

# Using Design of Experiment to Get The Best Ořechovka

Data Show 2019

Vojtěch FILIPEC



# whoami



nuclear physicist



professional bio:

- banks: applied statistics in credit risk: model development, monitoring, automation
- fintech, pharma, FMCG: consultancy in analytics



data scientist



# contents

## my project

1. ořechovka:
  - a. what is it
  - b. problem: many recipes
2. design of experiments
  - a. what is it
  - b. how to in four steps
3. optimal recipe

goal: a new useful method in your toolbox

## in general

1. the same problem elsewhere
2. what is DoE good for
3. data analysis

**What is ořechovka and what is the problem here?**



# What is ořechovka?



semi-sweet liqueur

simple basic recipe

many variations

the best one?

[illegible]



# Problem: Many variations

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	source	# nuts	alcohol	alcohol	alcohol	extraction	extraction	sirop		% alc. after mixing						
2			[l]	[%]	form	[weeks]	method	sugar [g]	water[g]		total sugar	clove [ks]	cinnamon	muscat (nut/blo	lemon peel	other spices
3	<a href="https://www.apetitonline">https://www.apetitonline</a>	24 ks	1	50	vodka, vínovice, t	6	s kusy koreni	500	500	28,2	1500	3	ano			
4	<a href="https://www.idnes.cz/hk">https://www.idnes.cz/hk</a>	1 kg	1,33586956	30	1l rum+ 200 ml H rada		jen palenka nebo vc. sirupu	250	500	20,3	1000	5	spetka	spetka	1 lzice, kandovana	
5	<a href="https://prima-receptar.c">https://prima-receptar.c</a>	tři hrsti	1	40	vodka-režná-slivo	3-4	jen palenka ("zalud. liker")	250	500	24,5	750	ano	ano	ano	ano	bady an
6	<a href="https://abece-daz-ahrad">https://abece-daz-ahrad</a>	15 ks	1	40	vodka	6	s kusy koreni	500	500	22,6	1500	4	cela		z 1/2 citronu	3 jalovce
7	<a href="https://www.toprecepty">https://www.toprecepty</a>	24 ks	1	40	konak, slivovice	6	s kusy koreni	500	500	22,6	1500	15 g	4 g			
8	<a href="https://www.receptyoni">https://www.receptyoni</a>	20 ks	1	40	zitna	3	s kusy koreni bez vanilky	200	500	24,9	600	1	kousek			5 jalovce, vanilka 15
9	<a href="https://www.iaktak.cz">https://www.iaktak.cz</a>	20 ks	1	80	lih	4	s kusy koreni i sirupem	750	1250	30,1	900	5 g	2 ks rozlamat		1/2 balicku pomé rozkrajeny citron	po
10	<a href="https://www.iaktak.cz/k">https://www.iaktak.cz/k</a>	doji macerace														
11																

Important ingredients are known

Looking for optimised combination

What combinations shall we try?



# Problem: Many variations

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	source	# nuts	alcohol	alcohol	alcohol	extraction	extraction	sirop		% alc. after mixing						
2			[l]	[%]	form	[weeks]	method	sugar [g]	water[g]		total sugar	clove [ks]	cinnamon	muscat (nut/blo	lemon peel	other spices
3	<a href="https://www.apetitonline">https://www.apetitonline</a>	24 ks	1	50	vodka, vínovice, t	6	s kusy koreni	500	500	28,2	1500	3	ano			
4	<a href="https://www.idnes.cz/hk">https://www.idnes.cz/hk</a>	1 kg	1,33586956	30	1l rum+ 200 ml H rada		jen palenka nebo vc. sirupu	250	500	20,3	1000	5	spetka	spetka	1 lzice, kandovana	
5	<a href="https://prima-receptar.c">https://prima-receptar.c</a>	tři hrsti	1	40	vodka-režná-slivo	3-4	jen palenka ("zalud. liker")	250	500	24,5	750	ano	ano	ano	ano	bady an
6	<a href="https://abece-daz-ahrad">https://abece-daz-ahrad</a>	15 ks	1	40	vodka	6	s kusy koreni	500	500	22,6	1500	4	cela		z 1/2 citronu	3 jalovce
7	<a href="https://www.toprecepty">https://www.toprecepty</a>	24 ks	1	40	konak, slivovice	6	s kusy koreni	500	500	22,6	1500	15 g	4 g			
8	<a href="https://www.recepty.onli">https://www.recepty.onli</a>	20 ks	1	40	zitna	3	s kusy koreni bez vanilky	200	500	24,9	600	1	kousek			5 jalovce, vanilka 15
9	<a href="https://www.iaktak.cz">https://www.iaktak.cz</a>	20 ks	1	80	lih	4	s kusy koreni i sirupem	750	1250	30,1	900	5 g	2 ks rozlamat		1/2 balicku pomer rozkrajeny citron	po
10	<a href="https://www.iaktak.cz/k">https://www.iaktak.cz/k</a>	doji macerace														
11																

Important ingredients are known

(crucial)

Looking for optimised combination

...and process description

What combinations shall we try?

naive approach: 34560 options



# Design of Experiments

What combinations shall we try?



