

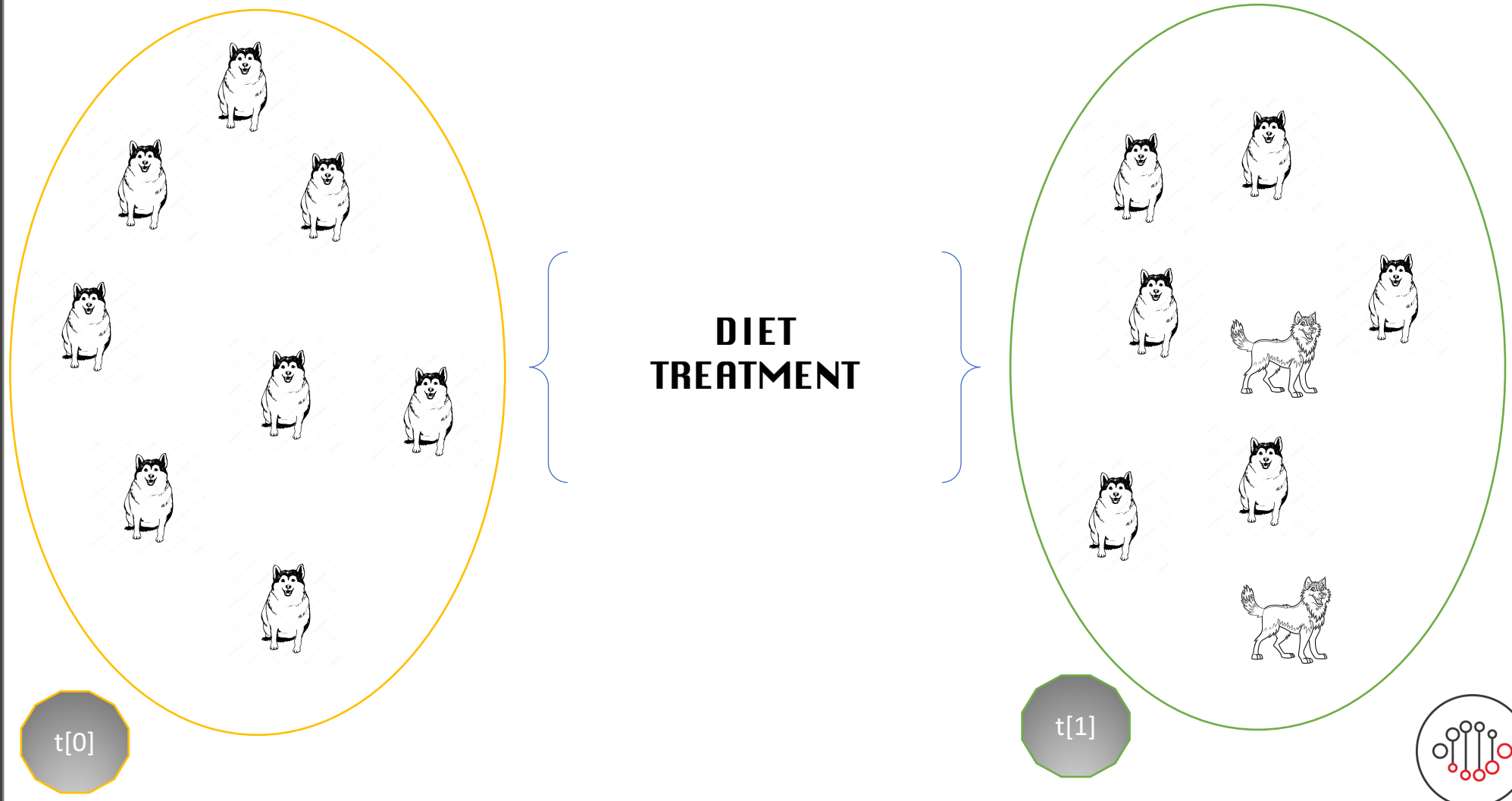
DATA LIONS

# Paired-sample t-test in epidemiology

*Szymon Moliński, 17.08.2017 r., Gdańsk*

# Paired-sample t-test use cases: dog overweight treatment

---



# Why are we performing this type of test?

---

Wikipedia: Les Twins during a shoot with photographer Shawn Welling.



Is therapy work?

Are samples changed through the time?

Are samples changed through the space?

