# INDEPENDENCE

## Durbin-Watson Test

$H_0$: not autocorrelations

$H_1$: autocorrelations

$$D = \frac{\sum_{i=2}^{n} (e_i - e_{i-1})^2}{\sum_{i=1}^{n} e_i^2}$$

THE RANGE IS $0 \leq D \leq 4$

$D = 2$ $H_0$ IS TRUE

$D < 2$ TOWARDS A POSITIVE AUTOCORRELATION

$D > 2$ TOWARDS A NEGATIVE AUTOCORRELATION

# if D<2

$H_0$: not autocorrelation

$H_1$: positive autocorrelation

On Durbin-Watson table we read the values of $d_L$ e $d_U$ (where“**n**” is the sample size and “**k**’” the number of explanatory variables and $\alpha$ =0.05)

$H_1$ if $D < d_L$

$H_0$ if $D > d_U$

| False $H_0$ | Incoclusive | True $H_0$ |
|---|---|---|

$r = 1$ 0 $1 < r < 0,5$ $d_L$ $0 < r < 0,5$ $d_U$ $r \to 0$ 2 $r = 0$

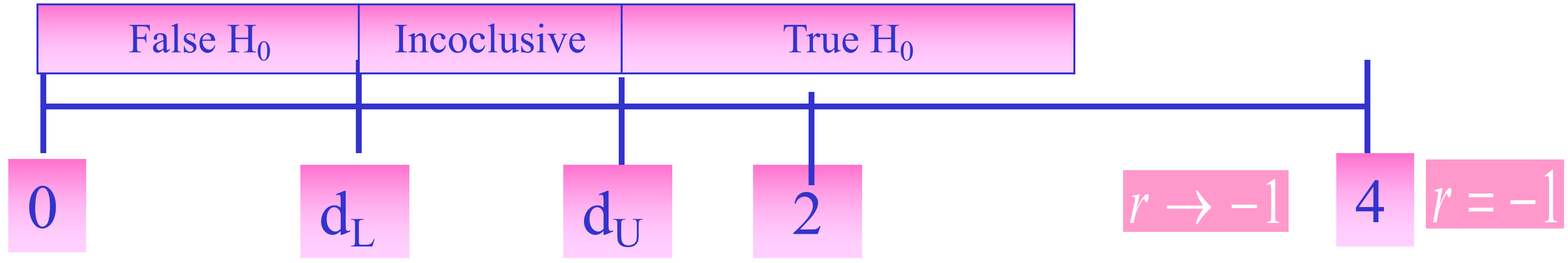
If D>2

$H_0$: not autocorrelation

$H_1$: negative autocorrelation

$H_1$ if (4-D) < $d_L$

$H_0$ if (4-D) > $d_U$



# Durbin-Watson

| Y | ei | $ei^2$ | $(e_i-e_{i-1})^2$ | Y* |
|---|---|---|---|---|
| 950 | -132.37 | 17522.03 | - | 1082.37 |
| 1600 | -121.37 | 14730.87 | 121.00 | 1721.37 |
| 1200 | -132.65 | 17594.91 | 127.13 | 1332.65 |
| 1500 | 10.80 | 116.62 | 20576.47 | 1489.20 |
| 950 | 8.21 | 67.39 | 6.71 | 941.79 |
| 1700 | -58.65 | 3439.33 | 4469.59 | 1758.65 |
| 1650 | 263.04 | 69189.62 | 103481.24 | 1386.96 |
| 935 | -15.31 | 234.42 | 77478.72 | 950.31 |
| 875 | -47.62 | 2267.74 | 1043.94 | 922.62 |
| 1150 | -45.26 | 2048.54 | 5.57 | 1195.26 |
| 1400 | 51.38 | 2639.82 | 9339.29 | 1348.62 |
| 1650 | 104.35 | 10889.80 | 2806.35 | 1545.65 |
| 2300 | 8.85 | 78.40 | 9120.25 | 2291.15 |
| 1800 | 164.89 | 27190.10 | 24348.48 | 1635.11 |
| 1400 | -28.50 | 812.01 | 37399.69 | 1428.50 |
| 1450 | -31.75 | 1007.80 | 10.56 | 1481.75 |
| 1100 | -403.05 | 162445.92 | 137863.69 | 1503.05 |
| 1700 | 182.04 | 33140.09 | 342330.31 | 1517.96 |
| 1200 | -201.87 | 40751.82 | 147390.73 | 1401.87 |
| 1150 | 18.64 | 347.42 | 48624.66 | 1131.36 |
| 1600 | -26.59 | 706.80 | 2045.30 | 1626.59 |
| | 365.28 | 133428.89 | 153558.18 | 1284.72 |
| | 218.80 | 47875.28 | 21454.93 | 981.20 |
| | -442.12 | 195470.80 | 436821.86 | 1242.12 |
| | 294.88 | 86953.74 | 543169.00 | 1455.12 |
| | | 870950.16 | 2123593.63 | |

$$D = \frac{\sum_{i=2}^{n}(e_i - e_{i-1})^2}{\sum_{i=1}^{n} e_i^{\,2}} = \frac{21235993.63}{870950.16} = 2.44$$