# Design of Experiments (DoE)

recall: seek optimal combination of known inputs

designed experiments → understand the process → optimise the inputs

output: quantified process description

cause-effect relation

*optimal* DoE: stronger assumptions for reduced number of runs



# How-to in four steps

1. describe

$$y = \beta_0 + \beta_1x + \beta_2x^2 + \beta_3m + \beta_4m^2 + \beta_5xm + \beta_6v + \beta_7d + \epsilon$$

2. generate designs



3. run



4. analyse





# 1. Describe + 2. Generate designs

inputs:

- reduction:  $12 \rightarrow 4$
- min's + max's
- assumed relations

input	symbol	unit	min	max
walnuts	<b>x</b>	g	330	550
sugar	<b>m</b>	g	600	1500
initial water	<b>v</b>	g	0	851
extraction	<b>d</b>	week	3	6



# 1. Describe + 2. Generate designs

inputs:

- reduction:  $12 \rightarrow 4$
- min's + max's
- assumed relations

designs:

- types of designs
- criteria
- software: R + SAS

input	symbol	unit	min	max
walnuts	<b>x</b>	g	330	550
sugar	<b>m</b>	g	600	1500
initial water	<b>v</b>	g	0	851
extraction	<b>d</b>	week	3	6

### 3. Run



## 4. Analyse



objective: taste rank (ordinal)

surprisingly hard task

univariate objective

details: [blog](#)

# The Best Recipe







# Final - and Ultimate - Recipe

ingredients:

- 550 g green walnuts (picked at June 24)
- 1 liter of 80 % alcohol
- 600 g plain sugar
- spices
- water to dilute the alcohol

steps:

- soak nuts (cleansed, sliced) to the alcohol;
- keep the jar lid untight at a sunny place, stir occasionally
- strain the nuts after 3-6 weeks
- add the spices and the sirup: put the water on to boil, dissolve the sugar, let it cool,
- let it ferment; the longer, the better

# Thank You

 [linkedin.com/in/vojtech-filipek/](https://www.linkedin.com/in/vojtech-filipek/)

 [github.com/vojtech-filipek](https://github.com/vojtech-filipek)



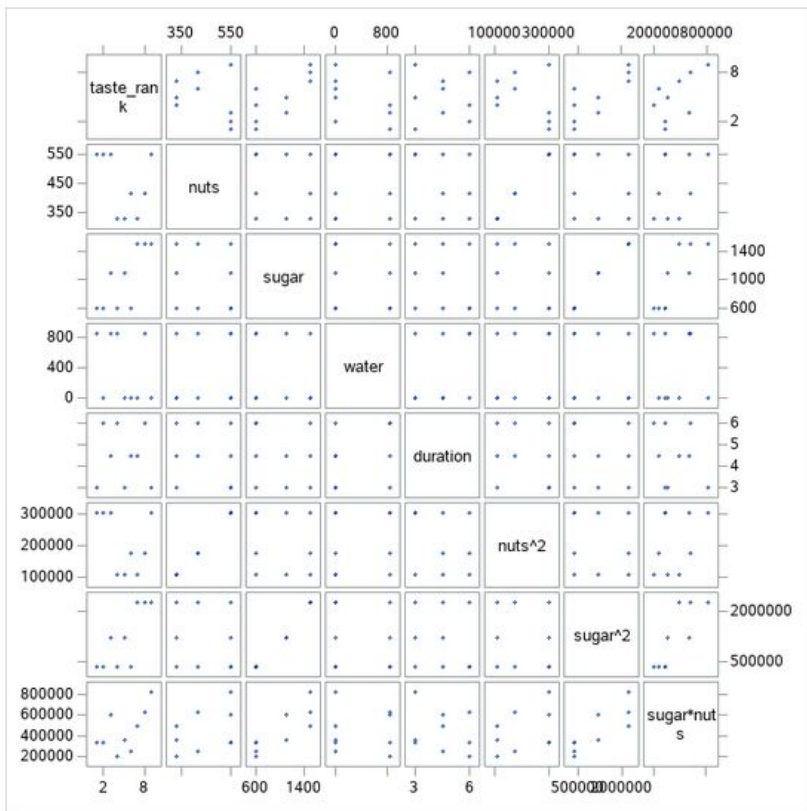
credits: <https://www.festisite.com/posters/uncle-sam-wants-you/>



## Additional Sources

- My blog about ořechovka:  
<https://vojtech-filipek.github.io/pages/orechovka1.html>
- NIST/SEMATECH e-Handbook of Statistical Methods:  
<https://www.itl.nist.gov/div898/handbook/pmd/section3/pmd3.htm>
- monography: Optimal Design of Experiments: A Case Study Approach:  
<https://www.wiley.com/en-us/Optimal+Design+of+Experiments%3A+A+Case+Study+Approach-p-9780470744611>

# Data analysis



taste rank  $\Rightarrow$  ordinal regression

(linear or beta regression)

non-trivial task:

- 9 batches (o.o.w. 3 extra sweet)
- variable significance
- (quasi-)complete separation issue
- $\exists$  of max. likelihood estimate

conclusion: diversity of recipes justified