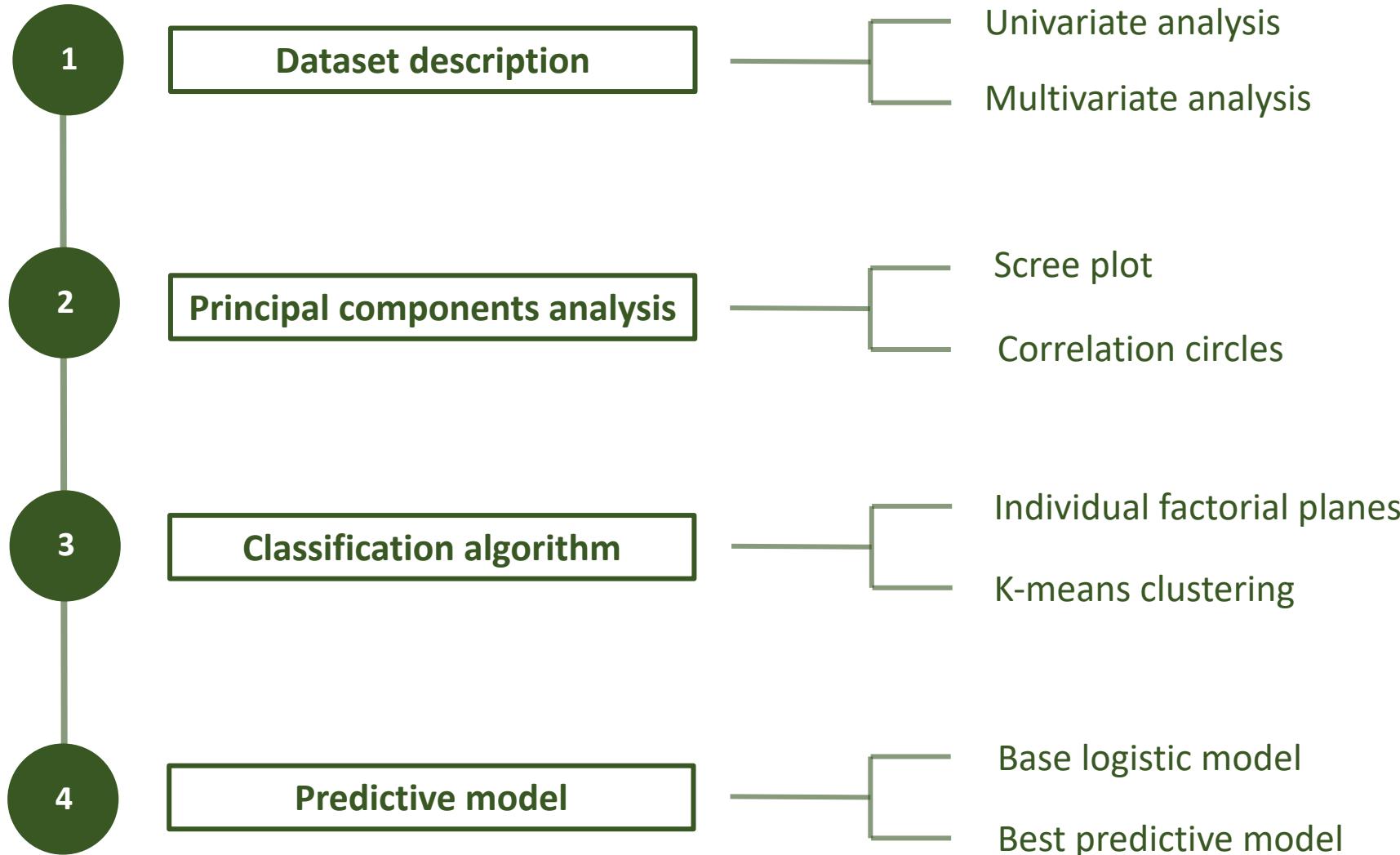


Detect counterfeit notes



Summary



Univariate analysis

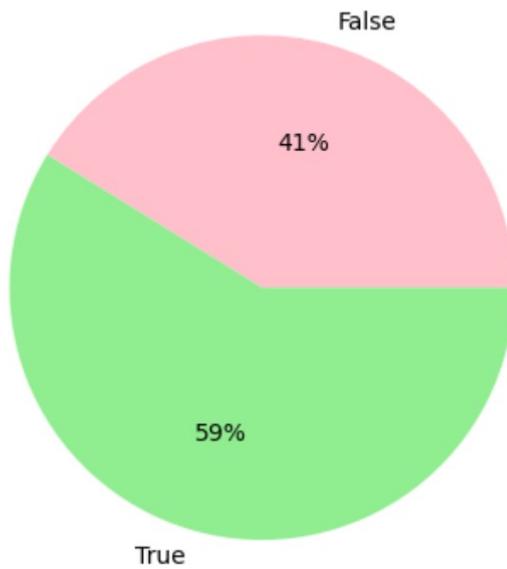
Dataset presentation

1 qualitative value: Boolean type



6 quantitative values: float type

True and false notes distribution



	diagonal	height_left	height_right	margin_low	margin_up	length
count	170.000000	170.000000	170.000000	170.000000	170.000000	170.000000
mean	171.940588	104.066353	103.928118	4.612118	3.170412	112.570412
std	0.305768	0.298185	0.330980	0.702103	0.236361	0.924448
min	171.040000	103.230000	103.140000	3.540000	2.270000	109.970000
25%	171.730000	103.842500	103.690000	4.050000	3.012500	111.855000
50%	171.945000	104.055000	103.950000	4.450000	3.170000	112.845000
75%	172.137500	104.287500	104.170000	5.127500	3.330000	113.287500
max	173.010000	104.860000	104.950000	6.280000	3.680000	113.980000

no null values

Univariate analysis

Diagonal

Height_left

Height_right

Margin_low

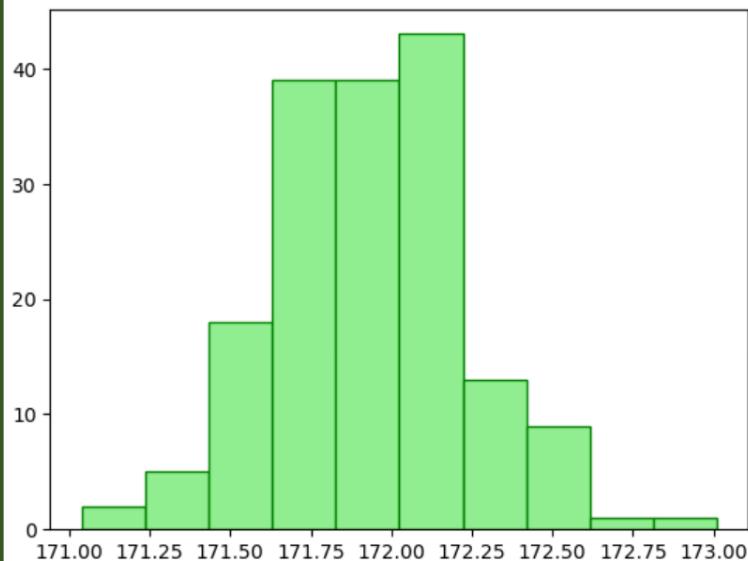
Margin_up

Length

Distribution

Distribution of diagonal values

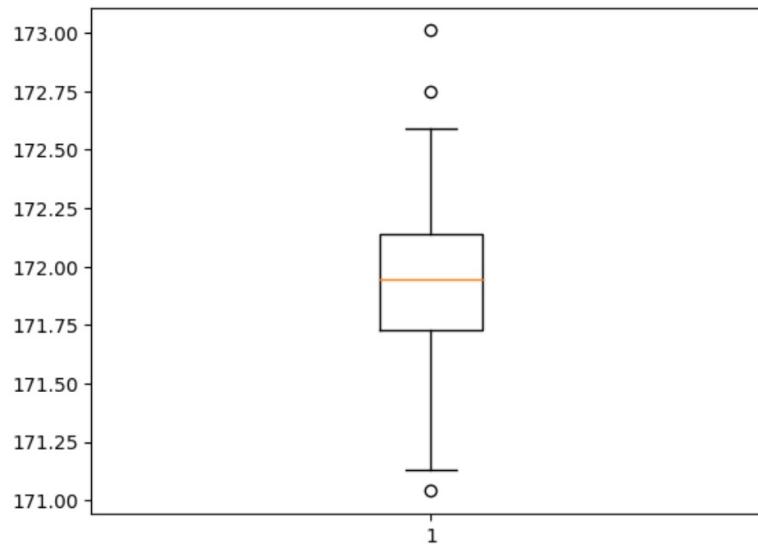
Normality test: 0.61 --> H0 accepted
Skewness: 0.2
Kurtosis: 0.59



Dispersion

diagonal

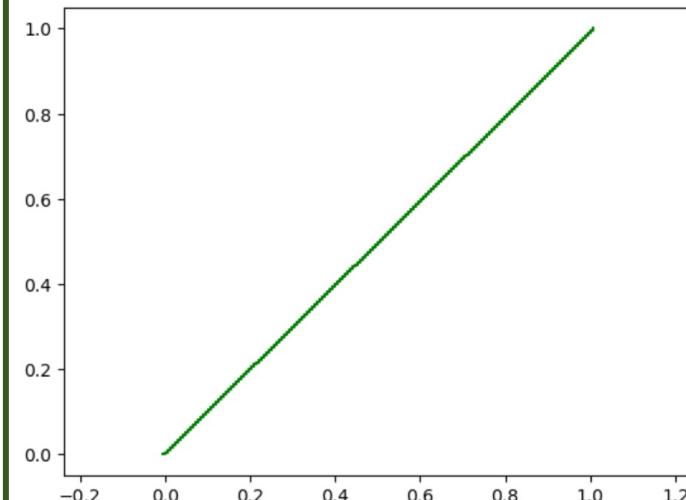
Outliers detected: 6
5 172.57
23 172.59
34 172.75
39 171.13
70 171.04
166 173.01
Name: diagonal, dtype: float64