# Durbin-Watson table

| $n$ | $k'=1$ | | $k'=2$ | | $k'=3$ | | $k'=4$ | | $k'=5$ | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $d_L$ | $d_U$ | $d_L$ | $d_U$ | $d_L$ | $d_U$ | $d_L$ | $d_U$ | $d_L$ | $d_U$ |
| 15 | 1.077 | 1.361 | 0.945 | 1.543 | 0.814 | 1.750 | 0.685 | 1.977 | 0.562 | 2.220 |
| 16 | 1.106 | 1.371 | 0.982 | 1.539 | 0.857 | 1.728 | 0.734 | 1.935 | 0.615 | 2.157 |
| 17 | 1.133 | 1.381 | 1.015 | 1.536 | 0.897 | 1.710 | 0.779 | 1.900 | 0.664 | 2.104 |
| 18 | 1.158 | 1.392 | 1.046 | 1.535 | 0.933 | 1.696 | 0.820 | 1.872 | 0.710 | 2.060 |
| 19 | 1.180 | 1.401 | 1.075 | 1.535 | 0.967 | 1.685 | 0.859 | 1.848 | 0.752 | 2.022 |
| 20 | 1.201 | 1.411 | 1.100 | 1.537 | 0.998 | 1.676 | 0.894 | 1.828 | 0.792 | 1.991 |
| 21 | 1.221 | 1.420 | 1.125 | 1.538 | 1.026 | 1.669 | 0.927 | 1.812 | 0.828 | 1.964 |
| 22 | 1.240 | 1.429 | 1.147 | 1.541 | 1.053 | 1.664 | 0.958 | 1.797 | 0.863 | 1.940 |
| 23 | 1.257 | 1.437 | 1.168 | 1.543 | 1.078 | 1.660 | 0.986 | 1.786 | 0.895 | 1.919 |
| 24 | 1.273 | 1.446 | 1.188 | 1.546 | 1.101 | 1.657 | 1.013 | 1.775 | 0.925 | 1.902 |
| 25 | 1.288 | 1.454 | 1.206 | 1.550 | 1.123 | 1.654 | 1.038 | 1.767 | 0.953 | 1.886 |
| 26 | 1.302 | 1.461 | 1.224 | 1.553 | 1.143 | 1.652 | 1.062 | 1.759 | 0.979 | 1.873 |
| 27 | 1.316 | 1.468 | 1.240 | 1.556 | 1.162 | 1.651 | 1.083 | 1.753 | 1.004 | 1.861 |
| 28 | 1.328 | 1.476 | 1.255 | 1.560 | 1.181 | 1.650 | 1.104 | 1.747 | 1.028 | 1.850 |
| 29 | 1.341 | 1.483 | 1.270 | 1.563 | 1.198 | 1.650 | 1.124 | 1.743 | 1.050 | 1.841 |
| 30 | 1.352 | 1.489 | 1.284 | 1.567 | 1.214 | 1.650 | 1.143 | 1.739 | 1.070 | 1.833 |

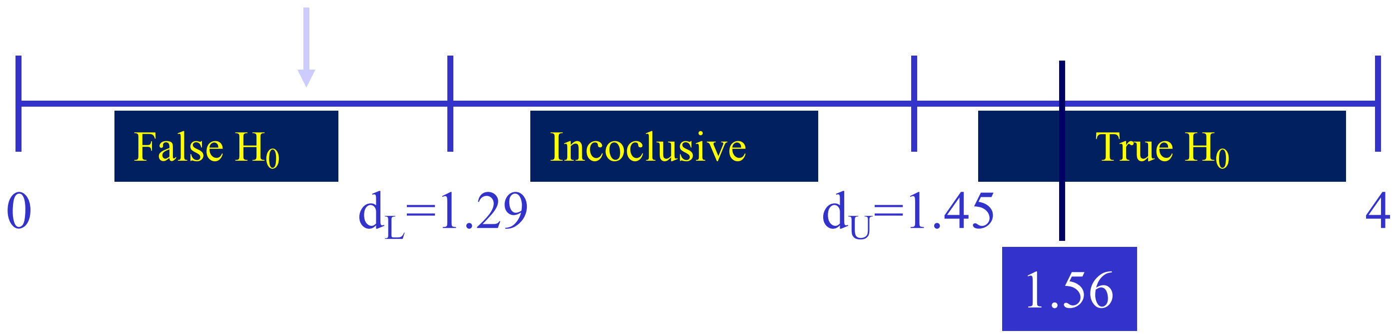
With one regressor and $\alpha=0.05$

# Durbin-Watson test

- D=2.44
- n = 25, k = 1
- $d_L = 1.29$ and $d_u = 1.45$
- (4-D) = (4-2.44 )=1.56 > $d_u$ = 1.45, $H_0$ is true and does not exist negative autocorrelation

