Practical Tests (R SoftWare)

Presented by Léandre Guenat, Julie Boisard, Eloise Ellegaard & Diego Gutierrez

Introduction to our project

The problematic and why ?

“How Cattle production is significantly different from other production type in France ?”

The main objective is to present the database and the results we found for the problematic while using statistical test and practices.

Why ?

Relationship with real world

In the dataset we are faced with real data that were collected in France so our analysis can and will reflect the situation in France and can be helpful on 2 different aspects

Environmental impact

What are the environmental impacts associated with cattle production compared other livestock, in terms of resource use and sustainability ?

Economical impact

How does the economic output of cattle production compare to other types of agricultural activities in terms of profitability, sustainability, and growth ?

The dataset

What is this about ?

R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE
UTASA	UTATO	UGBBO	UGBCA	UGBHE	UGBOV	UGBPO	UGBTO	SAFVD	SAFER	SAUTI	SUTOT	SFPTO	CHAFF
0.28	1.28	0	0	0	0	0	0	0	14	18	18	1	155810
0	1	0	0	0	0	0	0	3	15	17	17	1	72755
0.07	1.07	24	0	24	0	0	24	0	14	14	14	14	113872.32
0	3	2	0	35	34	0	35	0	18	18	18	16	311698
0	1	0	0	9	9	0	9	0	7	7	7	6	80631
6.78	7.78	0	0	0	0	0	0	0	15	15	15	1	603320
3.54	4.54	0	0	0	0	0	0	0	16	16	16	0	297151
0	1.5	0	0	0	0	0	0	0	22	22	22	0	133054
0	1	0	0	0	0	0	0	0	9	18	18	1	112667
0	1	0	0	9	9	0	9	6	7	12	12	10	113357
0	1	17	0	17	0	0	17	6	13	16	16	11	109450
0	1	13	0	13	0	0	13	0	8	11	11	11	42622.37
0	1	0	0	0	0	0	0	6	7	12	12	1	23979
2.38	4.38	0	0	0	0	0	0	0	4	4	4	1	205570.26
0	1.08	8	11	18	0	0	46	6	14	17	17	10	538979.11
0	1	5	0	5	0	0	5	0	22	22	22	4	202279

The kind of the dataset ?

Where does it come from?

- The Farm Accountancy Data Network (RICA) is a survey that collects data on farms to track their economic and technological evolution.

- This survey has been carried out in the member states of the European Union since 1968.

What types of information can be found in this data set ?

- Economic information
- Diversity of Farms
- Farm Structure
- Geographical and Temporal Scope
- Sustainability Indicators
- Utilization in Policy Analysis

What pre-processing was done on the dataset?

- The data was separated into different categories and also labeled with a specific name that is defined in the second part of the dataset

- We cleaned up the dataset to ensure that only the information we are interested in is available, for a more reliable and accurate result.

Main challenges of the dataset

The main challenge is to sort the data and to only use the right analysis and statistical tests to extract the right information out of the dataset

Main variables of the dataset

UGB

+

Revenues

+

Crop production

+

Animal production

Our columns

UGBBO :
UGBbovine

UGBHE :
UGBherbivore

UGBBO :
UGBovine

UGBPO :
UGBporcine

UGBCA :
UGBcaprine

PRODA :
animal production

CHAFF : revenue

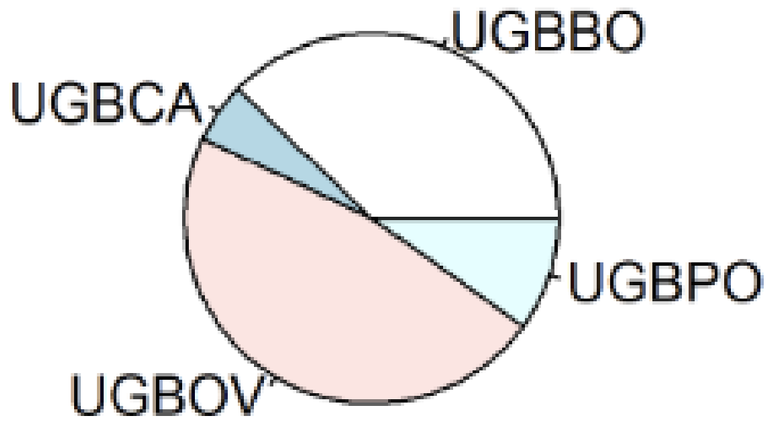
ZENVI : Zones with environmental constraints

