MaxDiff TEST analysis
(an example)

Danijel Kopčinović, IT Market


MAXDIFF TEST

U upitniku "MAXDIFF TEST" postavit ćemo Vam X pitanja s raznim varijantama, a Vi ćete u svakom pitanju odabrati Vama najbolju i najlošiju varijantu. Za ispunjavanje cijelog upitnika po našoj je procjeni dovoljno 15-20 minuta.

Kao znak zahvale za sudjelovanje u ovom upitniku, Y će Vam poslati mali poklon paket!

U ovom upitniku koristimo sljedeće osobne podatke:

- adresa e-pošte ispitanika
- IP adresa

Navedene osobne podatke koristit ćemo u svrhu identifikacije ispitanika i kreiranje profila ispitanika. Navedene podatke nećemo koristiti niti u koju drugu svrhu i nećemo ih dijeliti niti s kojom trećom stranom bez pismene privole ispitanika.

Više o politici zaštite osobnih podataka tvrtke Y pogledajte ovdje.

Prihvaćam pravila o zaštiti osobnih podataka ☐

Introduction

In this document we will present a test market research with the MaxDiff method. The research is based on the (assumed) wish of a local pizzeria to find out which new pizzas would its buyers prefer to see in its offering.

New pizzas that the respondents choose from are:

```
##  [1] miješana        povrtna         pikantna        4 vrste sira
##  [5] s tunom         ribarska        4 godišnja doba calzone
##  [9] slavonska       rukola/pršut    bolonjez        lovačka
## [13] losos           sa salamom
## 14 Levels: 4 godišnja doba 4 vrste sira bolonjez calzone losos ... slavonska
```

Except the main products (pizzas) that are in our focus, we will also take into account a possible influence of the "demographic" factors to the answers/results:

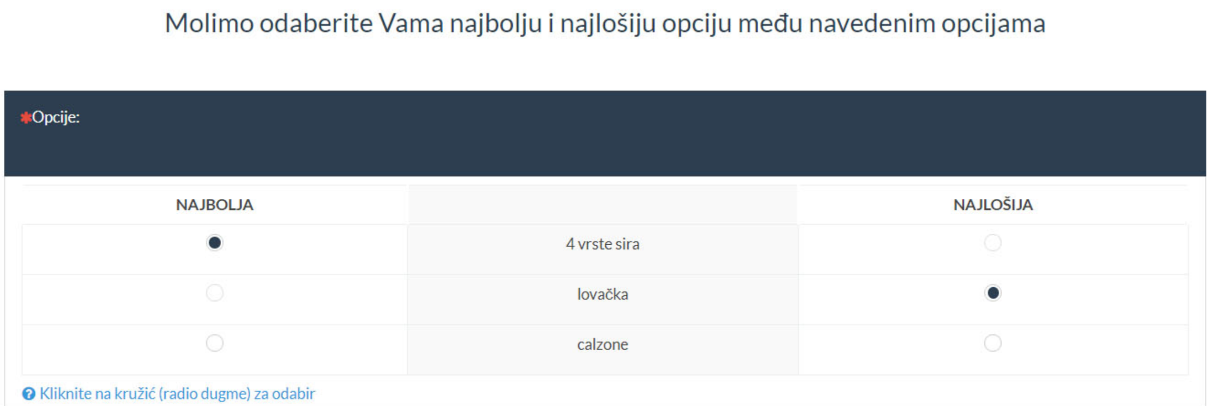
```
## $lokacija
## $lokacija$naziv
## [1] "lokacija"
##
## $lokacija$vrijednosti
## [1] "blizu pizzerije"     "daleko od pizzerije"
```

In a MaxDiff survey respondents choose the best and the worst product among a few alternatives. Let's have a look at some products combinations that we posed as questions.

```
##   task                                items best_choice worst_choice
## 1    1     4 vrste sira,slavonska,sa salamom           ?            ?
## 2    2            calzone,miješana,slavonska           ?            ?
## 3    3 4 godišnja doba,4 vrste sira,ribarska           ?            ?
## 4    4             bolonjez,calzone,ribarska           ?            ?
## 5    5              slavonska,ribarska,losos           ?            ?
```

Out of the set of all 364 products combinations with 3 alternatives (number of alternatives in a question is set as a design parameter), we have chosen 54 representative ones. With the chosen combinations it is possible to calculate (average or individual) value of each product for the respondents.