Concentration

Lorenz curve of diagonal

Gini index: 0.00098926502133323



Univariate analysis

Diagonal

Height_left

Height_right

Margin_low

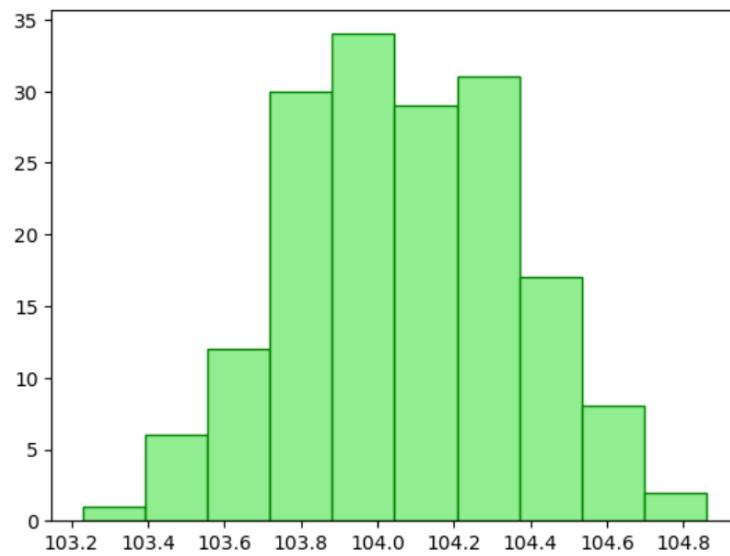
Margin_up

Length

Distribution

Distribution of height_left values

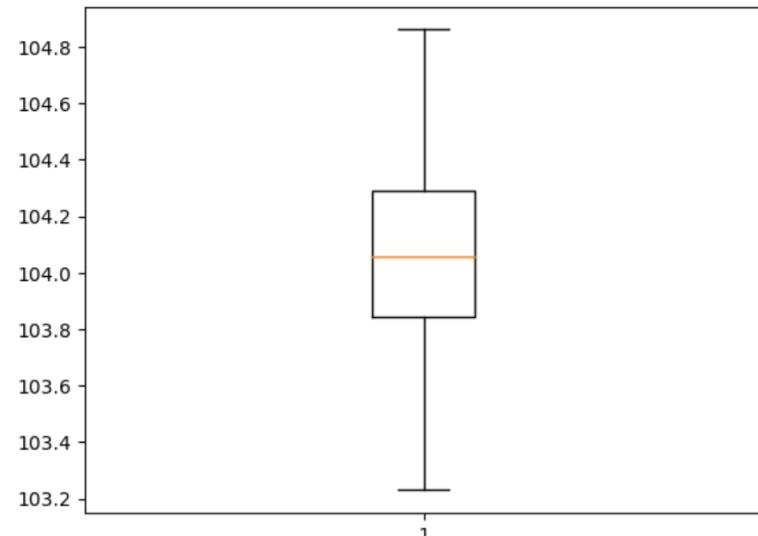
Normality test: 0.55 --> H0 accepted
Skewness: 0.03
Kurtosis: -0.46



Dispersion

height_left

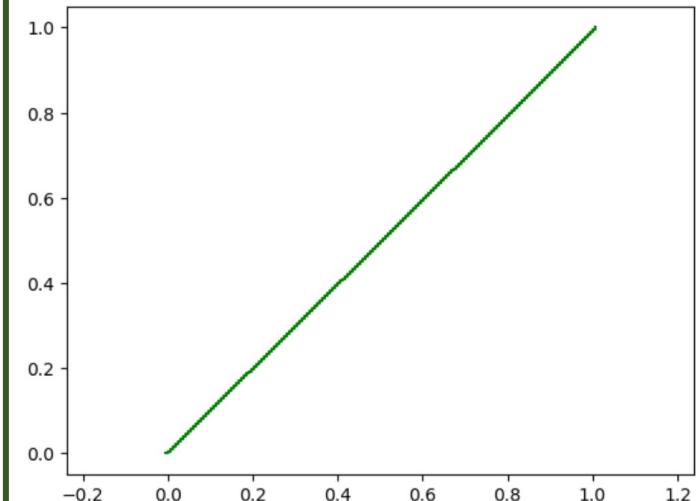
Outliers detected: 2
0 104.86
49 103.23
Name: height_left, dtype: float64



Concentration

Lorenz curve of height_left

Gini index: 0.0016290976983809324



Univariate analysis

Diagonal

Height_left

Height_right

Margin_low

Margin_up

Length

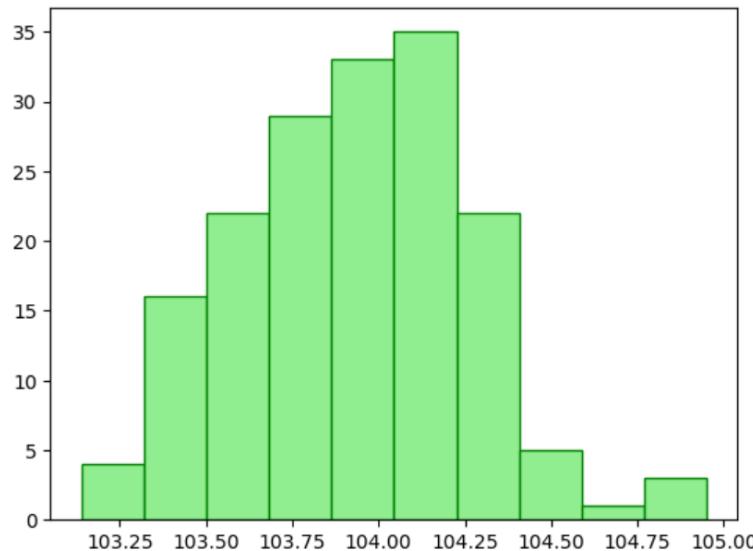
Distribution

Distribution of height_right values

Normality test: 0.16 --> H0 accepted

Skewness: 0.17

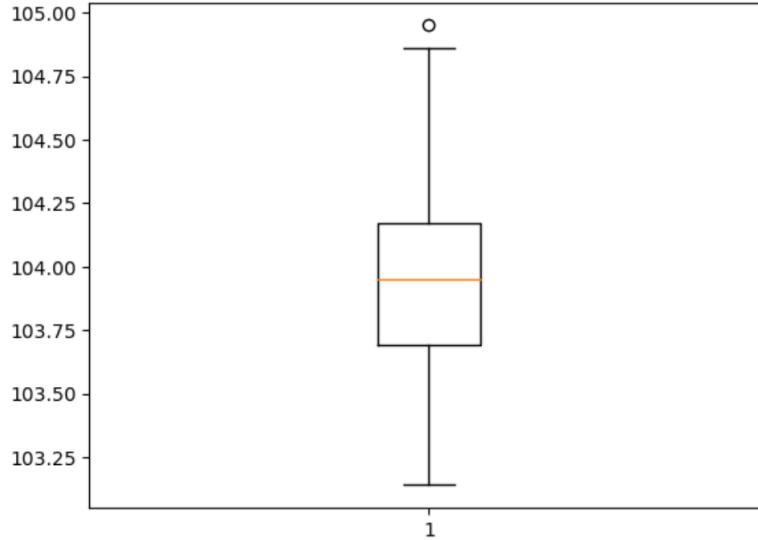
Kurtosis: -0.0



Dispersion

height_right

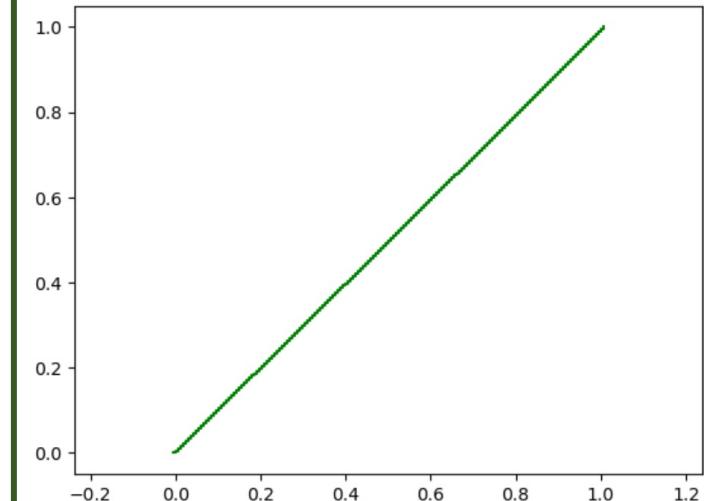
Outliers detected: 4
0 104.95
39 103.14
112 104.83
122 104.86
Name: height_right, dtype: float64



