy tables and ROC curves.

3. Explain the difference between correlation and regression and for each concept give an example when each would be useful to apply (8 marks)

Correlation provides a “unitless” measure of association between 2 variables, ranging from -1 (indicating perfect negative association) to 0 (no association) to $+1$ (perfect positive association). Both variables are treated equally in that neither is considered to be a predictor or an outcome (Crawford 2006). Thus, correlation has nothing to do with causality (i.e. there are no dependent and independent variables in correlation analysis). In contrast to correlation, regression considers one variable to be an outcome and the other to be a predictor variable. In addition, if the correlation is 1 (i.e. linear positive relationship), it does not mean that X has a strong effect on Y . An example of a strict deterministic linear relationship is $Salary = \$10000 + 0.0001 * age$. According to this equation, the salary is strictly related to age, so the correlation is 1 , but the effect is very small: a 1-year increase in age is associated to $\$0.0001$ increase in Salary. Therefore, if we are interested in finding good predictors, then we can start from looking at correlations between each potential predictor and Y and choose those independent variables that are highly correlated with Y as potential regressors. If we are interested in estimating the marginal effect of independent variables on Y , then we need to regress Y on independent variables. In most cases correlation and regression analysis go hand in hand with one another with correlation analysis preceding regression analysis. However, correlation analysis is sometimes self-sufficient in cases when a researcher hypothesizes the existence of positive or negative relationship between certain constructs measured on a scale (e.g. Likert scale), which is just one of the possible ways to operationalize this construct. Another situation in which correlation is more useful than regression is when we want to show that certain variables measure

the same aspect (such as car length and car weight) and thus can be combined in some way to be used further as a measure of this aspect. In such cases computing marginal effects of X on Y or predicting Y based on X may be unnecessary. On the contrary, in cases when we want to be able to predict Y based on the value of X, regression is useful (e.g. predict house price based on the number of bedrooms, house size and lot size).

The remaining part of the assignment is based on the fantasy student account data which is available on blackboard. This data consists of the application forms that students filled in for a student fantasy bank account, the offer that was made to them in terms of the overdraft limit they were allowed, whether they were to be given a credit card and if so what limit, whether they had free insurance of some sort and whether there was a special gift for them joining. The final piece of data is whether they would have accepted such an account. Remember this is fantasy data so students might not have been taking the decision too seriously and thus the results are likely not to be very indicative. You can get the data from LMS eFront.

- 4. The first thing is to describe the information about the students themselves. Pick out three interesting variables and use, histogram, pie chart and descriptive statistics to explain what, if anything is interesting about them. (There will be no marks for describing boring variables, like university attended - why is that boring?).**

Use correlation or any other techniques to describe the relationship between the gender of the student and other information about the student (25 marks)

We have picked up the following student characteristics: gender, marital status and their interests. It appeared that 52% of respondents are females, i.e. the sample is balanced with respect to gender. Most of respondents are single (75%), which makes it difficult to study the impact of each marital status on outcome variables. However, it will still be possible to recode the variable into single (1) vs. any other marital status (0) to see if experience of relationships impacts account acceptance.