The analysis

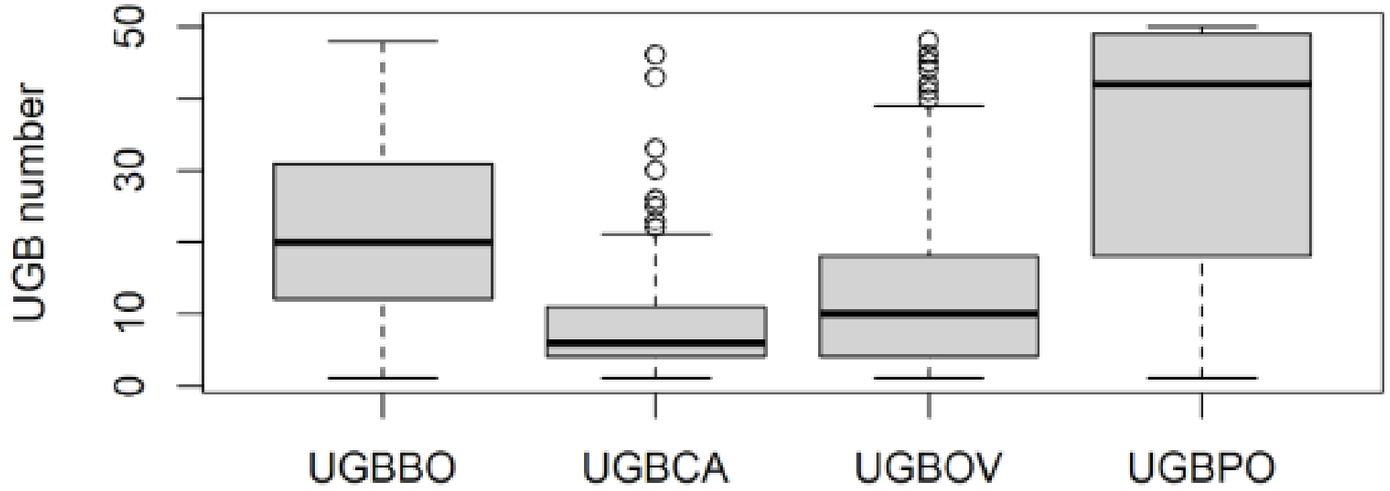
Repartiton of the different UGB

With these graphs we can see that the vast majority of UGB are for bovines,
It shows the importance of bovine production in France compared to other meat production

