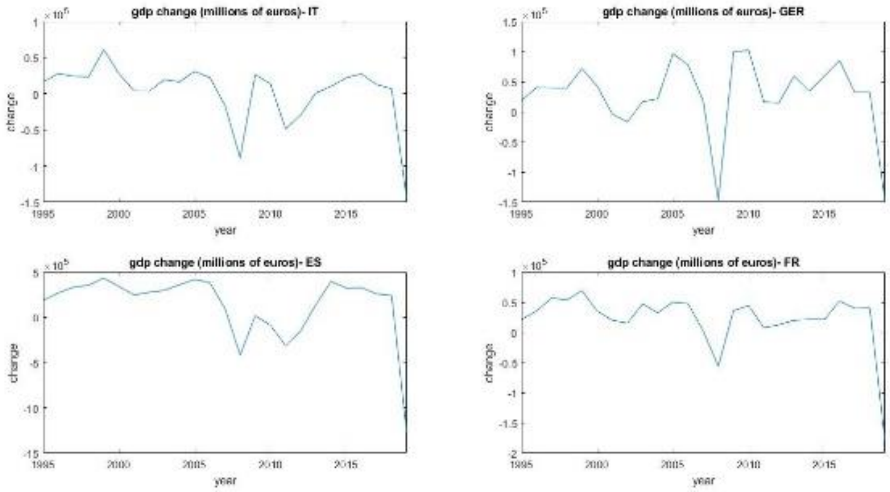
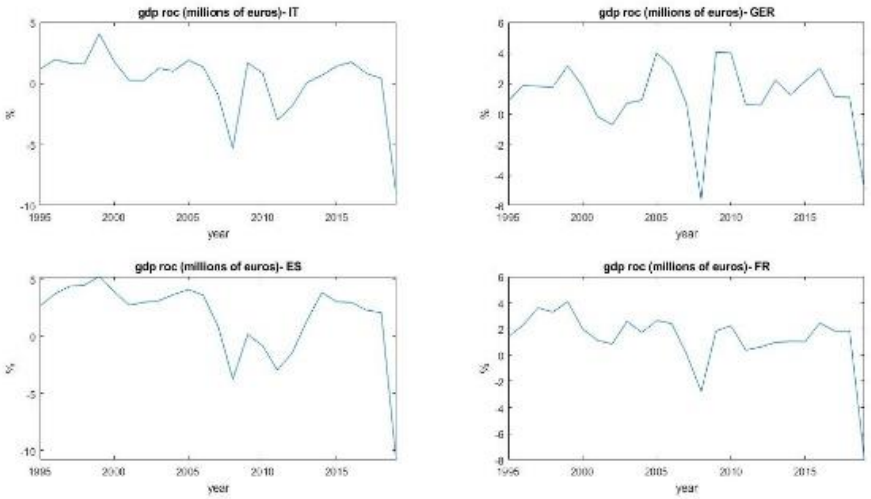
1.

The dataset is a time series from 1995 to 2020, given an annual frequency, of real Global Domestic Product (GDP) for: Italy, Germany, Spain and France.

a)





b)

GDP change and rate of change are the variable of interest, descriptive statistics let us summarize and compare relevant information about the dynamic of these countries, hence economy and wealth.

- Mean, median and mode are **location** statistics which measure the central tendency of GDP.
  The mean, $\bar{x}$, is given by the sum of the GDP changes values per year $GDP_{t\,;\,t-1}$ over the number of observations n, which is equal to the set of years which are considered:

$$\bar{x} = \frac{GDP_{1996\,;\,1995} + GDP_{1997\,;\,1996} + \ldots + GDP_{2020\,;\,2019}}{n} = \frac{\sum_{i=96\,;95}^{20\,;\,19} GDPi}{25}$$