Concentration

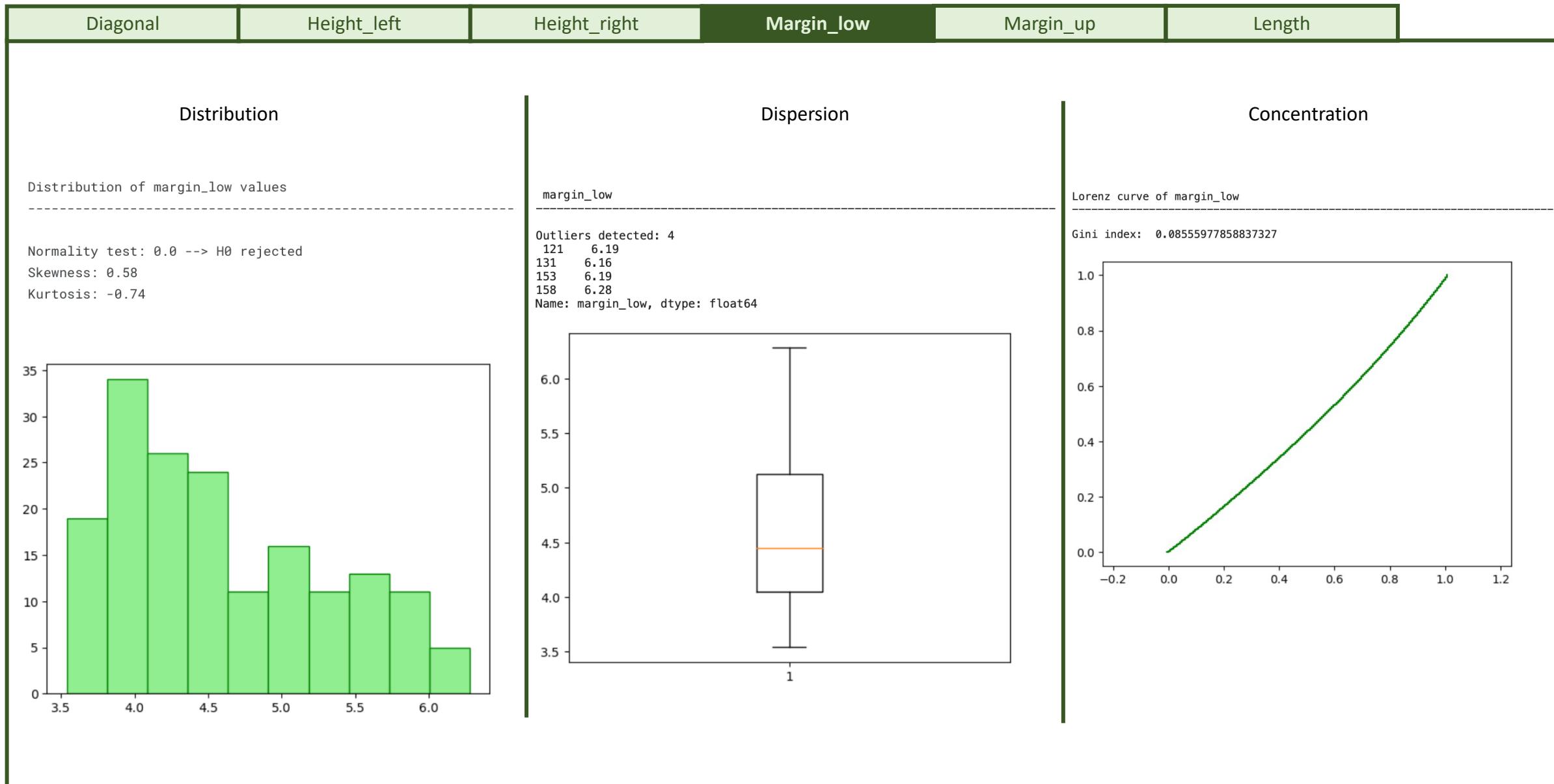
Lorenz curve of height_right

Gini index: 0.0017930612541421365



Univariate analysis

1 - Dataset description



Univariate analysis

Diagonal

Height_left

Height_right

Margin_low

Margin_up

Length

Distribution

Dispersion

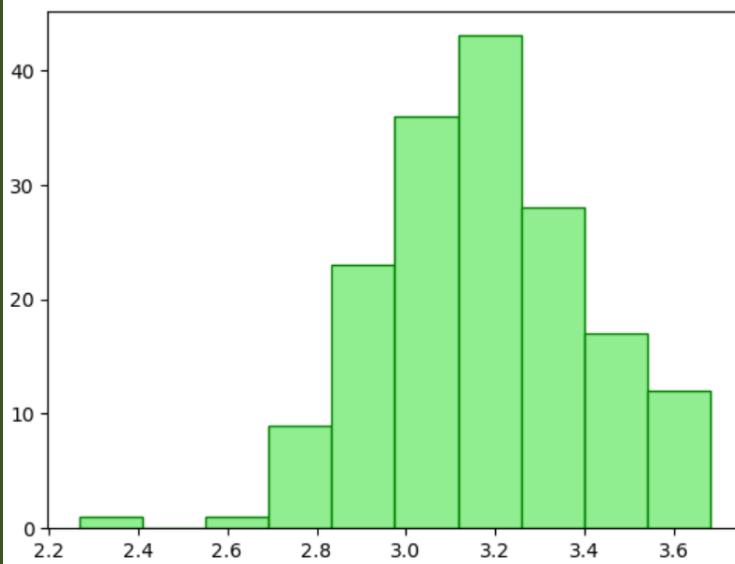
Concentration

Distribution of margin_up values

Normality test: 0.2 --> H0 accepted

Skewness: -0.2

Kurtosis: 0.55

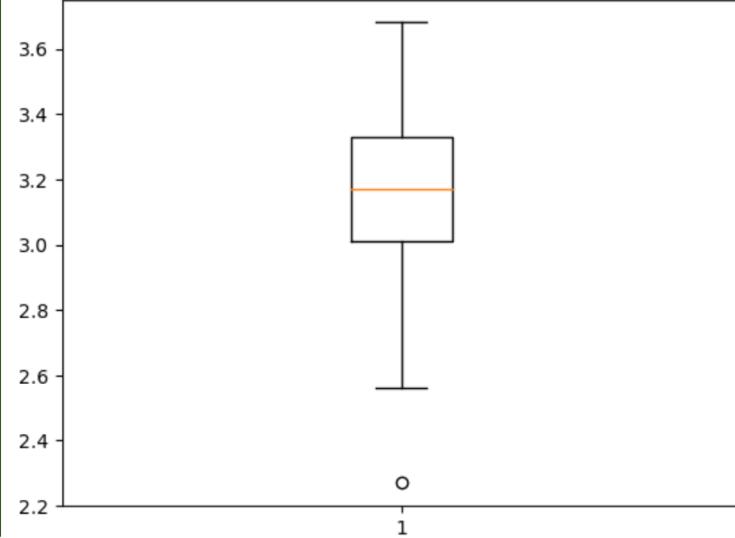


margin_up

Outliers detected: 6

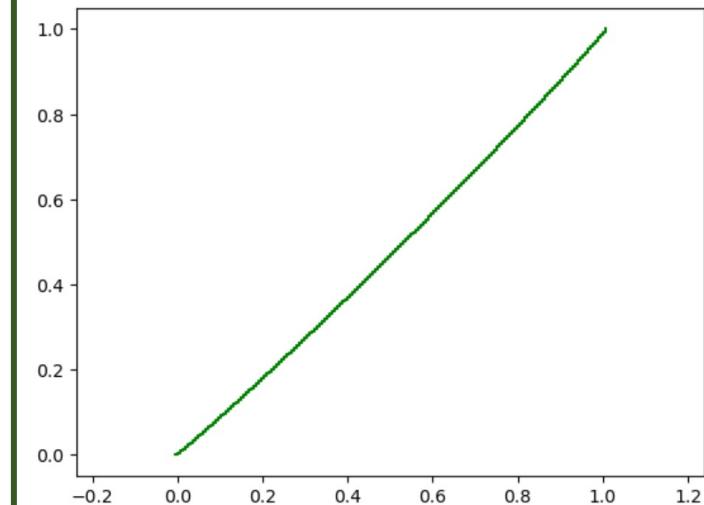
4	2.27
15	2.56
106	3.67
120	3.68
129	3.66
149	3.65

Name: margin_up, dtype: float64



Lorenz curve of margin_up

Gini index: 0.04163136876547724



Univariate analysis

Diagonal

Height_left

Height_right

Margin_low

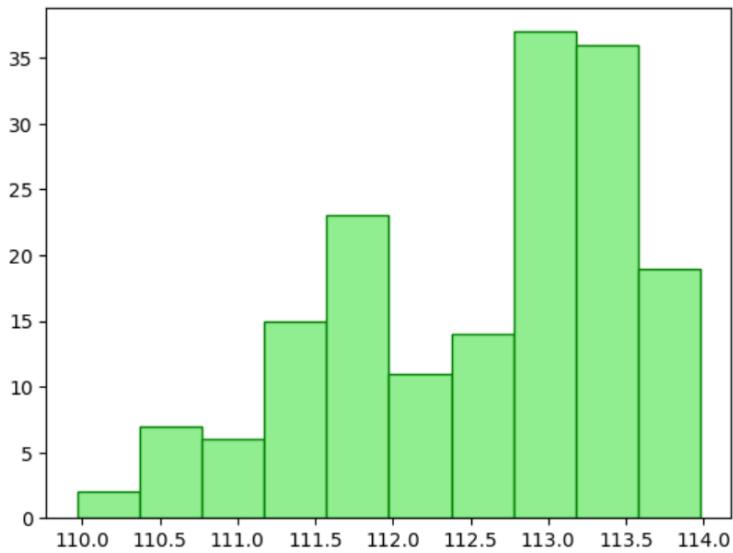
Margin_up

Length

Distribution

Distribution of length values

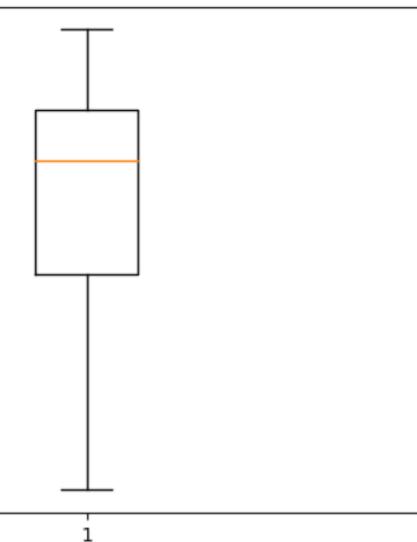
Normality test: 0.0 --> H0 rejected
Skewness: -0.66
Kurtosis: -0.53



Dispersion

length

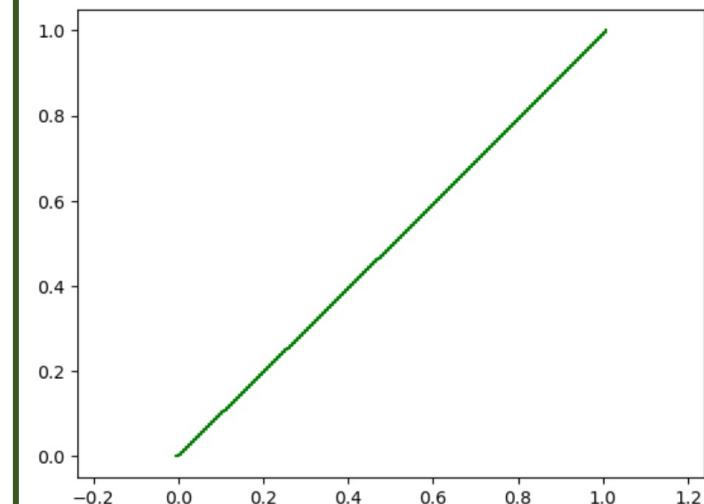
Outliers detected: 7
137 110.64
142 110.53
147 110.48
151 109.97
159 110.31
161 110.61
167 110.68
Name: length, dtype: float64



