Iris Setosa

CAS
Applied Data Science

The Mutation of Iris Setosa
A Criminal Story

C. Cris *, M. Freunek *
* Swiss Federal Institute of Intellectual Property, 3013 Bern

Initial Situation and Summary

- Company v4Setosa owns a ***patent*** claiming ***protection*** *on* ***gene sequence*** of Iris Setosa
- Iris Setosa found on a farmers field, **treated with a product** from Sonte Manto, show ***possibly larger setal leaves*** than normal, possibly indicating ***a gene mutation*** and therewith a ***patent infringement of the patent of v4Setosa***
- **Two data sets available**:
  - ✓ A reference data set of definitely non-mutated Iris Setosa
  - ✓ A mixed data-set of farmers Iris Setosa, Virginica and Versicolor

➢ ***Hypothesis test (t-test)*** shows significant evidence of gene mutation of Iris Setosa setal leave

Task and 0-Hypothesis

By means of *the* ***reference data set*** of ***non-mutated*** Iris Setosa investigate, if there is a mutation of the farmer Iris Setosa setal size.

0-Hypothesis: There is NO mutation of farmers Iris Setosa (increasing size of setal leaves) treated with Sonte Manto.
1-Hypothesis: There is a mutation of farmers Iris Setosa (increasing size of setal leaves) treated with Sonte Manto.

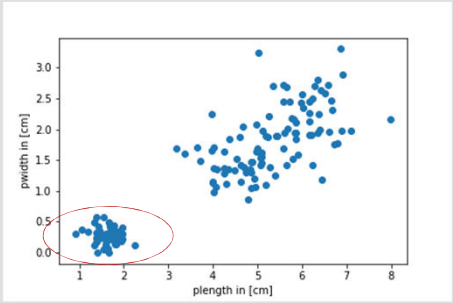
Methods / Procedure / Test

1. Identifying Iris Setosa in farmers data set:

	slength	swidth	plength	pwidth
0	7.512300	3.472900	1.892329	0.304084
1	6.305038	3.356097	1.647476	0.279527
2	4.818646	4.575827	1.380733	0.165094
3	6.024839	4.042857	1.585929	0.203635
4	5.792077	3.278056	2.251739	0.113221

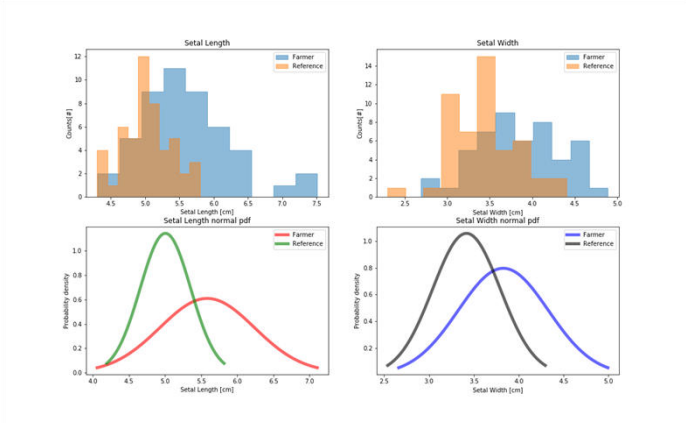
Remark: Farmer data through random generator:
np.random.normal(x,y, size = pts) *
np.random.normal(1.1,0.1, size = pts)

Scatterplot yields Iris Setosa



Filter for Setosa:
- p_l < 2.5 cm
- p_w < 0.7 cm

2. T-test of two samples, not paired, samples differ in standard deviation (.ttest_ind(equal=False))



		Slenght [cm]	Swidth [cm]
Reference data	mean	5.01	3.42
	sd	0.35	0.38
	p-test normal distribution	0.91	0.39
Farmers data	mean	5.59	3.82
	sd	0.65	0.50
	p-test normal distribution	0.03	0.78

3. T-Test yields:

	p-value
setal length	$7.6 \cdot 10^{-7}$
setal width	$1.7 \cdot 10^{-5}$

Conclusions

- Normal distribution of reference and farmer data
- reference and farmer data differ in SD
- Iris Setosa in farmer data can be filtered via petal size
- p-test with p << 0.05 clearly rejects the 0-hypothesis and supports 1-hypothesis