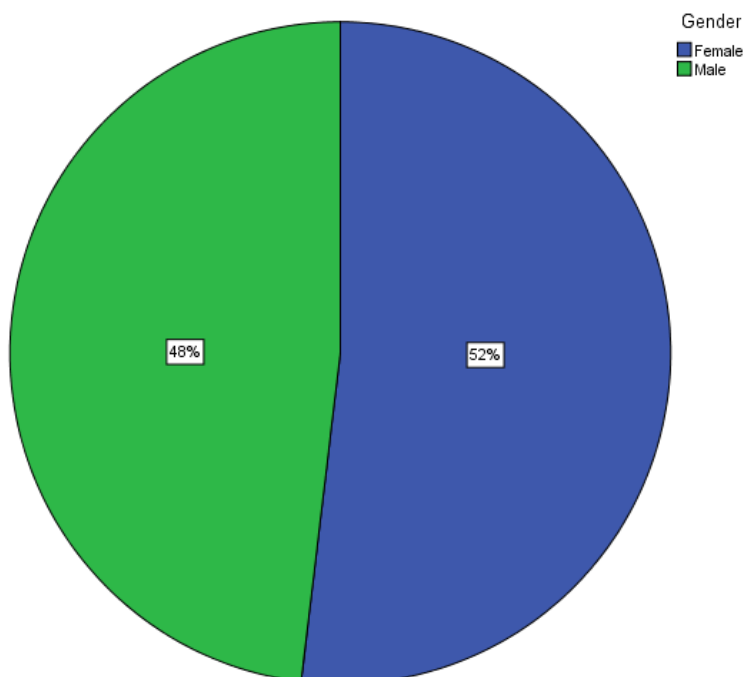


Figure 1. Gender: pie chart

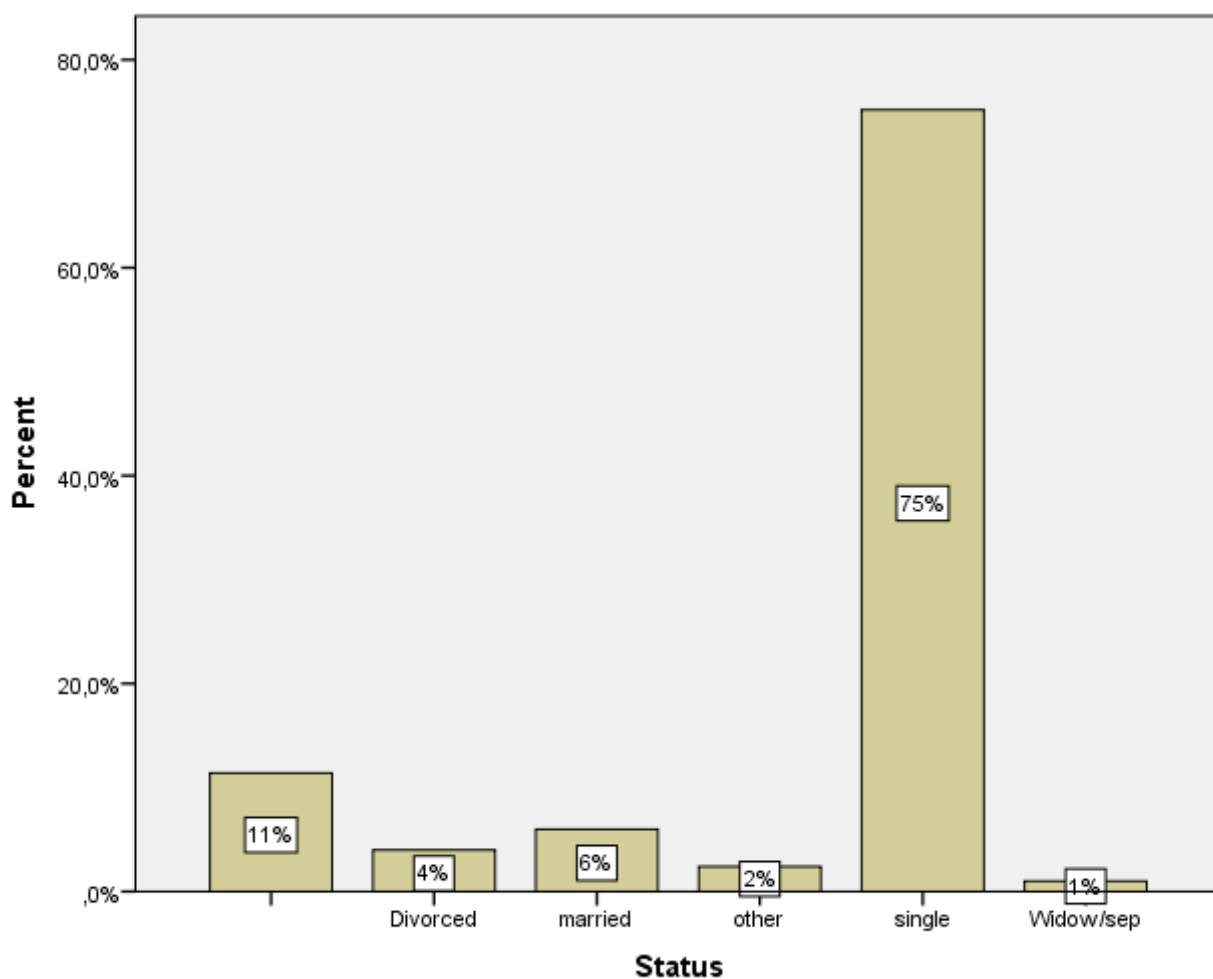


Figure 2. Bar chart: marital status

Share of respondents that mentioned each of the fields of interest

	Column N %
Music	62.8%
Sport	53.6%
Cinema	52.6%
Travel	49.4%
Clubbing	45.2%
Cars	25.6%
Beer	24.8%
DIY	9.8%
CountryWest	6.4%
ern	
Gardening	5.8%

In order to compare males and females we have used independent-samples t-test (to compare means) and z-test (to compare proportions). Significantly higher percentages (at 5% significance level) are highlighted: females are more interested in travelling, while males – in cars, beer and DIY.





Consider a supermarket (e.g., United Supermarkets) that must make weekly shipments of a certain product from two different warehouses to four different store [25 points, 5 points for each]

- a. How could a quantitative approach to decision making be used to solve this problem?
- b. What would be the uncontrollable inputs for which data must be gathered?
- c. What would be the decision variables of the mathematical model? the objective function? the constraints?
- d. Is the model deterministic or stochastic?
- e. Suggest assumptions that could be made to simplify the model.

Quantitative and Qualitative analysis

- a. Explain at least three differences between quantitative and qualitative analysis from the manager's point of view. [9 points]
- b. Please provide one business problem which requires quantitative and qualitative analysis respectively. (i.e., one for quantitative analysis and one for qualitative analysis) [6 points]
- c. Can quantitative and qualitative analysis be used together to solve a business problem? If so, please provide an example. [5 points]
- d. Please provide at least three reasons why we need quantitative analysis. Explain each of them briefly [5 points]

Пример аналитического кейса