All combinations were divided to 3 questionnaires, each with 6 questions, each with 3 options/alternatives.

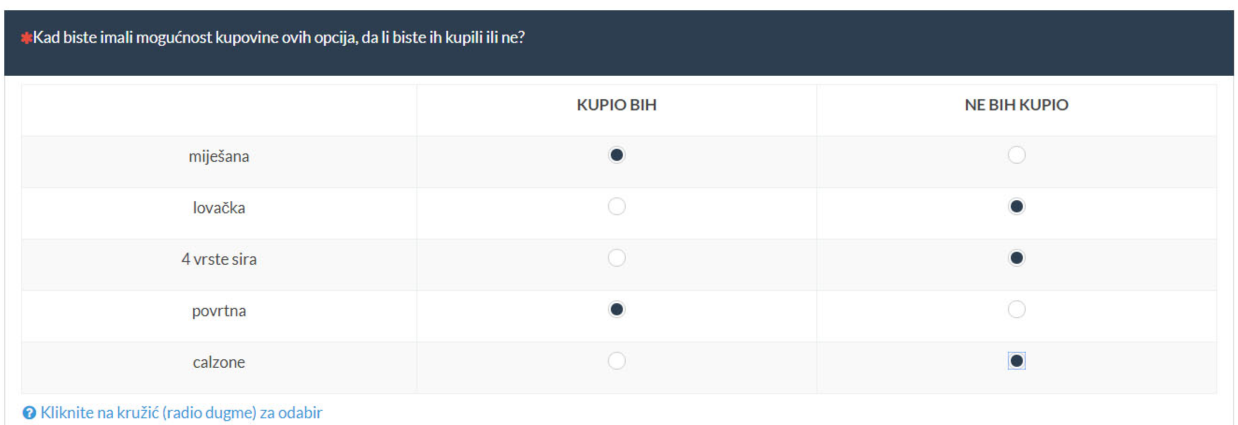

Molimo odaberite Vama najbolju i najlošiju opciju među navedenim opcijama

*Opcije:

NAJBOLJA		NAJLOŠIJA
●	4 vrste sira	○
○	lovačka	●
○	calzone	○

Kliknite na kružić (radio dugme) za odabir

For a better estimation of the values of different products, we added a question in which a respondent expresses general positive or negative tendency to a few given products. The products given in this question are chosen related to the previous respondent's choices, so that a span, from the pizzas most often chosen as best to the pizzas most often chosen as worst, is given. In this question products are not compared with each other but all against the so called threshold option, which can mean different things in different contexts, but generally means the point between want/don't want, would buy/wouldn't buy, am interested/am not interested etc. This question helps us to better level the answers of all the respondents against the threshold option and thus against each other, giving us better and more relevant product values estimates.


*Kad biste imali mogućnost kupovine ovih opcija, da li biste ih kupili ili ne?

	KUPIO BIH	NE BIH KUPIO
miješana	●	○
lovačka	○	●
4 vrste sira	○	●
povrtna	●	○
calzone	○	●

Kliknite na kružić (radio dugme) za odabir

Survey results were computer generated, with certain random parameters, for 200 respondents.

Simple analysis - counting

There are multiple ways to analyze the data with a MaxDiff method. The simplest approach is to count how many times some product was chosen as the best and as the worst. Here is the list of the products from our survey with the number of times chosen as best/worst:

alt	chosen_as_best	chosen_as_worst
4 godišnja doba	161	20
4 vrste sira	164	36
bolonjez	7	175
calzone	18	148
losos	130	43
lovačka	24	154
miješana	70	19
pikantna	**229**	**3**
povrtna	115	61
ribarska	129	47
rukola/pršut	**7**	**215**
s tunom	52	68
sa salamom	56	59
slavonska	38	152

We can see that pikantna was chosen most times as the best option, and rukola/pršut most times as the worst option.