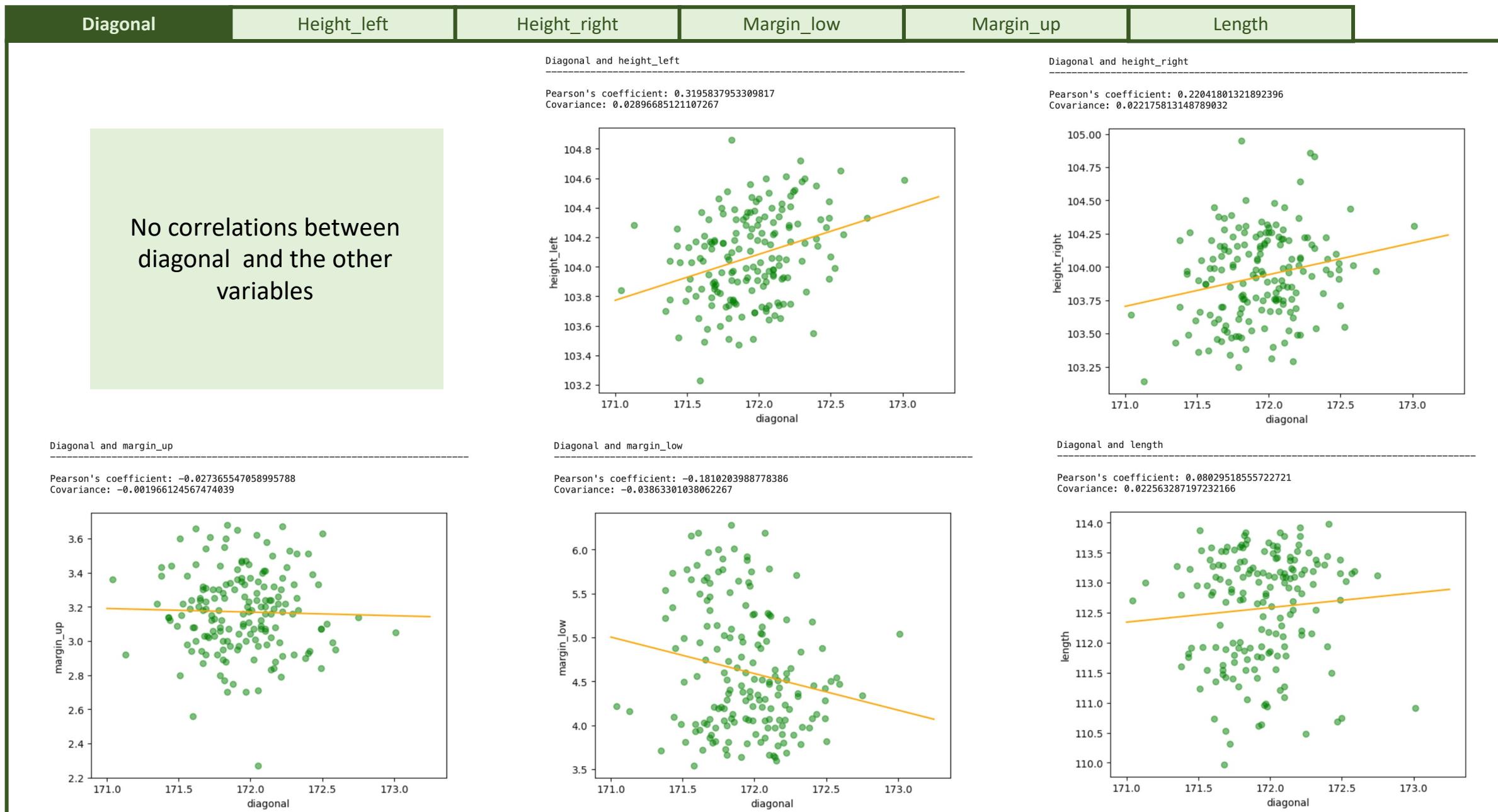
Concentration

Lorenz curve of length

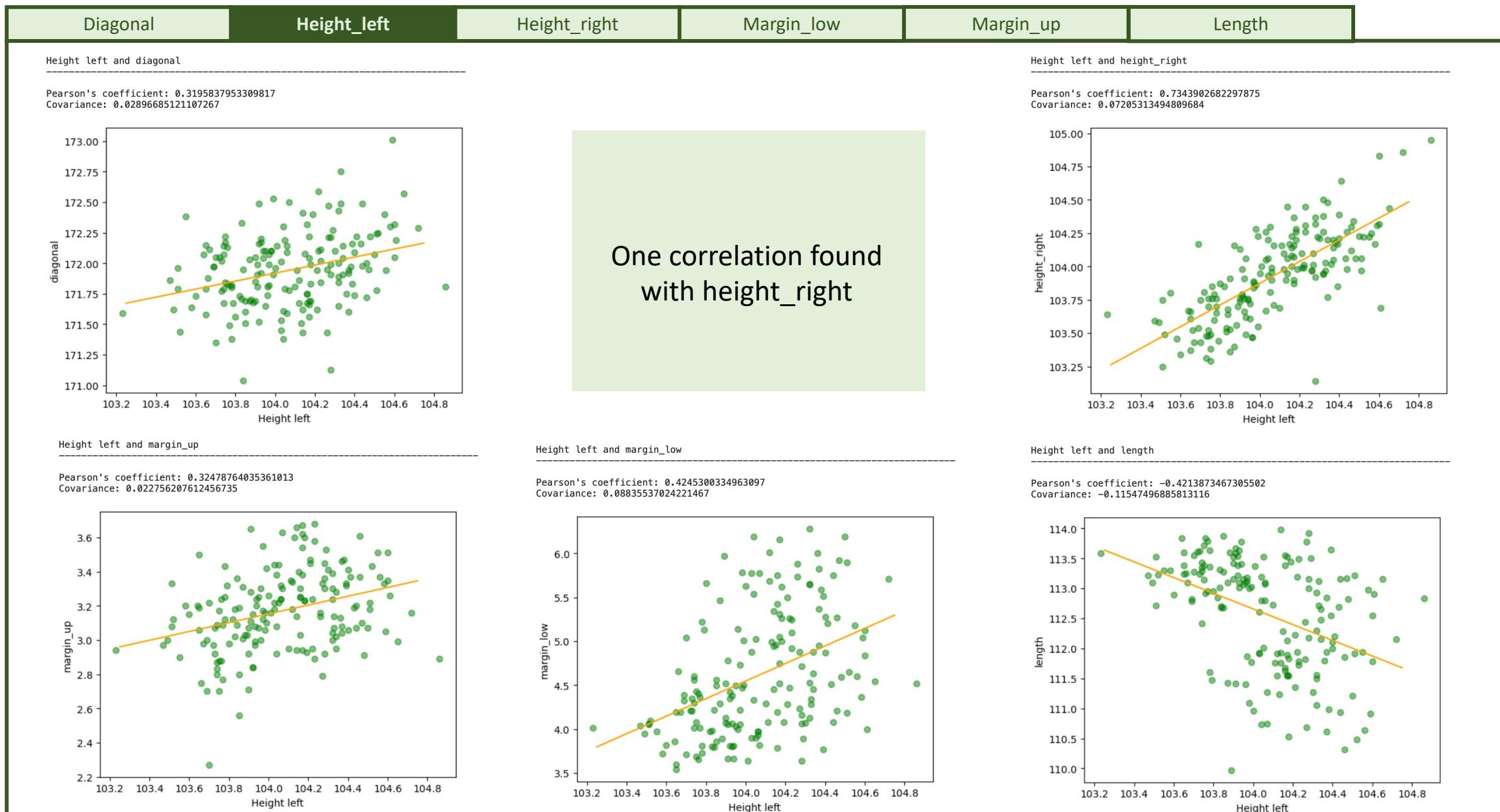
Gini index: 0.004581784398901112



Multivariate analysis – Quantitative variables



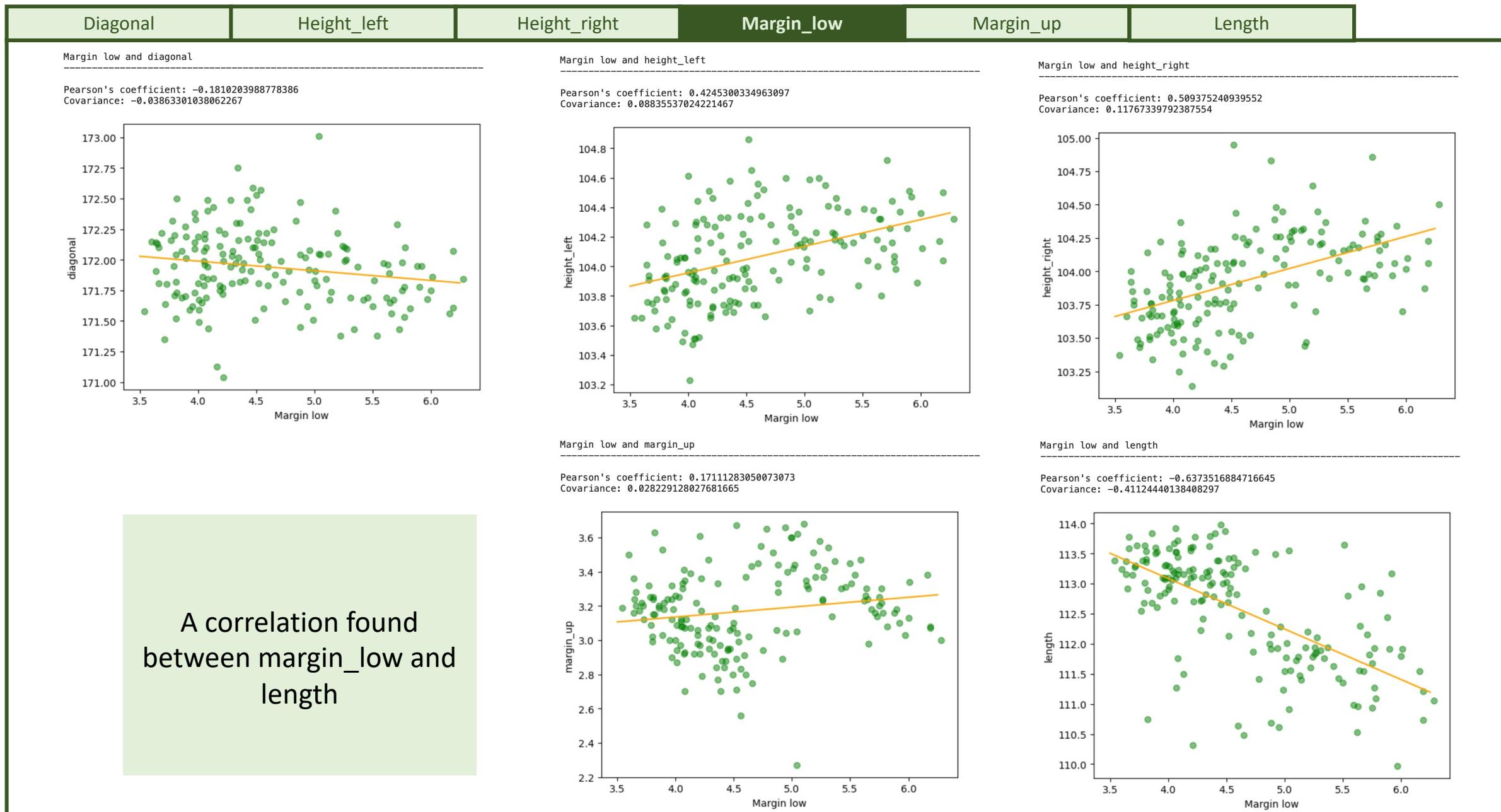
Multivariate analysis – Quantitative variables