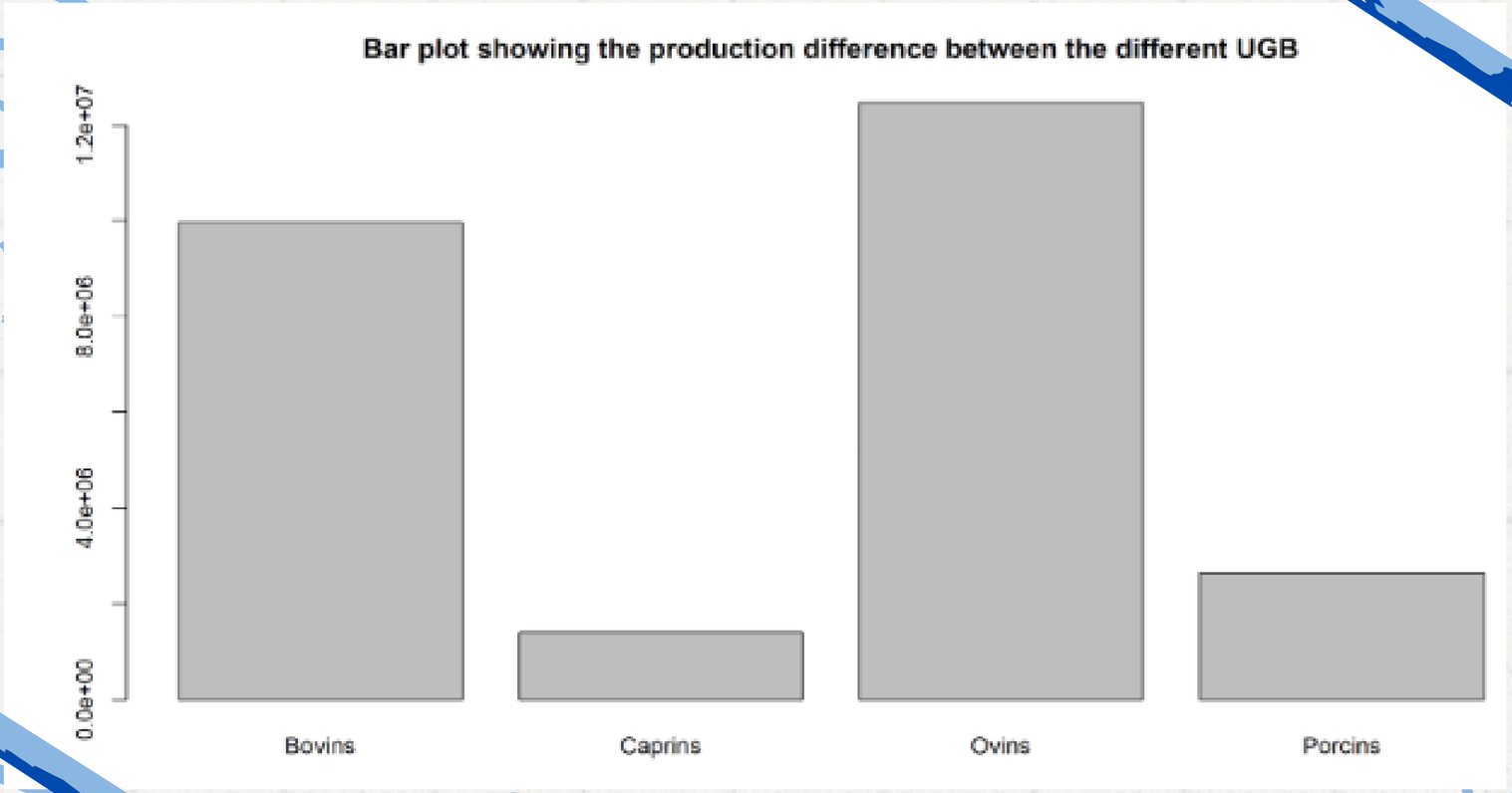
Repartition of each UGB



Statistical comparison of each relevant UGB



Difference of production

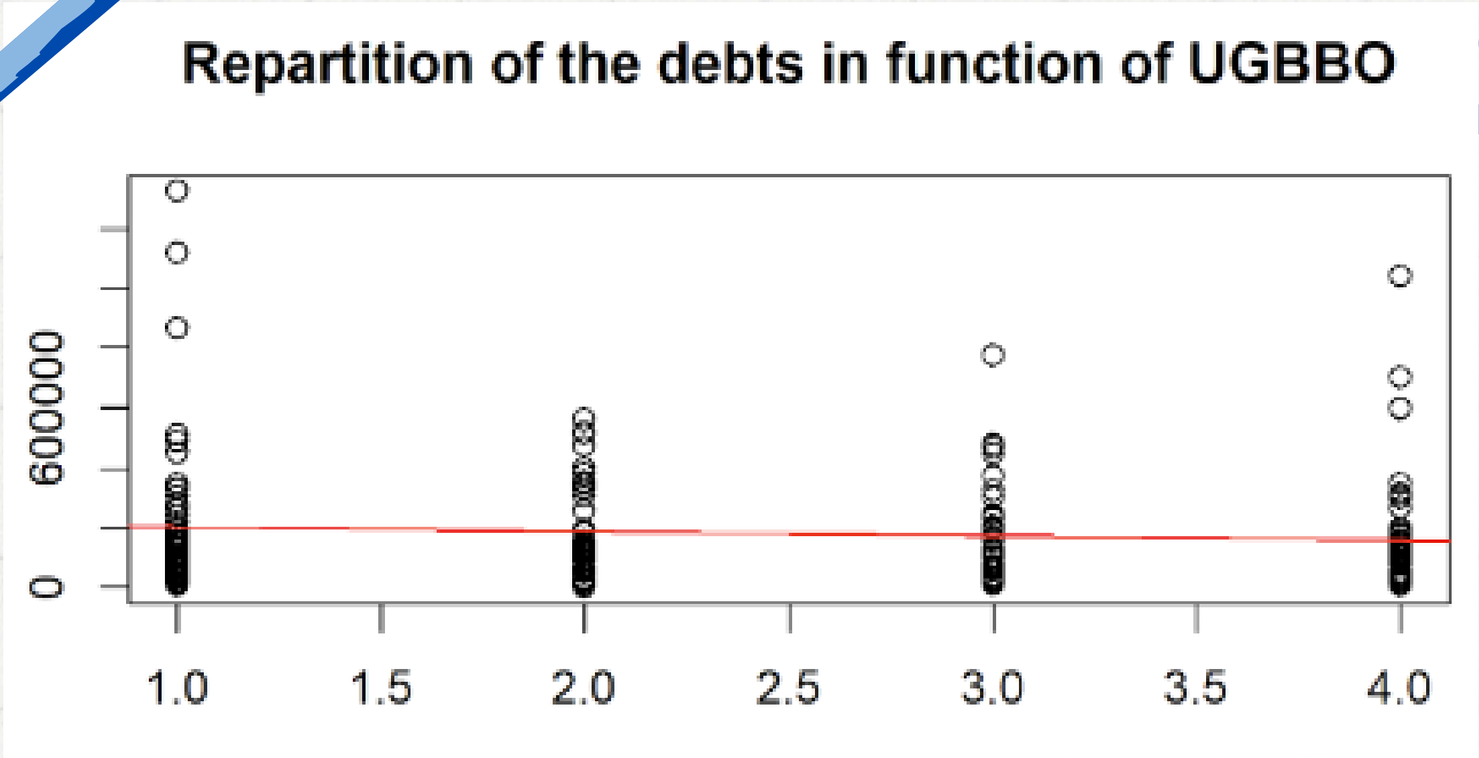


In the graph, we can see that there are more ovines production than bovines production even though we have more farm producing beef

Debt depending of the UGB

As we can see on the graph the number of ugb isn’t representative of the debt farmers have.

There aren’t any correlation between the amount of debt and the number of UGB bovine



Shapiro Wilk test UGB bovine

```
        Shapiro-Wilk normality test

data:  New_RICA_clean$UGBBO
W = 0.85188, p-value = 5.911e-12
```

Shapiro Wilk test UGB Ovine

```
        Shapiro-Wilk normality test

data:  New_RICA_clean$UGBOV
W = 0.71515, p-value < 2.2e-16
```

Shapiro Wilk UGB Caprine

```
        Shapiro-Wilk normality test

data:  New_RICA_clean$UGBCA
W = 0.80781, p-value = 8.073e-14
```

Shapiro Wilk test UGB Porcine

```
        Shapiro-Wilk normality test

data:  New_RICA_clean$UGBPO
W = 0.86949, p-value = 4.198e-11
```

Shapiro Wilk test UGB herbivore

```
        Shapiro-Wilk normality test

data:  New_RICA$UGBHE
W = 0.97125, p-value < 2.2e-16
```

```
data:  New_RICA_clean_Dettes$UGBBO and New_RICA_clean_Dettes$UGBHE
F = 2.9615, num df = 172, denom df = 172, p-value = 3.775e-12
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 2.194165 3.997108
sample estimates:
ratio of variances
          2.961472
```

Variances

```
data:  New_RICA_clean_Dettes$UGBOV and New_RICA_clean_Dettes$UGBHE
F = 2.9522, num df = 172, denom df = 172, p-value = 4.326e-12
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 2.187263 3.984533
sample estimates:
ratio of variances
          2.952155
```

The results suggest that there might be more variation in the production or management conditions affecting the UGBBO and UGBOV groups compared to UGBHE

Are bovine more susceptible to be in zones with more environmental constraints ?

```
        Pearson's Chi-squared test

data:  RICA_Data_CSV$UGBBO and RICA_Data_CSV$ZENVI
X-squared = 79.674, df = 48, p-value = 0.002751
```