Multinomial logit model analysis

More advanced analysis includes data (respondents' choices) modelling and then calculating the coefficients (values) for each product according to the multinomial logit model. After applying this analysis to our data, we get the following summary:

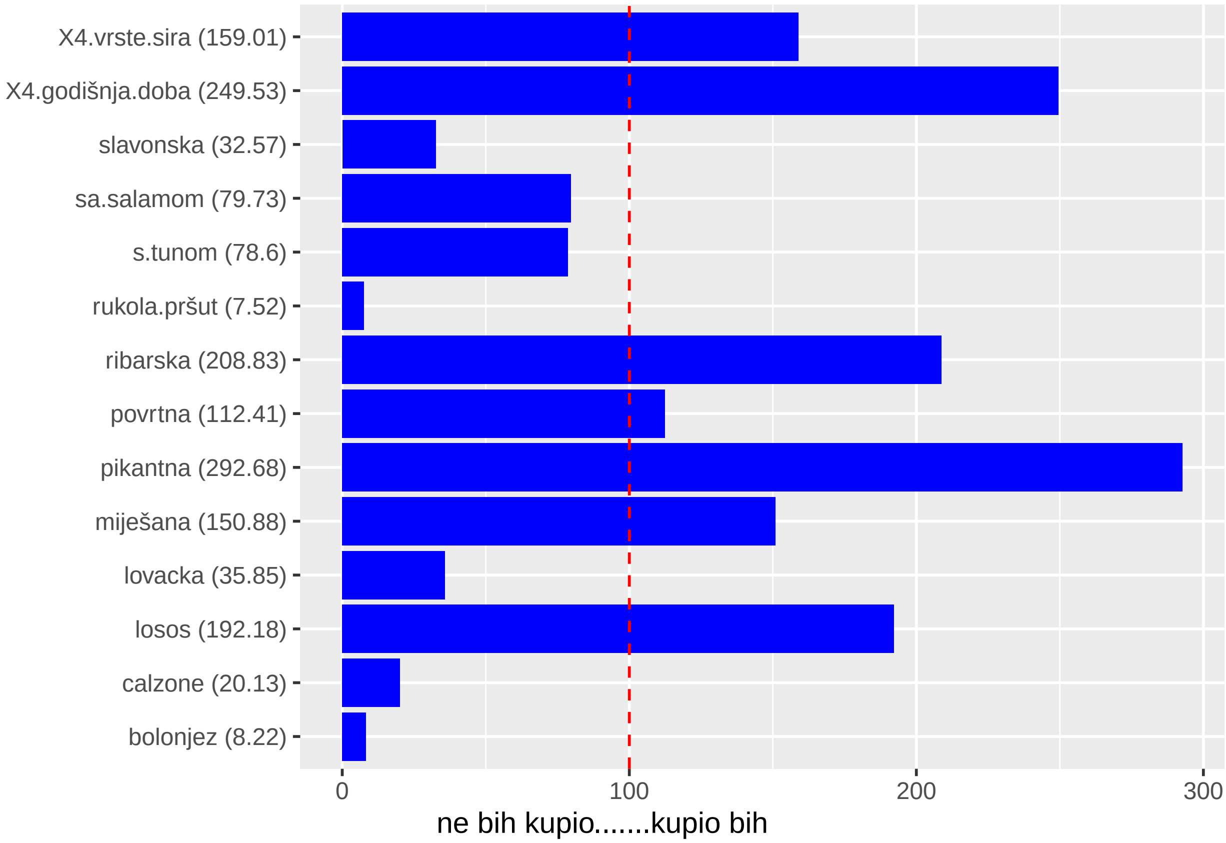
```
##                      1st Qu.       Mean    3rd Qu.
## miješana         -0.888382307  0.2349791  1.3505137
## povrtna          -1.180309157 -0.2887543  0.6152998
## pikantna          2.700681397  3.9116654  5.0833724
## X4.vrste.sira    -0.599027521  0.3436022  1.2806869
## s.tunom          -1.906194196 -0.8122667  0.2935417
## ribarska         -0.009438742  1.0520189  2.1115724
## X4.godišnja.doba  0.596211851  1.8214363  3.0371697
## calzone          -3.448424673 -2.4087209 -1.3417371
## slavonska        -2.991635218 -1.8822479 -0.7353968
## rukola.pršut     -4.475042336 -3.4380242 -2.3552538
## bolonjez         -4.468180912 -3.3461902 -2.1785240
## lovačka          -2.751588330 -1.7738617 -0.7688110
## losos            -0.206722580  0.8012456  1.8071983
## sa.salamom       -1.744238546 -0.7930006  0.1748956
```