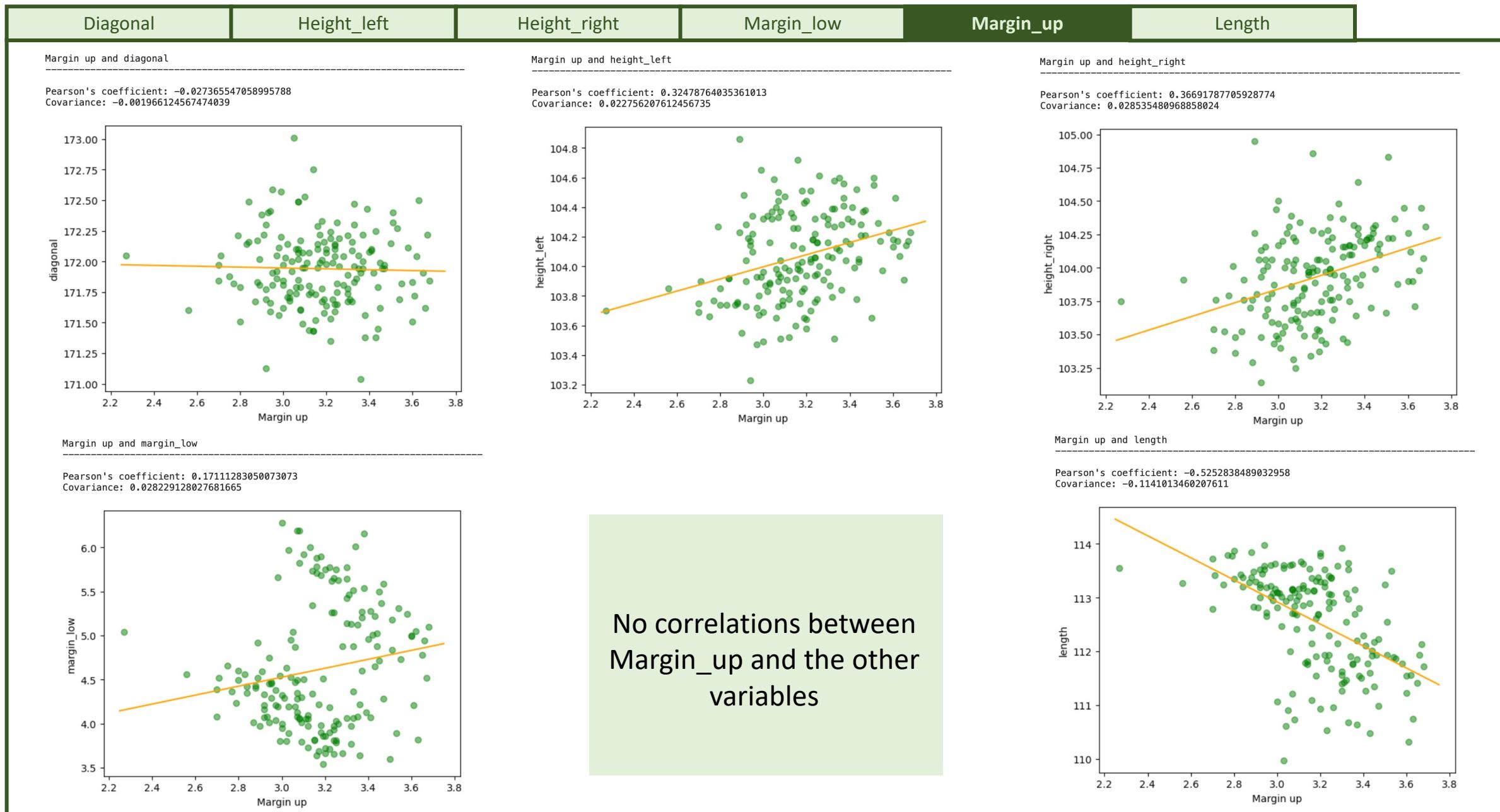
Multivariate analysis – Quantitative variables



Multivariate analysis – Quantitative variables

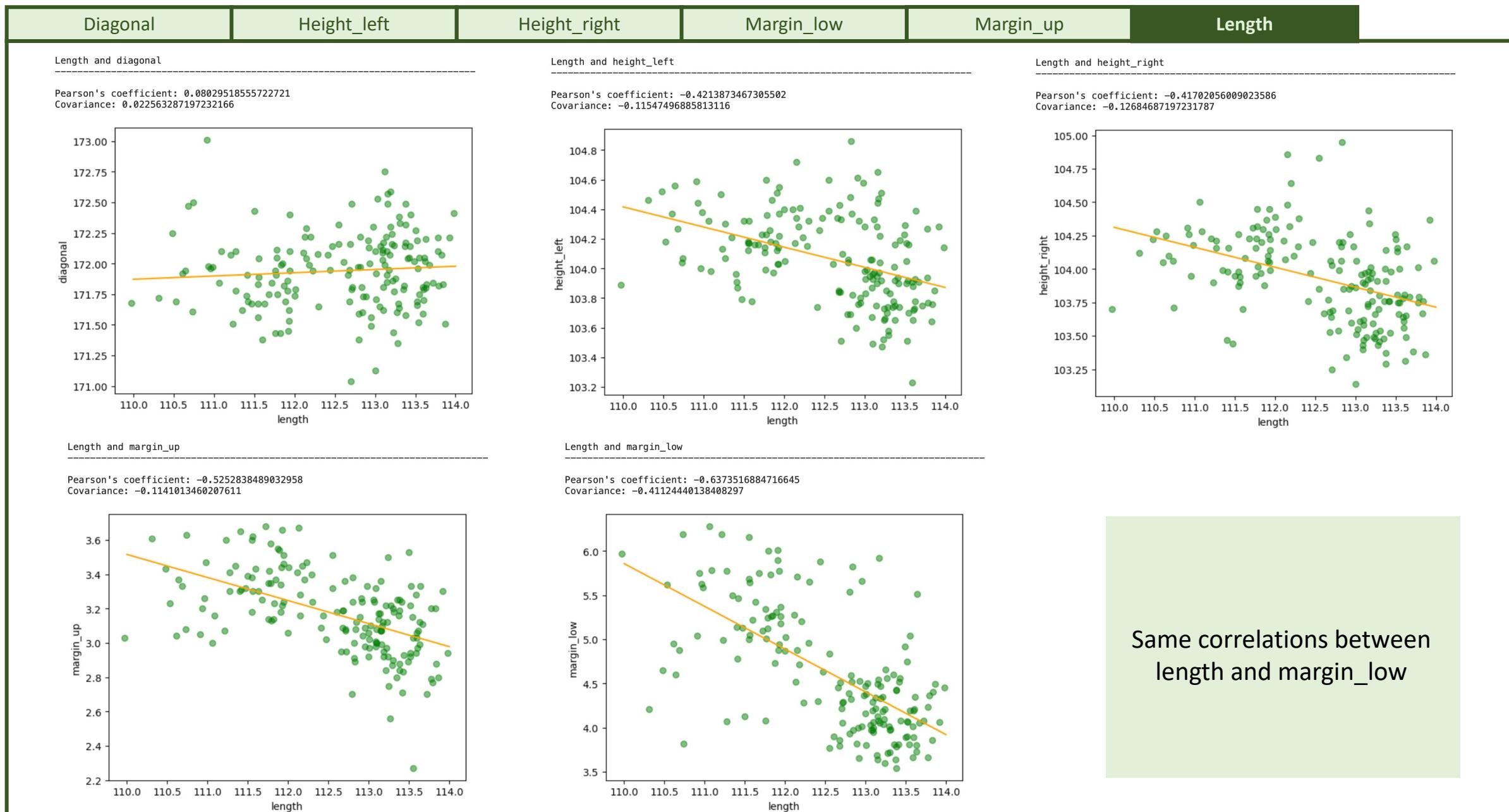


Multivariate analysis – Quantitative variables



Multivariate analysis – Quantitative variables

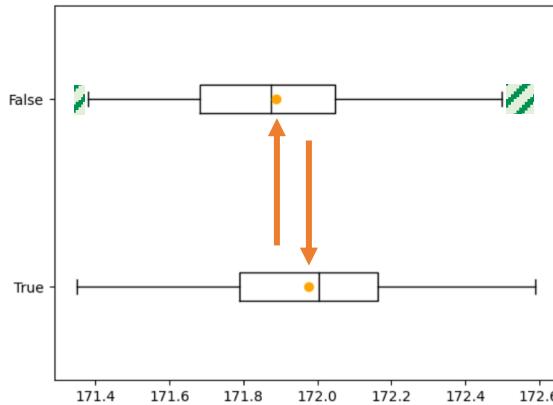
1 - Dataset description



Multivariate analysis – Qualitative Vs Quantitative variables

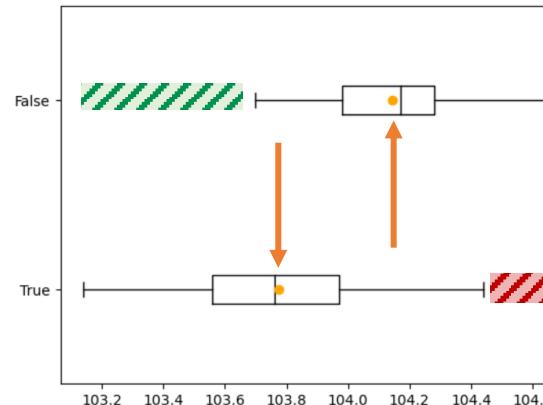
Is_genuine and diagonal

Eta squared: 0.019 --> Small relationship



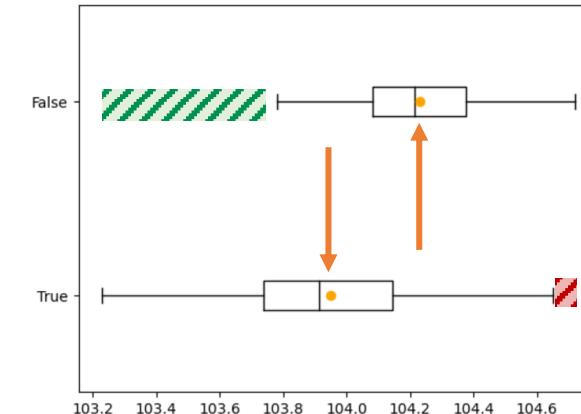
Is_genuine and height_right

Eta squared: 0.304 --> Large relationship



Is_genuine and height_left

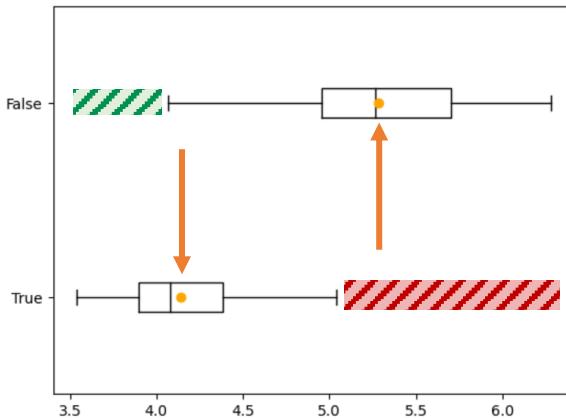
Eta squared: 0.213 --> Large relationship