Bovine

```
        Pearson's Chi-squared test

data:  RICA_Data_CSV$UGBOV and RICA_Data_CSV$ZENVI
X-squared = 57.615, df = 45, p-value = 0.09829
```

Ovine

```
        Pearson's Chi-squared test

data:  RICA_Data_CSV$UGBCA and RICA_Data_CSV$ZENVI
X-squared = 78.916, df = 28, p-value = 9.67e-07
```

Caprine

```
        Pearson's Chi-squared test

data:  RICA_Data_CSV$UGBPO and RICA_Data_CSV$ZENVI
X-squared = 32.806, df = 50, p-value = 0.9713
```

Porcine

```
        Pearson's Chi-squared test

data:  RICA_Data_CSV$UGBHE and RICA_Data_CSV$ZENVI
X-squared = 87.415, df = 48, p-value = 0.00044
```

Herbivores

Bovines, Herbivores, and Caprins show significant associations with the variable “ZENVI”, suggesting that these livestock types are more susceptible to be in zones with more environmental constraints

```
        One Sample t-test

data:  New_RICA$UGBBO
t = 102.43, df = 3112, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 21.11254 21.93661
sample estimates:
mean of x
 21.52457
```

Student's T test UGB bovine

```
        One Sample t-test

data:  New_RICA$UGBOV
t = 29.796, df = 617, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 11.35572 12.95820
sample estimates:
mean of x
 12.15696
```

Student's T test UGB Ovine

```
        One Sample t-test

data:  New_RICA$UGBCA
t = 14.924, df = 172, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 7.338135 9.575160
sample estimates:
mean of x
 8.456647
```

Student's T test UGB Caprine

```
        One Sample t-test

data:  New_RICA$UGBPO
t = 37.779, df = 420, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 31.78065 35.26923
sample estimates:
mean of x
 33.52494
```

Student's T test UGB Porcine

```
        One Sample t-test

data:  New_RICA$UGBHE
t = 105.89, df = 3616, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 20.50698 21.28069
sample estimates:
mean of x
 20.89383
```

Student’s T test UGB herbivore

Not normaly distributed ? An alternative : The Wilcoxon signed rank test

```
        Wilcoxon signed rank test with continuity correction

data:  New_RICA$UGBPO
V = 88831, p-value < 2.2e-16
alternative hypothesis: true location is not equal to 0
```

Porcine

```
        Wilcoxon signed rank test with continuity correction

data:  New_RICA$UGBHE
V = 6543153, p-value < 2.2e-16
alternative hypothesis: true location is not equal to 0
```

Herbivore

The Wilcoxon test is a non-parametric statistical test used to compare two independent samples. It is mainly used when data do not follow a normal distribution.

Not normaly distributed ? An alternative : The Wilcoxon signed rank test

```
        Wilcoxon signed rank test with continuity correction

data:  New_RICA$UGBBO
V = 4846941, p-value < 2.2e-16
alternative hypothesis: true location is not equal to 0
```

Bovine

```
        Wilcoxon signed rank test with continuity correction

data:  New_RICA$UGBOV
V = 191271, p-value < 2.2e-16
alternative hypothesis: true location is not equal to 0
```

Ovine

```
        Wilcoxon signed rank test with continuity correction

data:  New_RICA$UGBCA
V = 15051, p-value < 2.2e-16
alternative hypothesis: true location is not equal to 0
```

Caprine

Conclusion

Summary

To summarize all the graph, we can see that the bovine production is different than others because there are a lot of small farms that produces the most meat in France while other farms are big but they aren't many.

Summary

To summarize:

- Shapiro-Wilk: no normality
- Variance: no significant difference
- Chi-squared test: Bovine, herbivores, caprins are more susceptible to be in zone with more environmental constraint.
- t-test: not relevant because of the non-normality of the data
- Wilcoxon: an alternative to t-test

Results

With all the statistical and graphical data, we saw that the bovine production is significantly different than other animal production due to its nature and productivity

Limits of our dataset

- Excel and R are less efficient with big dataset
- When cleaning the dataset we found negative values where there shouldn't be any.
- Since our values are not distributed normally,the t-test is not accurate so we had to do a Wilcoxon test.
- We were limited by time and columns.

Thank you !