The given summary contains a lot of information which is not so easy to interpret directly. But to clarify a bit, here are a few explanations derived directly from the summary:

- 'calzone' has a negative (mean) coefficient -2.409, which means that the respondents value it less than the basic threshold option representing the would/would not buy point (coefficients are always calculated with respect to the first/basic declared option, which is in this case the threshold option) - to put it simply, respondents would not buy 'calzone', and due to a relatively big negative coefficient, this decision is pretty strong,
- 'pikantna' and 'ribarska' both have positive (mean) coefficients (3.912, 1.052) which means that the respondent value both of them more than the basic threshold option, but 'pikantna' is valued more than 'ribarska' - to put it simply, respondents would buy both of them, but 'pikantna' rather than 'ribarska',

- the first and third quantiles define the span containing (this is somewhat simplified) each coefficient with the 75% probability.

If we normalize and scale all the coefficients so that the threshold option (would/would not buy) maps to the number 100, it is easier to interpret their values and corresponding ratios:



Now we can see that the most valued product is pikantna and that its value is 1.84 times more than the value of X4.vrste.sira.

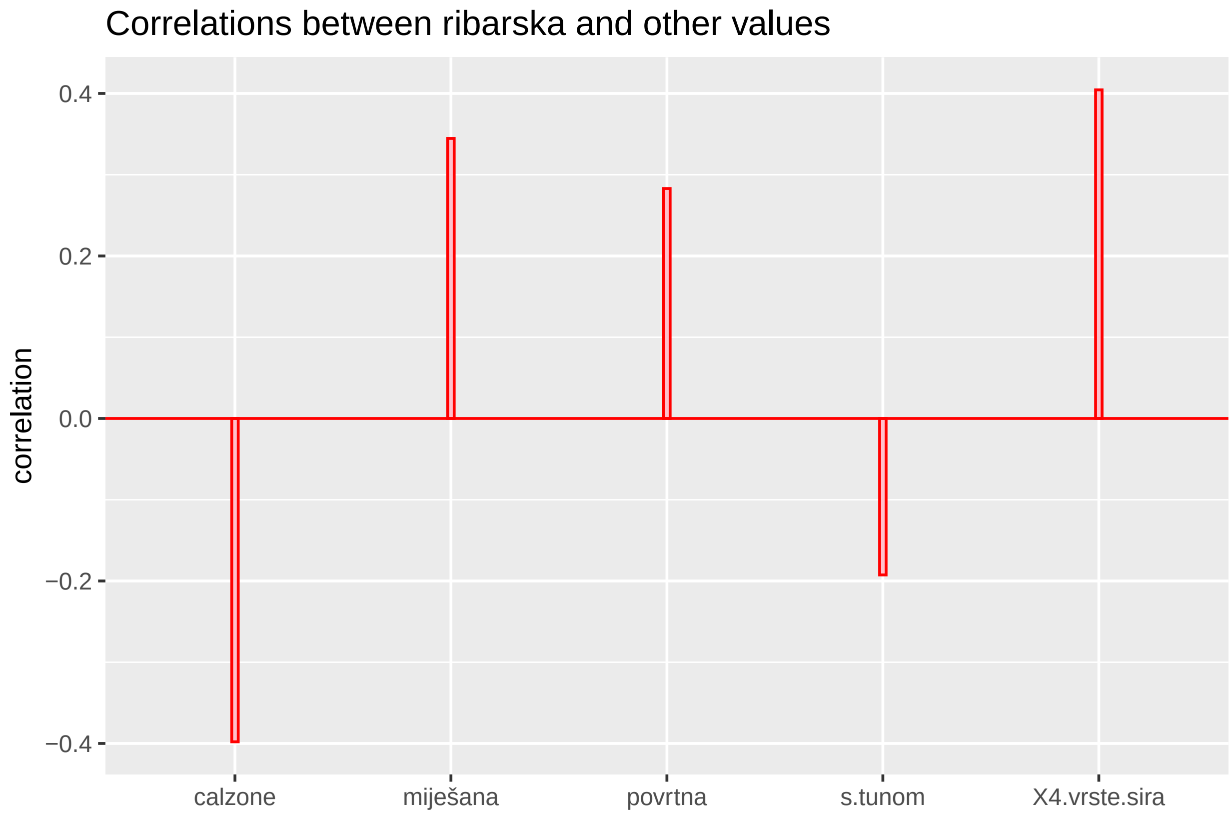
These results show the main advantages of the MaxDiff analysis compared to a simple product "ordering" from the best to the worst which is sometimes (too often :)) used for business analyses. MaxDiff shows us:

- *relations between products* - how much is each product more or less valuable to the respondents,
- product positions *related to (key!) point marking the positive/negative attitude* - would or would not buy, am or am not interested...

This way we get a complete picture of all the products from the respondents' perspective and we can make quality business decisions.

Products correlation analysis

Another element that we can analyze is the correlation between the products/products values.



For example, in the graph shown above, we can see that *ribarska* is positively correlated with *miješana, povrtna, X4.vrste.sira*, and negatively correlated with *s.tunom, calzone.* This means that if respondents put more value to the *ribarska* pizza, then they put more value also to *miješana, povrtna, X4.vrste.sira*, while they put less value to *s.tunom, calzone.*

Results of the analysis and coefficients estimation is the easiest to understand with the share prediction calculation, which we will now show.

Share prediction

Once we have all the product coefficients estimates, we can compare different product combinations and predict the sales shares. Predicted sales shares are probabilities that a certain product will be chosen/sold compared to other products in the tested combination.

As an example we will pick a combination of a few randomly selected products and calculate their predicted sales shares:

```
##          [,1]       [,2]       [,3]      [,4]           [,5]
## opcije  "miješana" "pikantna" "povrtna" "4 vrste sira" "s tunom"
## share % "12.43"    "45.54"    "12.02"   "14.00"        "16.02"
```

In the table above we can see how would respondents choose between 5 different pizzas. Each pizza gets the predicted share (percentage) in the total sales, according to the corresponding coefficients estimates.

Most often chosen option (among the given pizzas, not in general) is *pikantna* **with the predicted sales share 45.54%**.

Checking the model on the existing data

With the share predictions for different products combinations, we can also test how would our model predict the choices for (some) of the combinations given in the survey:

```
##                     choice
## respondent/question true best predicted best true worst predicted worst
##                28/5         2              2          1               3
##               185/6         2              2          3               3
##               168/1         1              1          2               2
##                68/5         2              2          1               3
##               146/6         2              2          1               1
##               176/6         2              2          3               3
##                86/1         1              1          3               3
##                84/5         3              3          1               1
##                10/1         2              2          3               3
##                87/5         2              2          1               1
##               153/1         2              2          1               1
##               158/1         1              1          3               2
##                28/6         3              3          1               2
##                44/1         2              2          1               1
##                42/6         2              2          1               1
##               179/5         2              2          1               1
##                28/1         2              2          1               1
##               112/5         2              2          3               3
##                68/1         2              2          1               1
##                44/6         3              3          2               2
```

```
## [1] "Overall correct predictions:90%"
```

The overall prediction correctness gives us confidence that our model works well.

Respondents segmentation

Taking into account different factors (demographic or other), we can divide the respondents into segments and then model how each segment (instead of average or individual) values the products. The number of segments we believe exists in the respondents set must be given in advance as a parameter. Assuming that there are 3 different respondents segments we build the multinomial logit model with LCA (Latent Class Analysis) and get the following summary:

```
##
##                      segment_1   segment_2  segment_3
##   miješana           0.1148834  0.09101120 -0.3130916
##   povrtna           -0.1178750 -0.28466952 -0.3599366
##   pikantna           2.4660650  1.47187209  1.4795165
##   X4.vrste.sira      0.2263574  0.06130955 -0.3630211
##   s.tunom           -0.4368493  0.57051495  0.5078137
##   ribarska           0.6289285  0.60133842  0.6285399
##   X4.godišnja.doba   1.0681058  1.06660231  1.0389167
##   calzone           -1.5812423 -1.12379359 -0.7151136
##   slavonska         -1.1868617 -0.72700958 -1.0868358
##   rukola.pršut      -2.3095288 -1.76848126 -1.3268809
##   bolonjez          -2.1264829 -1.47366018 -2.2016732
##   lovačka           -1.0616948 -0.99635986 -0.8848496
##   losos              0.4854090  0.56463342  0.5819652
```

```
##   sa.salamom       -0.6502203 -0.02659289 -0.5247803
```