The first step before the time-series analysis (e.g.: ANOVA)



# Calculations: steps

---

0. Analyze the properties of data.

1. Calculate the difference between sample at time  $t[1]$  and  $t[0]$  (before and after the treatment).

2. Calculate the mean difference.

3. Calculate the standard deviation of differences, calculate the standard error of mean difference.

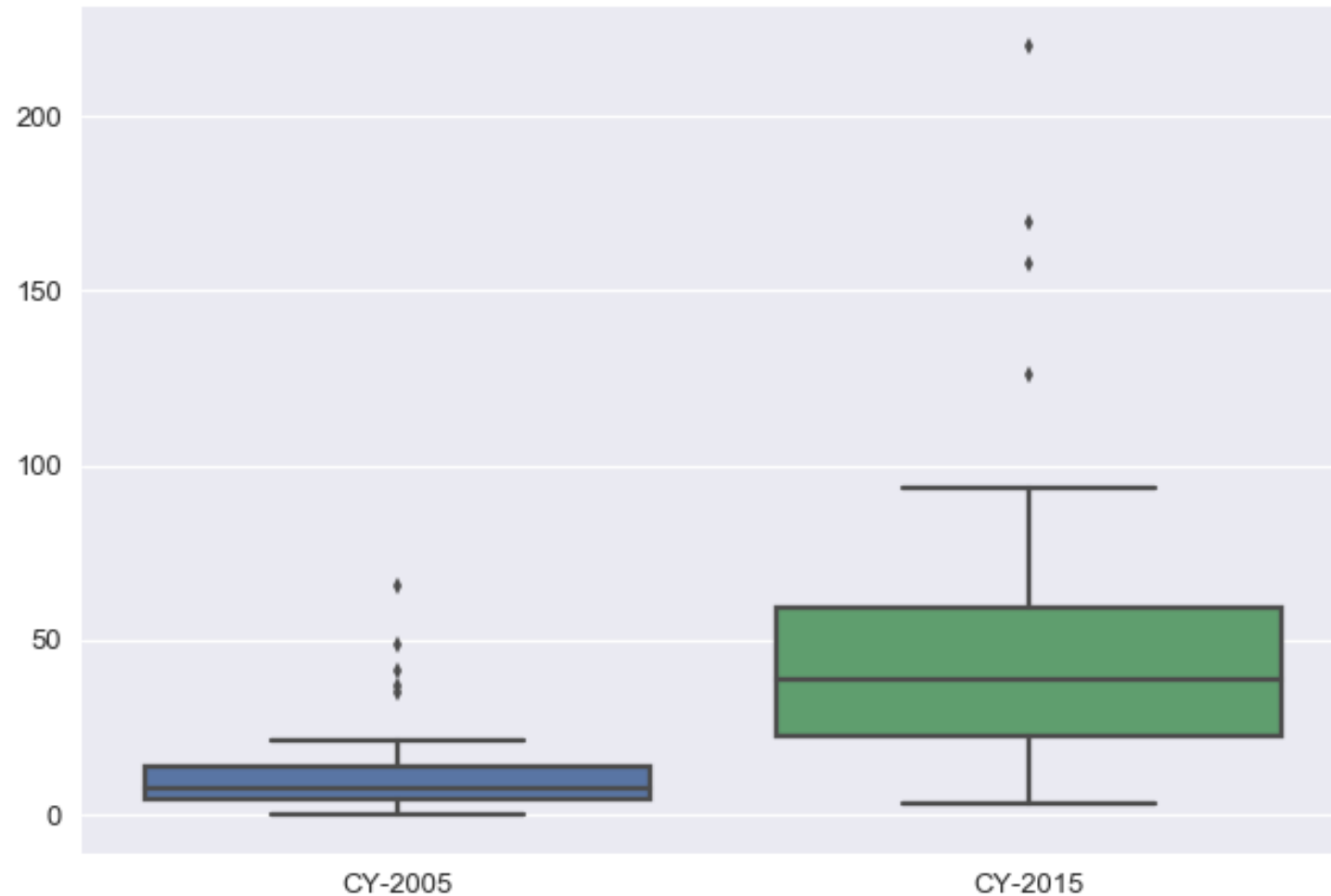
4. Calculate the t-statistics.

5. Compare t-distribution with tables for getting the value of T.  
This will give the p-value for paired t-test.



# Calculations: step 0 – data analysis

---



# EXERCISE

---

From the next slide we'll perform an exercise to understand what is going on behind the paired t-test.

Exercise is written for the MS Excel platform. You will be working in the document named *paired\_t\_test\_ex.xlsx*

Two datasets are prepared. You will check if the diet helps overweight dogs or not. Good luck!

- The part of the exercise is to find some kind of functions in the Excel software. The list (not sorted) of all needed functions:
- *AVERAGE* (ŚREDNIA), *SQRT* (PIERWIASTEK), *COUNT* (ILE.LICZB), *ABS* (MODUŁ.LICZBY), *T.DIST.2T* (ROZKŁAD.T.DS)





# Calculations: step 1 – calculate the difference between paired samples

1) Difference between pairs of observations

$$d_i = t[1]_i - t[0]_i$$

t[0] - observation before treatment  
t[1] – observation after treatment

D2	=[@[After diet]]-[@[Before diet]]				
	A	B	C	D	E
1	ID	Before diet	After diet	Difference d = (after - before)	(d - mean(d))^2
2	1	25	27	2	17,4
3	2	29	22	-7	23,2
4	3	32	36	4	38,1
5	4	38	33	-5	7,9
6	5	28	24	-4	3,3
7	6	37	25	-12	96,4
8	7	30	24	-6	14,5
9	8	27	36	9	124,9
10	9	37	24	-13	117,0