→ Individually,
none can identify
a counterfeit
notes

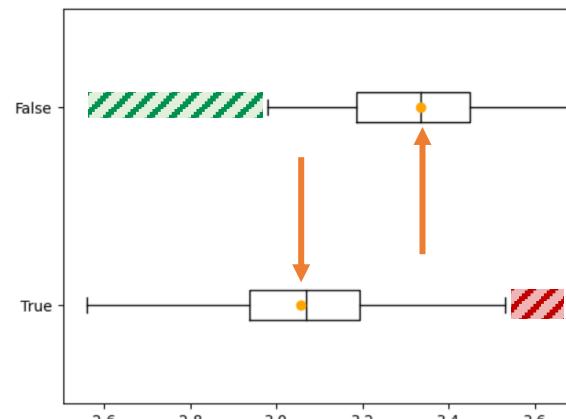
Is_genuine and margin_low

Eta squared: 0.64 --> Large relationship



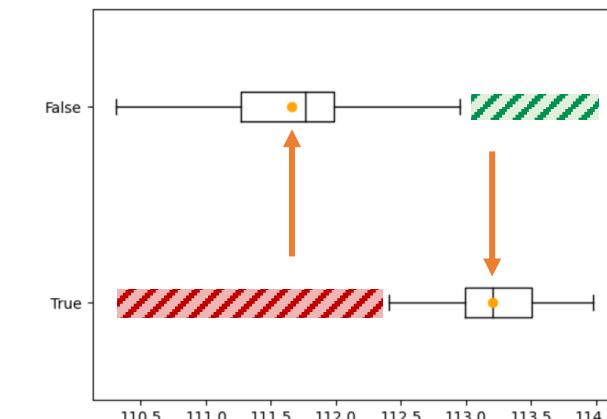
Is_genuine and margin_up

Eta squared: 0.34 --> Large relationship



Is_genuine and length

Eta squared: 0.682 --> Large relationship

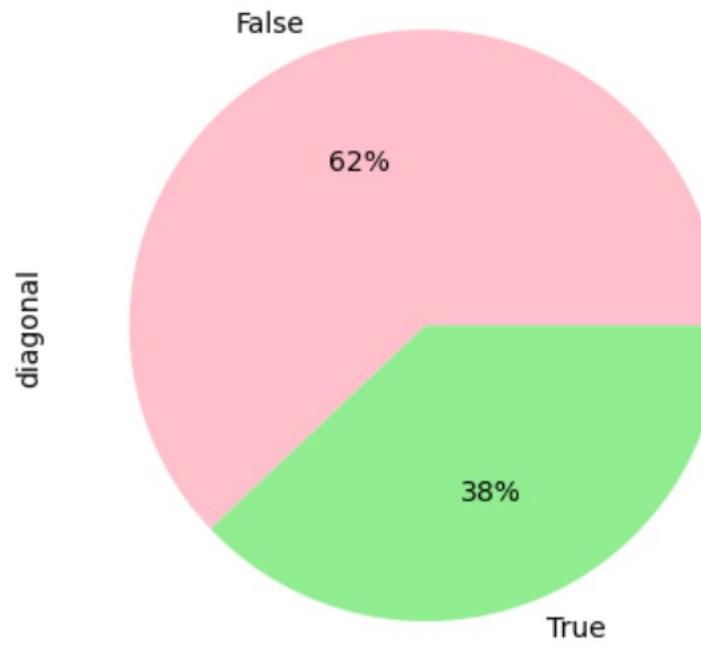


**Outliers
authenticity
distribution**

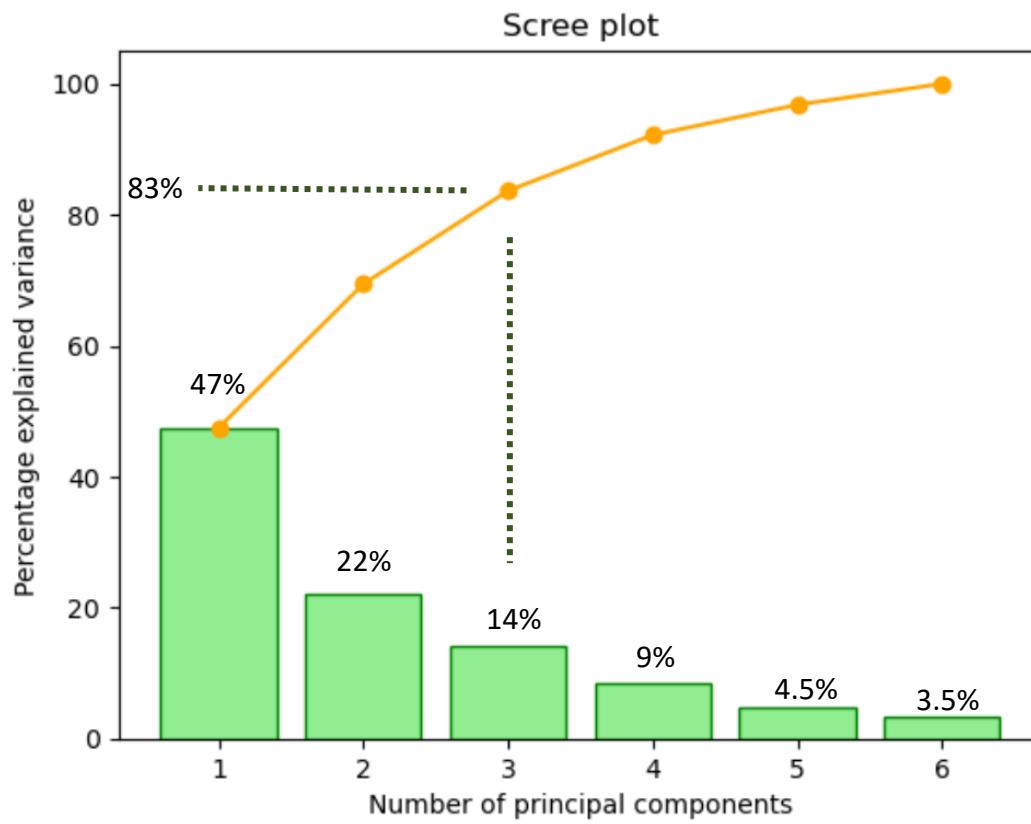
Multivariate analysis – Qualitative Vs Quantitative variables

```
is_genuine  
False    18  
True     11  
Name: diagonal, dtype: int64
```

True and false notes distribution with an outlier value detected

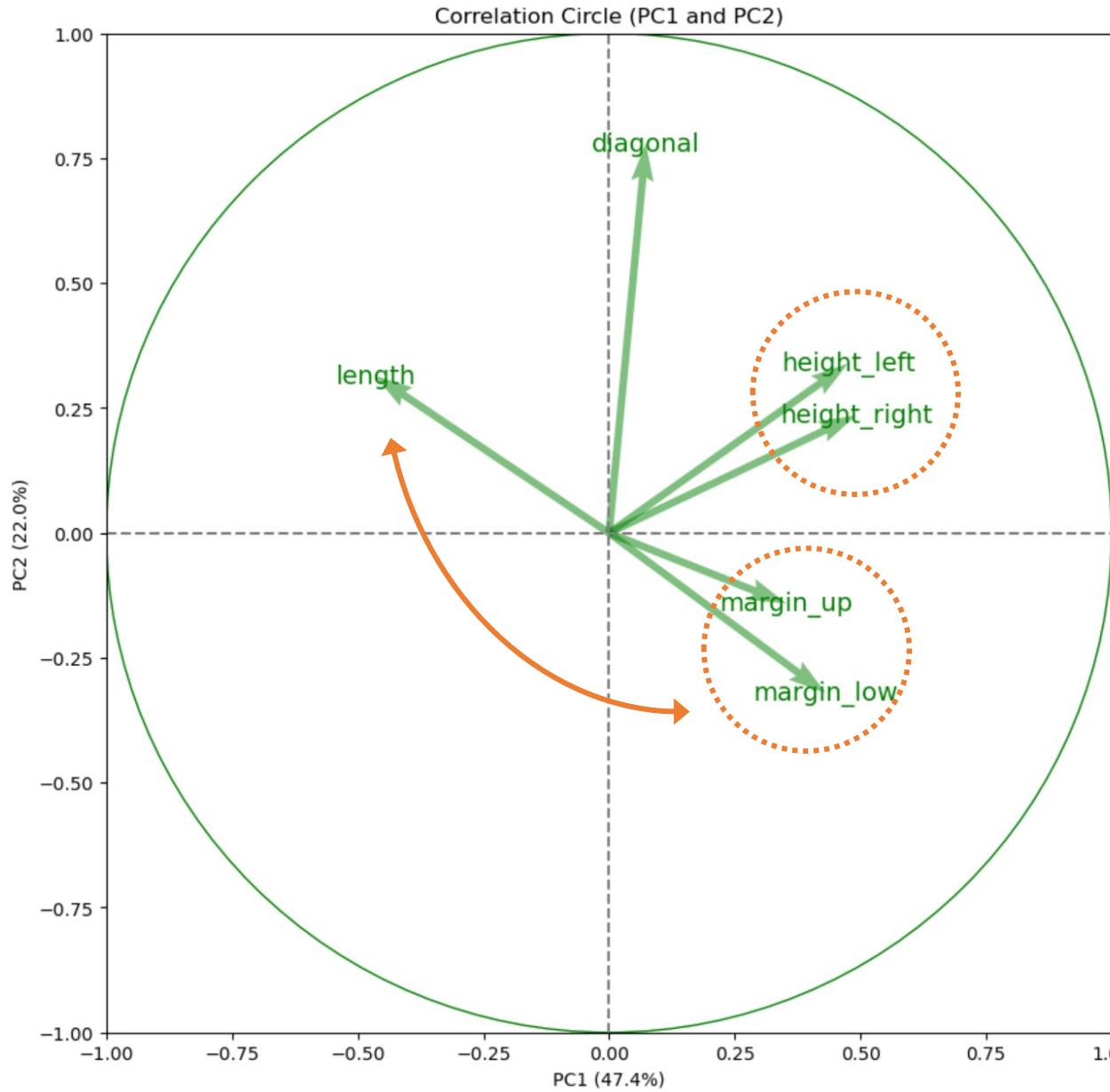


Scree plot



→ Satisfying results for PC1 and PC2 which represent **69%** of the total variance

Correlation circle



On PC1 plane: ↗ length + margins + height

On PC2 plane: ↗ diagonal

Direction and strength shared by:

- Height_left & right
- Margin_up and right

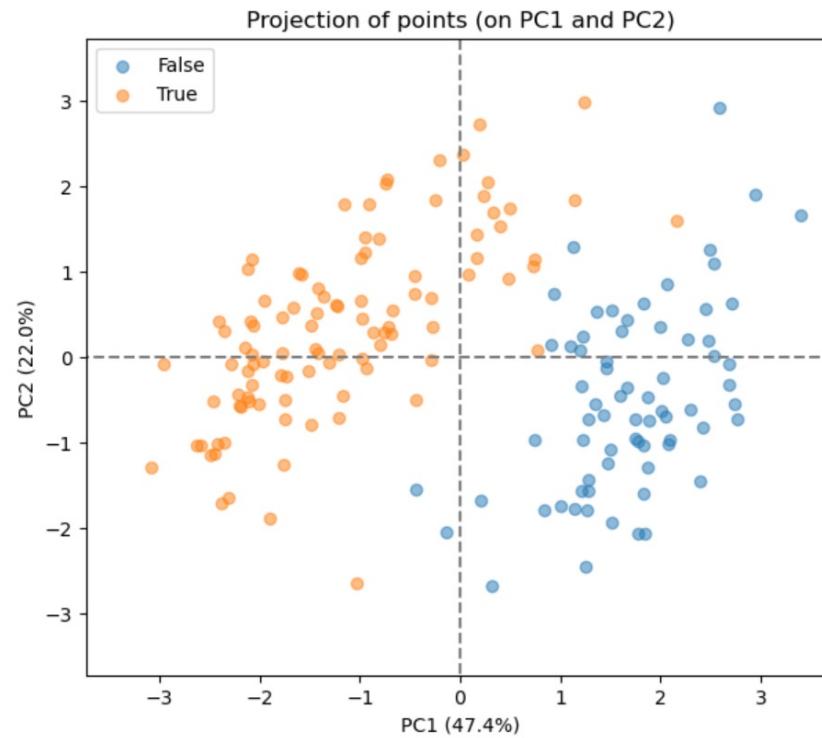
→ Keeping both of them in a model should be challenged

Margin_low and length are negatively correlated

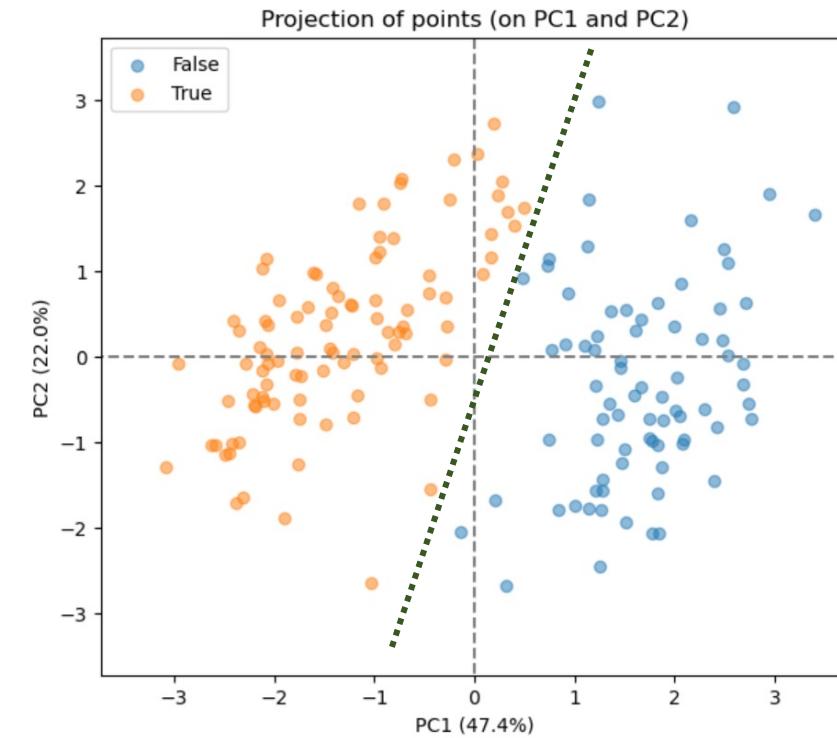
→ Multicollinearity issue

Individual factorial planes – is_genuine Vs K-means

Is_genuine distribution ✓



K-means with 2 groups ☐

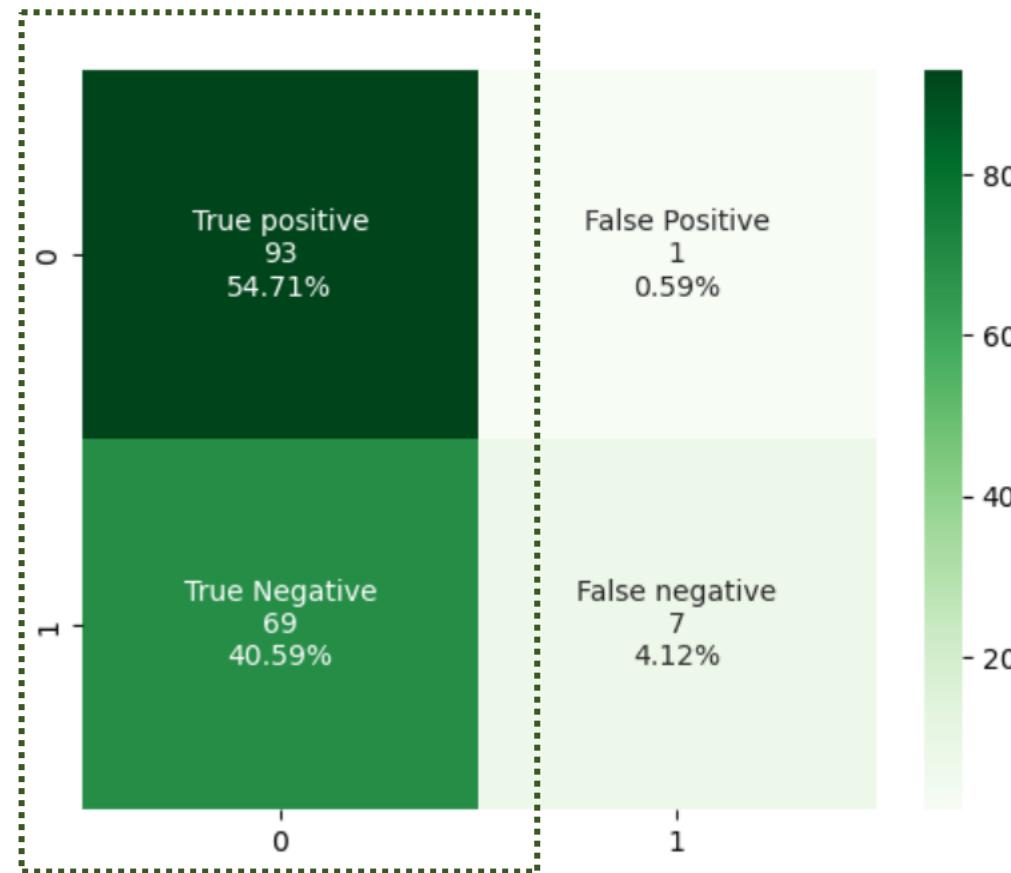


PC1 seams to discriminate well enough the dataset

Contrary to PC2 which cannot be consider at all

Accuracy of k-means

K-means accuracy



K-means classify the authenticity of notes with a **95% accuracy**

Base logistic model

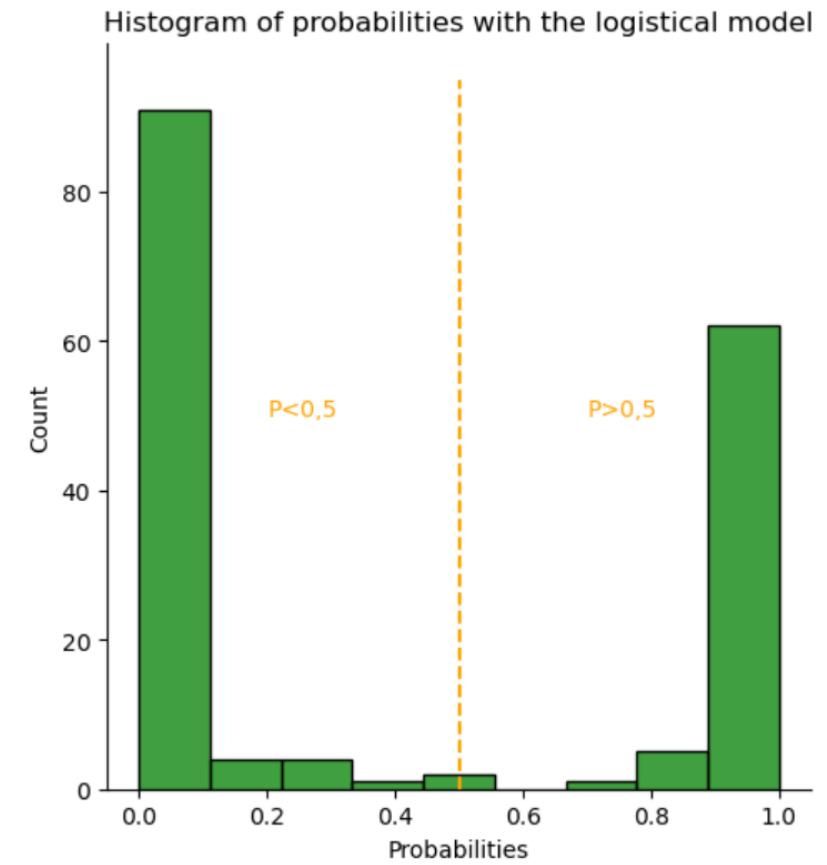
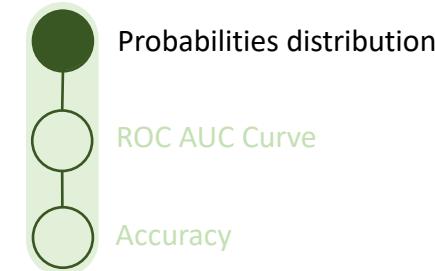
is_genuine ~ diagonal + height_left + height_right + margin_up + [margin_low] + length



Deleted: Collinearity issue prevent smf.logit to finish

Logit Regression Results							
Dep. Variable:	is_genuine	No. Observations:	170	Df Residuals:	164	Df Model:	5
Model:	Logit	Method:	MLE	Date:	Sat, 09 Sep 2023	Pseudo R-squ.:	0.8162
Time:	11:33:18	converged:	True	LL-Null:	-21.171	Log-Likelihood:	-115.17
Covariance Type:	nonrobust			LLR p-value:	1.042e-38		
	coef	std err	z	P> z	[0.025	0.975]	
Intercept	388.8639	258.205	1.506	0.132	-117.209	894.937	
diagonal	-3.2829	1.485	-2.211	0.027	-6.193	-0.373	
height_left	1.4221	2.429	0.585	0.558	-3.339	6.183	
height_right	3.7925	2.024	1.874	0.061	-0.174	7.759	
margin_up	5.6453	2.702	2.090	0.037	0.350	10.940	
length	-3.4256	0.731	-4.686	0.000	-4.858	-1.993	