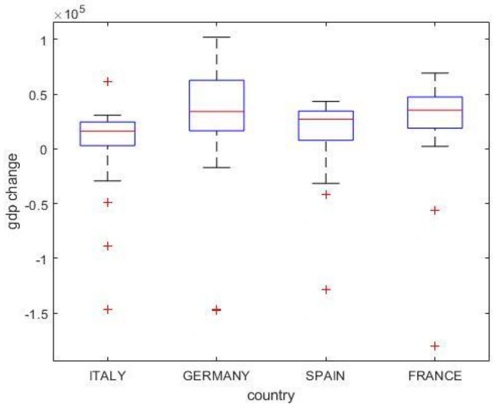
According to rate of changes, the highest GDP rate of change was registered for Spain (1.66%), second position for France (1.26%), third one for Germany (1.16%) and lowest one for Italy (0.22%). The median is the middle observation of an ordered set of observations, whose size is an odd number, or it can be computed by following formula when the size is an even number:

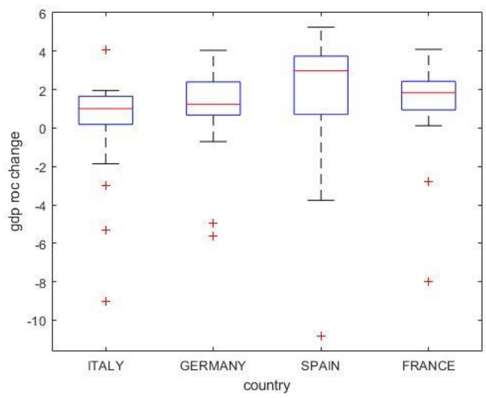
$$GDPmedian = 0.50(n+1)th\ ordered\ position$$

The mode is the most frequent observation, and it may not exist, in fact MatLab returns the lowest value among the observations. Although it is more common for categorical data, it could also be a useful statistic for numerical observations-

	ITALY	GERMANY	SPAIN	FRANCE
Mean	2670,1	28059,61	13858,06	21929,95
Median	16095,2	34084,6	27062,7	35380,6
Roc mean	0,21%	1,16%	1,66%	1,27%
Roc median	1,01%	1,23%	2,97%	1,84%
Mode	-1.46	-1.47	-1.28	-1.80
Roc mode	-8.99%	-5.64%	-10.82%	-7.98%

- Among **dispersion** statistics, which quantify how much an observation set could vary from its mean:
  Range = max – min
  The range is defined as the largest observation value minus the smallest one.
  $interquartile\ range = Q\_3 - Q\_1$
  interquartile range measures the spread in the middle 50% of data by computing the third quartile minus the first quartile and it allows to draw a boxplot which describes the shape of a distribution.





The graphs show, for each country in GDP change and rate of change: the first quartile, which is the lowest black line; the median like red line in the middle; third quartile as higher black line and outliers like red plus symbol, which register maxima and minima.
Variance, S^2, and standard deviation, S, measure the distance between the mean and each of data values:

$$S^2 = \frac{\sum_{i=96\,;\,95}^{20\,;\,19} (GDPi - \bar{x})^2}{n-1}$$

$$S = \sqrt{\frac{\sum_{i=96\,;\,95}^{20\,;\,19}(GDPi - \bar{x})^2}{n-1}}$$

	ITALY	GERMANY	SPAIN	FRANCE
Range	2.07	2.50	1.72	2.49
Interquantile range	2.16	4.61	2.68	2.86
Variance	1 825 462 844.26	3 781 755 590.01	1 388 233 426.75	2 373 903 145.41
Standard deviation	4 2725.43	61 495.98	37 259.01	48 722.72
Rate of change range	13.06	2.50	16.07	12.08
Rate of change interquantile range	1.46%	1.73%	3.03%	1.48%
Rate of change variance	7.04%	5.37%	12.14%	5.52%
Rate of change standard deviation	2.65%	2.31%	3.48%	2.52%