We can see that the segments differ (for more than 0.5) in valuation of e.g. *pikantna, s.tunom, rukola.pršut, bolonjez, sa.salamom.*

Since one of the goals of the segmentation is assigning each respondent to some segment (with the highest probability), let's see this result too (the numbers given are the numbers of the respondents belonging to a certain segment):

```
## [1] "Segment 1"
##   [1]   1   4   5   6  10  12  15  20  23  24  25  26  27  36  37  38  39  40
##  [19]  41  42  44  46  48  50  51  52  53  55  57  58  59  62  67  68  69  70
##  [37]  72  73  77  79  80  82  83  84  90  93  95  96  99 100 101 102 105 109
##  [55] 110 111 116 117 119 124 126 128 129 130 131 135 137 138 139 140 142 145
##  [73] 147 148 149 150 151 153 155 156 157 158 160 162 163 165 169 172 173 174
##  [91] 178 179 181 182 183 185 186 189 190 191 194 196 197 199
## [1] "Segment 2"
##  [1]   2   3   7   8   9  11  13  14  16  17  18  19  21  22  28  29  30  31  32
## [20]  33  34  35  43  45  47  49  54  56  60  61  63  64  65  66  71  74  75  76
## [39]  78  81  85  86  87  88  89  91  92  94  97  98 103 104 106 107 108 112 113
## [58] 114 115 118 120 121 122 123 125 127 132 133 134 136 141 143 144 146 152 154
## [77] 159 161 164 166 167 168 170 171 175 176 177 180 184 187 188 192 193 195 198
## [96] 200
## [1] "Segment 3"
## integer(0)
```

Since one of the segments is empty (no respondents in it), we conclude that the number of segments that we can identify with the given data and model is 2.

With the given segmentation we can e.g. create different products that will be the most interesting (best valued) by each segment, thus increasing buyers satisfaction and income/profit.

TURF analysis

With an offer that contains more products we will increase the probability that at least some product will be sold. To calculate which products should be offered to achieve the highest number of sales, we can use TURF (total unduplicated reach and frequency) analysis.

Let's assume that we want to put 3 products in an offer. For each combination/offer containing 3 products and for each respondent from our survey, we will calculate whether he would buy at least one product from the offer (reach - is he 'reached' with at least one product) and how many (if any) products from the offer would he buy (frequency - how many times is he 'reached' with the offer). The respondent's buying decision for a product is based on the probability of sales for that product compared with the other products in the offer and the basic threshold option marking the buying/not buying boundary.

Let's have a look at a few rows from the reach & frequency table for our products and offers (ordered decreasing by reach):

```
## # A tibble: 6 x 3
##   combination                        reach  freq
##   <chr>                              <dbl> <dbl>
## 1 pikantna, rukola/pršut, bolonjez     200   200
## 2 pikantna, calzone, slavonska         197   199
## 3 pikantna, slavonska, lovačka         196   197
## 4 pikantna, slavonska, rukola/pršut    196   197
## 5 pikantna, lovačka, sa salamom        195   200
## 6 pikantna, s tunom, slavonska         195   199
```

We can see that the offer pikantna, rukola/pršut, bolonjez would reach (at least one product would be bought by) the most respondents, 200 of them, which is 100% of the total number of respondents, with a total of 200 (potential) sales.

This way we can build an offer of products that will maximize the sales.

Conclusion

MaxDiff analysis is a powerful tool for gaining knowledge of your customers and their values and creating the best products for them. This approach always leads to the business improvement, higher incomes and higher customer satisfaction.

For more information contact **danijel.kopcinovic@itmarket.hr**.