- **Pseudo R-squared** at 81% = Great amount of outcome variable variance
- **LLR p-value** well below 0.05 so we reject the null hypothesis i.e. our model is relevant
- **P>|z|** for height_left the null hypothesis cannot be rejected, the coefficient may not be relevant in this model
- **Confidence Intervals** for height_left and right contain 0, the null hypothesis cannot be rejected i.e. the coefficient is likely to equal 0



Base logistic model

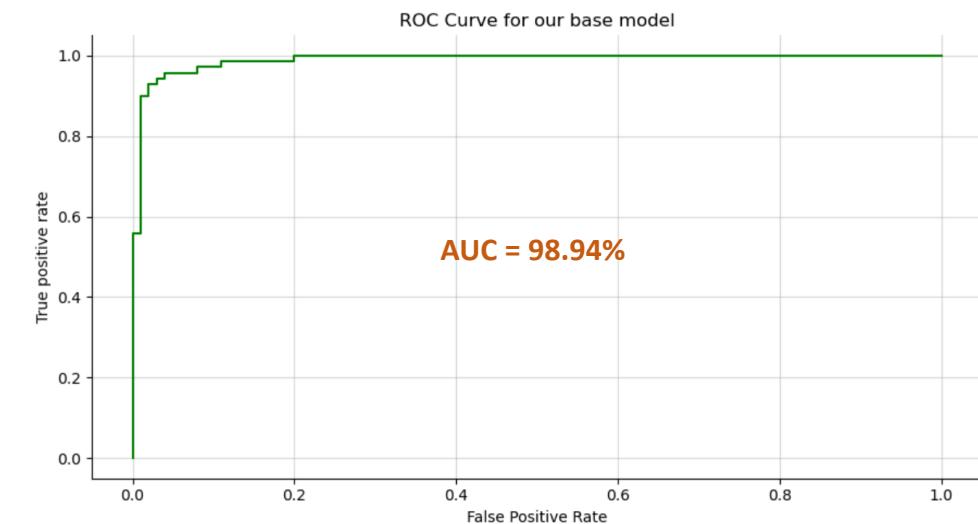
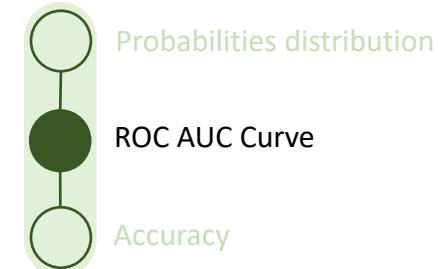
is_genuine ~ diagonal + height_left + height_right + margin_up + [margin_low] + length



Deleted: Collinearity issue prevent smf.logit to finish

Logit Regression Results						
Dep. Variable:	is_genuine	No. Observations:	170			
Model:	Logit	Df Residuals:	164			
Method:	MLE	Df Model:	5			
Date:	Sat, 09 Sep 2023	Pseudo R-squ.:	0.8162			
Time:	11:33:18	Log-Likelihood:	-21.171			
converged:	True	LL-Null:	-115.17			
Covariance Type:	nonrobust	LLR p-value:	1.042e-38			
	coef	std err	z	P> z	[0.025	0.975]
Intercept	388.8639	258.205	1.506	0.132	-117.209	894.937
diagonal	-3.2829	1.485	-2.211	0.027	-6.193	-0.373
height_left	1.4221	2.429	0.585	0.558	-3.339	6.183
height_right	3.7925	2.024	1.874	0.061	-0.174	7.759
margin_up	5.6453	2.702	2.090	0.037	0.350	10.940
length	-3.4256	0.731	-4.686	0.000	-4.858	-1.993

- **Pseudo R-squared** at 81% = Great amount of outcome variable variance
- **LLR p-value** well below 0.05 so we reject the null hypothesis our model is relevant
- **P>|z|** for height_left the null hypothesis cannot be rejected, the coefficient may not be relevant in this model
- **Confidence Intervals** for height_left and right contain 0, the null hypothesis cannot be rejected that is to say the coefficient is likely to equal 0



Base logistic model

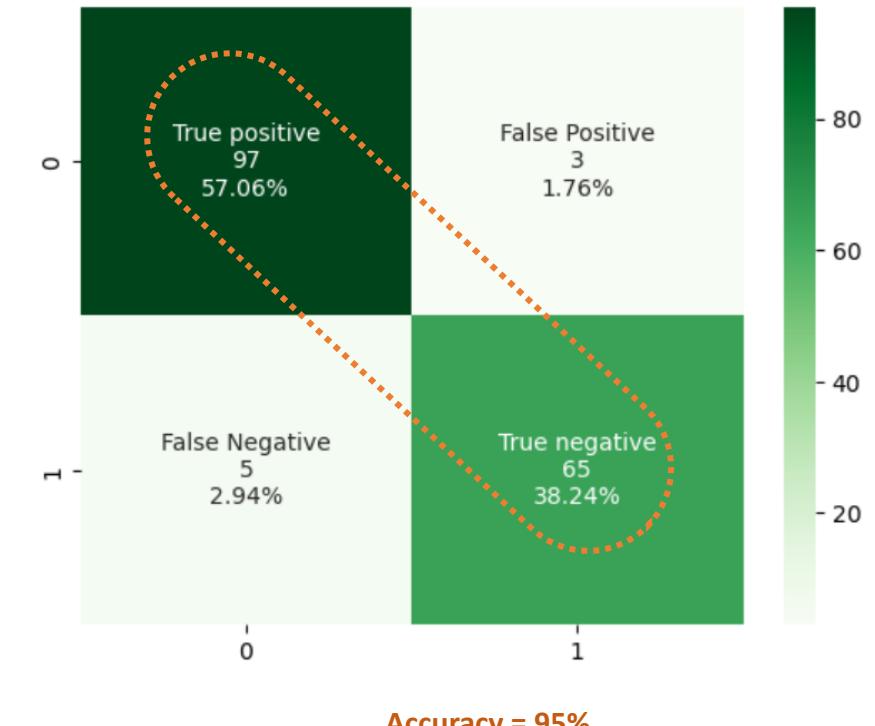
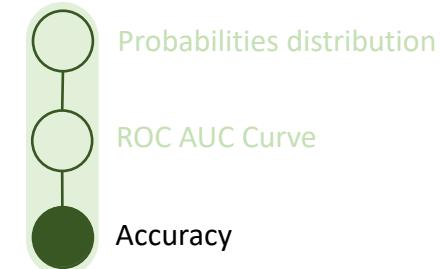
is_genuine ~ diagonal + height_left + height_right + margin_up + [margin_low] + length



Deleted: Collinearity issue prevent smf.logit to finish

Logit Regression Results						
Dep. Variable:	is_genuine	No. Observations:	170			
Model:	Logit	Df Residuals:	164			
Method:	MLE	Df Model:	5			
Date:	Sat, 09 Sep 2023	Pseudo R-squ.:	0.8162			
Time:	11:33:18	Log-Likelihood:	-21.171			
converged:	True	LL-Null:	-115.17			
Covariance Type:	nonrobust	LLR p-value:	1.042e-38			
	coef	std err	z	P> z	[0.025	0.975]
Intercept	388.8639	258.205	1.506	0.132	-117.209	894.937
diagonal	-3.2829	1.485	-2.211	0.027	-6.193	-0.373
height_left	1.4221	2.429	0.585	0.558	-3.339	6.183
height_right	3.7925	2.024	1.874	0.061	-0.174	7.759
margin_up	5.6453	2.702	2.090	0.037	0.350	10.940
length	-3.4256	0.731	-4.686	0.000	-4.858	-1.993

- **Pseudo R-squared** at 81% = Great amount of outcome variable variance
- **LLR p-value** well below 0.05 so we reject the null hypothesis that is to say our model is relevant
- **P>|z|** for height_left the null hypothesis cannot be rejected, the coefficient may not be relevant in this model
- **Confidence Intervals** for height_left and right contain 0, the null hypothesis cannot be rejected that is to say the coefficient is likely to equal 0



Best predictive model

Original Dataset + potential interesting variables

	is_genuine	diagonal	height_left	height_right	margin_low	margin_up	length	height_right2	margin_up2	margin_low2	length2
0	0	171.81	104.86	104.95	4.52	2.89	112.83	11014.5025	8.3521	20.4304	12730.6089
1	0	171.67	103.74	103.70	4.01	2.87	113.29	10753.6900	8.2369	16.0801	12834.6241
2	0	171.83	103.76	103.76	4.40	2.88	113.84	10766.1376	8.2944	19.3600	12959.5456
3	0	171.80	103.78	103.65	3.73	3.12	113.63	10743.3225	9.7344	13.9129	12911.7769
4	0	172.05	103.70	103.75	5.04	2.27	113.55	10764.0625	5.1529	25.4016	12893.6025

Loop on all the logit model tested below



3 relevant indicators gathered

model	Accuracy	R-squared	AUC	Global
-------	----------	-----------	-----	--------

→ Best model overall

0	is_genuine ~ diagonal + height_left + margin_up + margin_low	0.982353	0.948803	0.999143	97.68
1	is_genuine ~ diagonal + height_left + height_right + margin_up + margin_low2	0.976471	0.951875	0.999429	97.59
2	is_genuine ~ diagonal + height_left + height_right + margin_up + margin_low	0.976471	0.951386	0.999429	97.58
3	is_genuine ~ diagonal + height_left + height_right2 + margin_up2 + margin_low	0.976471	0.950692	0.999429	97.55
4	is_genuine ~ diagonal + height_left + height_right + margin_up + length	0.952941	0.816181	0.989429	91.95
5	is_genuine ~ diagonal + height_left + height_right2 + margin_up + length	0.952941	0.816067	0.989429	91.95
6	is_genuine ~ diagonal + height_left + height_right + margin_up + length2	0.952941	0.815979	0.989429	91.94
7	is_genuine ~ diagonal + height_left + margin_up + length	0.947059	0.795263	0.987571	91.00
8	is_genuine ~ margin_up + length	0.952941	0.749024	0.973143	89.17
9	is_genuine ~ length	0.947059	0.723584	0.967714	87.95
10	is_genuine ~ diagonal + height_left + height_right + margin_up	0.876471	0.538301	0.937286	78.40