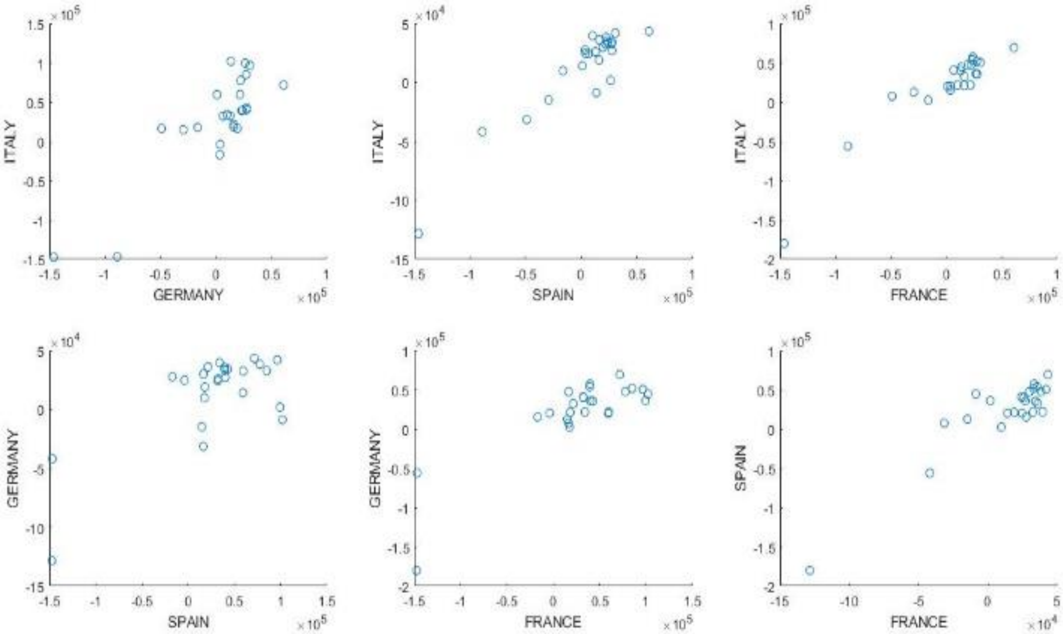
- Computing the difference among these observations is useful to observe dynamics, in absolute value or percentage, that is not hold by GDP sum, because it quantifies a GDP variation which can be plotted.
  Similar dynamics of these countries often imply dependence among them. The linear dependence can be quantified by **correlation** statistics like ρ_{AB} given A and B as countries.
  ρ_{AB} ranges from -1, whose meaning is "negative linear relationship", to +1, whose meaning is "positive linear relationship" and when $\rho_{AB} = 0$ there is no linear relationship but it does not mean there is not any kind of dependence.

$$\rho_{AB} = \frac{S_{AB}}{S_A S_B}$$

S_{AB} is the covariance.

$$S_{AB} = \frac{\sum_{96\,;\,95}^{20\,;\,19} (A_i - \bar{x}_A)(B_i - \bar{x}_B)}{n-1}$$

The scatter plot is a graphical tool for linear dependence representation which uses dots to show values for two different variables, which are GDP changes of countries. The position of each dot on the horizontal and vertical axis indicates values for an individual data point



ρ_{AB}	ITALY	GERMANY	SPAIN	FRANCE
ITALY	1	0.8574	0.9327	0.9387
GERMANY	0.8574	1	0.6632	0.8411
SPAIN	0.9327	0.6632	1	0.9077
FRANCE	0.9387	0.8411	0.9077	1

$\%\rho_{AB}$	ITALY	GERMANY	SPAIN	FRANCE
ITALY	1	0.8430	0.9345	0.9472
GERMANY	0.8430	1	0.6632	0.8184
SPAIN	0.9345	0.6632	1	0.8977
FRANCE	0.9472	0.8184	0.8977	1

c)

If μ is the unknow population mean, so GDP change mean of a country, a **confidence interval** can be computed given its value from 1995 to 2020, so a sample of 25 observations. The confidence interval is a set of number where the population mean lies $1 - \alpha$ times.

$$0 < \alpha < 1 .$$

$$C.I. = [\ \bar{x} - \frac{S}{\sqrt{n}}\ t_{n-1,\alpha/2};\ \bar{x} + \frac{S}{\sqrt{n}}\ t_{n-1,\alpha/2}\ ]$$