The Students, while proceeding with their assignment, are to rehearse and demonstrate skills in selecting, implementing and interpreting quantitative analyses as well as deriving implications thereof to a given problem solving. It must have a maximum length of 25 pages. The term project concerns customer behavior and store loyalty on the FMCG (Fast-moving consumer goods)” market. The project has two phases. The first phase consists in sketching a questionnaire outline proposition, while in the second phase students are expected to deliver a quantitative analysis report. Both phases should answer to one or more decision problems of your choice and a number of research questions as well. (see *Marketing Decision Problems* and *Marketing Research Problem as referred to* by Malhotra, Birks & Wills, page 55). MDP and MRP shouldn’t necessarily be identical for both phases of the assignment.

As a future expert in analysis of quantitative survey data, you have been approached by a decision-maker operating in the FMCG (select one of your choice). The decision maker wants you to:

- **Specify the decision problems with its associated research questions and compile a coherent questionnaire** (show/explain 1-2 pages for the underlying operationalization) that answers the se-



lected survey questions. The Decision problem, the Research questions, the explanation of the operationalization as well as the the questionnaire are to be uploaded.

- **An Analysis Report** that, within the topic “FMCG (Fast-moving consumer goods) Store Loyalty” and based on a questionnaire (attached) analyze one or more decision problems and issues of your choice. Decision problems and research questions does not need, as mentioned above, to be identical with those which led to the group's questionnaire outline proposition. You may need to revise in line with the uploaded questionnaire and datafile in order to solve your decision problem and research questions

The Analysis Report must indispensably include the following parts (page numbers in parentheses are only indicative)

Introductory part (ca. 2 pages)

- Explain the MDP and the associated research questions. Specify what information is needed to answer those questions. Show, either in this part or in the group's conclusion, how the selected research questions contribute to the chosen decisions problem solving.