## Calculations: step 2 – calculate the Standard Deviation

2.1 ) Mean difference	$\bar{d} = \frac{\sum_{i=1}^n d_i}{n}$	n – number of samples
2.2 ) Sum of each (Difference – Mean Difference) ^ 2	$SS_{\bar{d}} = \sum_{i=1}^n (d_i - \bar{d})^2$	
2.3 ) Standard Deviation	$s_d = \sqrt{\frac{SS_{\bar{d}}}{n - 1}}$	







## Calculations: step 3 – calculate the standard error and degrees of freedom

---

3.1) Standard error of differences	$SE(\bar{d}) = \frac{s_d}{\sqrt{n}}$	
3.2 ) Degrees of freedom DF	$DF = n - 1$	

## Calculations: step 4 - calculate the t-statistics and P-value

---

4.1) t-statistics	$T = \frac{\bar{d}}{SE(\bar{d})}$	
-------------------	-----------------------------------	--





Calculations: step 4.2 – use tables of the t-distribution to compare value of T to the  $t_{n-1}$  distribution. This will give p-value for the paired t-test. In MS EXCEL you should use T.DIST.2T function to get the result.

Step 1:	Step 2:	Step 3:
Count difference d(1)	Count mean difference mean(d)	Sum of (difference - mean_difference)^2
2	-2,18	1985,38
Step 4:	State the null hypothesis H[0]	Diet program is not working; Average weight is the same
Step 5:	Formulate analysis plan	Significance level should be smaller than 0.05
Step 6:	Analyze sample data	
Step 6.1:	Step 6.2:	Step 6.3:
s (standard deviation of differences)	SE (standard error)	DF (degrees of freedom)
$s = \sqrt{\text{SUM}(d - \text{mean}(d))^2 / (\text{number of pairs in sample} - 1)}$	$SE = s / \sqrt{\text{number of pairs in sample}}$	$DF = (\text{number of pairs in sample} - 1)$
6,365371881	0,900199524	49
Step 7:		
t (t-statistic test)		
$t = [(\text{mean}(\text{after}) - \text{mean}(\text{before}) - D) / SE = (\text{mean}(\text{difference}) - \text{expected difference}) / SE$		
-2,42168535		
Step 8:		
Calculate P-value from t		
1,92%		
Diet has statistically significant effect on the dogs weight		





# Calculations: it can be done faster in MS Excel!

---

Go to:

Tools -> Data Analysis -> T-test: Paired Two Sample for Means

Enjoy!

T-test: Paired Two Sample for Means		
	Variable 1	Variable 2
Mean	32.28	30.1
Variance	22.77714286	29.68367347
Observations	50	50
Pearson's correlation	0.22965177	
H(0)	0	
df	49	
t Stat	2.42168535	
P(T<=t) one tail	0.009597746	
T test one tail	1.676550893	
P(T<=t) two tails	0.019195491	
T test two tails	2.009575237	





# PROGRAMMING EXCERCISE

---

At the end you have to create simple Java Script module which is calculating the t statistics.

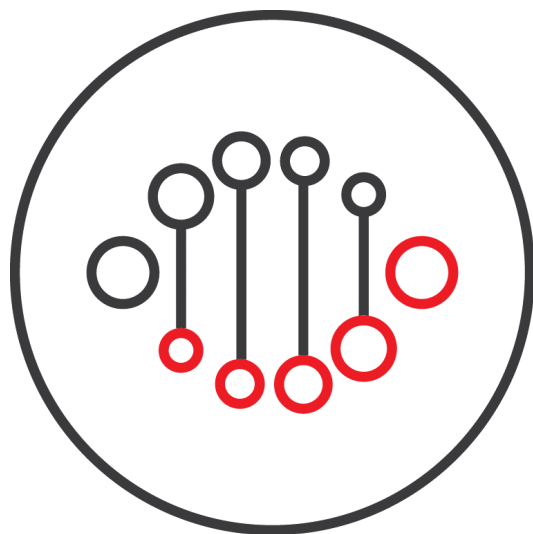
Excercise has two parts:

- 1) Calculations (up to t-statististics),
- 2) Reading data from the table (t-distribution tables) – the second part will be exploited later, during the next workshop, so do not be worry about that.

Steps to do:

- 1) Read JSON file with 'CY-2005' object and 'CY-2015' object (each object is created from the ID key and value for this key), file name is data.json,
- 2) Perform all calculations: for each ID subtract var1 value from var2 value.
- 3) Calculate the mean difference.
- 4) Calculate Standard Deviation and Standard Error.
- 5) Calculate t-statistics.





DATA LIONS

---

*Empathy and Curiosity are our driving factors*