$t_{n-1,\alpha/2}$ is the Student's t quantile distribution whose degrees of freedom are $n - 1$ and $\alpha/2$ is the percentile, but if the population standard deviation σ was known, Gaussian quantiles could be used.
The Student-t quantiles are considered for inferential statistic of population mean whether the population variance is unknown, because the results of the ratio between sample mean, whose distribution is Normal, and sample variance, whose one is Chi-Square, is the Student-T distribution.
The variables must be independent and identically distributed (i.i.d.) , whether they are not distributed like a Normal too this result is hold only asymptotically because Central Limit Theorem can be applied. Two variables, for example A and B, are independent if A does not affect B and vice versa. If there is the same likelihood for A and B, they have the same population mean, they are identically distributed.

	Confidence interval		
Country	**90%**	**95%**	**99%**
ITALY	$[-0.493881, 1.322614]$	[-0.681285, 1.510018]	[-1.070430, 1.899163]
GERMANY	[1.718220, 3.304357]	[1.554581, 3.467996]	[1.214784, 3.807793]
SPAIN	[1.187312, 3.572502]	[0.941237, 3.818577]	[0.430260, 4.329553]
FRANCE	[1.881913, 3.490470]	[1.715961, 3.656422]	[1.371361, 4.001021]

Although the population mean is unknown, **hypothesis** test can be used as tool of statistical inference. There are two hypotheses: null (H_0) and alternative (H_1)

H_0 is considered true unless sufficient evidence to the contrary is obtained, if $\mu_0 = 0$ is supposed to be the null hypothesis $\mu_1 \neq 0$ is the alternative one.
H_1 is the hypothesis against which the null hypothesis is tested, and which is considered true till H_0 can be considered true.
Given some hypotheses, pivotal statistic can be considered as: $t = \frac{\bar{x}-\mu_0}{S/\sqrt{n}}$

The test under bilateral hypothesis can be rejected if $t > t_{n-1,\alpha/2}$ or $t < -t_{n-1,\alpha/2}$ and not rejected otherwise.
$I\ type\ error$ is fulfilled if $H_0$is rejected although it is true, and its probability is α, while the
$II\ type\ error$ is for failure of H_0 rejection.
As alternative way to test hypotheses, $\alpha/2$ can be compared to the p-value which is equal to the upper tail probability of the sample mean estimate.
If the p-value is lower than $\alpha/2$, H_0 can be rejected and vice versa.
According to Italy, Germany, Spain and France data whether GDP rate of change mean is assumed to be equal to zero:

	Confidence percentage		
	90%	**95%**	**99%**
ITALY	$t_{IT} = 0.414367 < 1.710882$ p-value=0.341143>0.05 H_0cannot be rejected	$t_{IT} = 0.414367 < 2.063899$ p-value=0.341143>0.025 H_0cannot be rejected	$t_{IT} = 0.414367 < 2.796940$ p-value=0.341143>0.005000 H_0cannot be rejected
GERMANY	$t_{GER} = 2.51128 > 1.710882$ p-value=0.009583<0.05 H_0can be rejected	$t_{GER} = 2.51128 > 2.063899$ p-value=0.009583<0.025 H_0can be rejected	$t_{GER} = 2.511289 < 2.796940$ p-value=0.009583>0.005 H_0cannot be rejected
SPAIN	$t_{ES} = 2.379907 > 1.710882$ p-value=0.012805<0.05 H_0can be rejected	$t_{ES} = 2.379907 > 2.063899$ p-value=0.012805<0.025 H_0can be rejected	$t_{ES} = 2.379907 < 2.796940$ p-value=0.012805>0.005000 H_0cannot be rejected
FRANCE	$t_{FR} = 2.68619 > 1.710882$ p-value=0.006454<0.05 H_0can be rejected	$t_{FR} = 2.686191 > 2.063899$ p-value=0.006454<0.025 H_0can be rejected	$t_{FR} = 2.686191 < 2.796940$ p-value=0.006454>0.005 H_0cannot be rejected