Measurement part (ca. 4 page)

- Show how the group’s need for information can be operationalized from the questionnaire and datavariabes. Evaluate (positive and negative criticism) the content and measurements in the questionnaire as well as data collection methods and response in the uploaded survey (mainly reliability and validity issues). It is though sufficient to limit your assessments to the individual questions involved in your analysis (see Analysis plan).

Analysis plan (ca. 1 page)

- Set up an analysis plan that outlines the analysis methods used to answer the above research questions. The group can freely set up the specific content of the analysis plan, as long as they make sure it involves at least one analysis of each of the following analytical techniques: representativity test (nonparametric chi-square test) , frequency table, crosstable (including chi-square test), bar or pie chart, t-test, ANOVA, correlation, multiple regression , factor analysis, reliability test on the composite scale (alpha analysis), cluster and discriminant analysis

Analysis part (ca. 17 pages)

- Download the data file Survey Data FO2016. Perform and interpret the analysis. To demonstrate interpretation skills of analysis results, reference should be made in the report to the relevant parts of the output and at least one analysis output is to be referred to for each of the above (Analysis plan) mentioned analysis. These outputs can be included in the main text or as an appendix..

Conclusion (ca. 1 page)



- Highlight the report’s "significant" results and specify its conclusions’ importance in relation to the assignment.

The questionnaire and its operationalization must be attached to the report!

11 Порядок формирования оценок по дисциплине

- Результирующая оценка=0.3*Оценка за экзамен+0.7*Накопленная оценка
- Накопленная оценка=0.2*(Оценка за д.з 1+Оценка за д.з 2+Оценка за д.з. 3)+0.05*(Оценка за к.р 1+Оценка за к.р. 2)+0.3*Оценка за работу на семинарах

12 Учебно-методическое и информационное обеспечение дисциплины

12.1 Основная литература

1. Bryman A., Bell E. Business research methods. – Oxford University Press, USA, 2015.
2. Zikmund W. G. et al. Business research methods. – Cengage Learning, 2013.
3. Fox J. Applied regression analysis and generalized linear models. – Sage Publications, 2015.
4. Sachs L. Applied statistics: a handbook of techniques. – Springer Science & Business Media, 2012.

12.2 Дополнительная литература

5. Janssens W., De Pelsmacker P., Van Kenhove P. Marketing research with SPSS. – Pearson Education, 2008.
6. Malhotra N. K. Marketing research: An applied orientation, 5/e. – Pearson Education India, 2008.
7. Mazzocchi M. Statistics for marketing and consumer research. – Sage, 2008.
8. Rossi P. E., Allenby G. M. Bayesian statistics and marketing //Marketing Science. – 2003. – Т. 22. – №. 3. – С. 304-328.

12.2 Дистанционная поддержка дисциплины

Материалы дисциплины размещаются в LMS eFront

13 Материально-техническое обеспечение дисциплины

- PCs with IBM SPSS statistics installed on them
- VNC software for screen sharing.

14 Особенности организации обучения для лиц с ограниченными возможностями здоровья

В случае необходимости, обучающимся из числа лиц с ограниченными возможностями здоровья (по заявлению обучающегося) могут предлагаться следующие варианты восприятия учебной информации с учетом их индивидуальных психофизических



особенностей, в том числе с применением электронного обучения и дистанционных технологий:

1) *для лиц с нарушениями зрения*: в печатной форме увеличенным шрифтом; в форме электронного документа; в форме аудиофайла (перевод учебных материалов в аудиоформат); индивидуальные консультации с привлечением тифлосурдопереводчика; индивидуальные задания и консультации.

2) *для лиц с нарушениями слуха*: в печатной форме; в форме электронного документа; видеоматериалы с субтитрами; индивидуальные консультации с привлечением сурдопереводчика; индивидуальные задания и консультации.

3) *для лиц с нарушениями опорно-двигательного аппарата*: в печатной форме; в форме электронного документа; в форме аудиофайла; индивидуальные задания и консультации.