True $H_0$ since

(4-D) > $d_U$



# Normality

- Histogram
- Box-plot
- Normal Probability Plot (PP PLOT) or QQ Plot
- Statistical Tests (Jarque-Bera, Kolmogorov-Smirnov, Shapiro-Wilk)

# Histogram and normality tests

The histogram is a graph which displays the probability distribution of a character (residuals). In this context we want know if the distribution of residuals can be approximated by a normal curve graphically. It is also possible to calculate some statistical tests of hypotheses to be tested , for which:

$H_0$: normal

$H_1$: not normal

Statistical tests:

Kolmogorov –Smirnov

Jarque – Bera

Shapiro – Wilk

# Normality: Kolmogorov-Smirnov (D) test

$$H_0 : F(X) = F_0(X)$$
$$H_1 : F(X) \neq F_0(X)$$

$$D = sup\,|\,\hat{F}_n(X) - F_0(X)\,|$$
$$where$$
$$\hat{F}_n = \frac{1}{n}\sum_i I_{(Xi)\leq X}$$

F(X) is the cumulative distribution function (cdf) of X
In H0 we refer to the normal cdf.
Ho is true if the cdfs are similar and, then, the difference is close to zero.

# *Normality: Shapiro-Wilk (W) test*

$$\begin{cases} H_0: \textit{normal distribution} \\ H_1: \textit{otherwise} \end{cases}$$

$$W = \frac{\left[\sum_{i=1}^{n} a_i x_i\right]^2}{\sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2}$$

$a_i$ is weight which dependens on sample size

*0<W<1*,respectively, false and true H0

The index W is however highly asymmetric.

The threshold values of such statistic are reported in special tables

## Normality: Jarque-Bera(JB) test

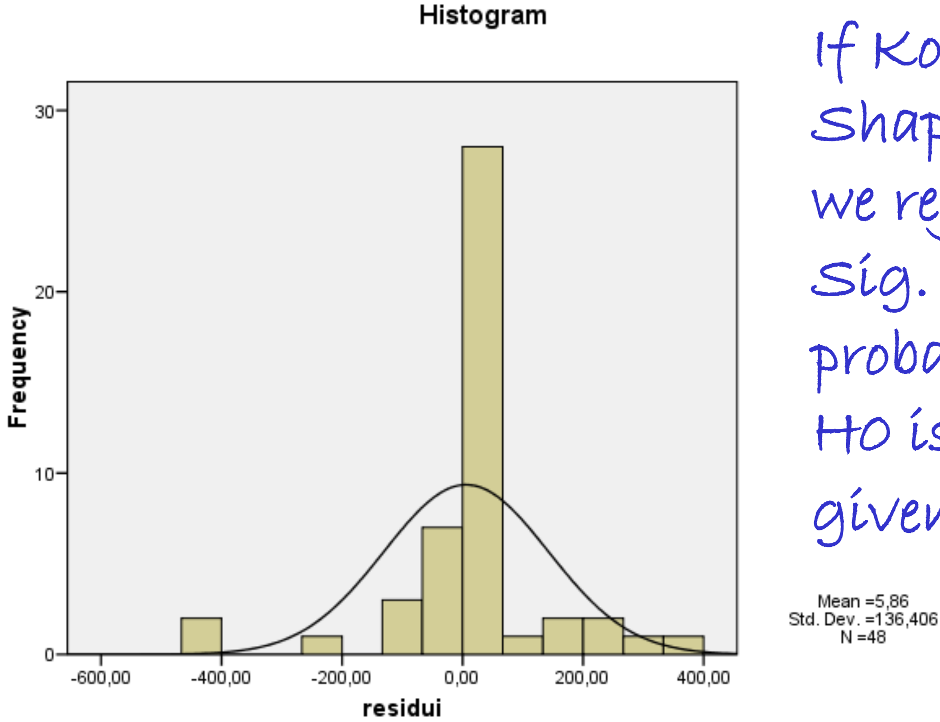
$$H_0 : normal$$

$$H_1 : otherwise$$

$$JB = n\left[\frac{S^2}{6} + \frac{(K-3)^2}{24}\right]$$

- S = Skewness
- K = Kurtosis
- JB is distributed as $\chi^2$ with 2 d.o.f.
- JB> $\chi_2^2$ false $H_0$

# NORMALITY: HISTOGRAM AND TESTS




If Kolmogorov – Smirnov and Shapiro Wilk are lower than 1, we reject H0.
Sig. is the p-value, or the probability of rejecting H0, when H0 is true (has to be greater of a given α, to accept H0).

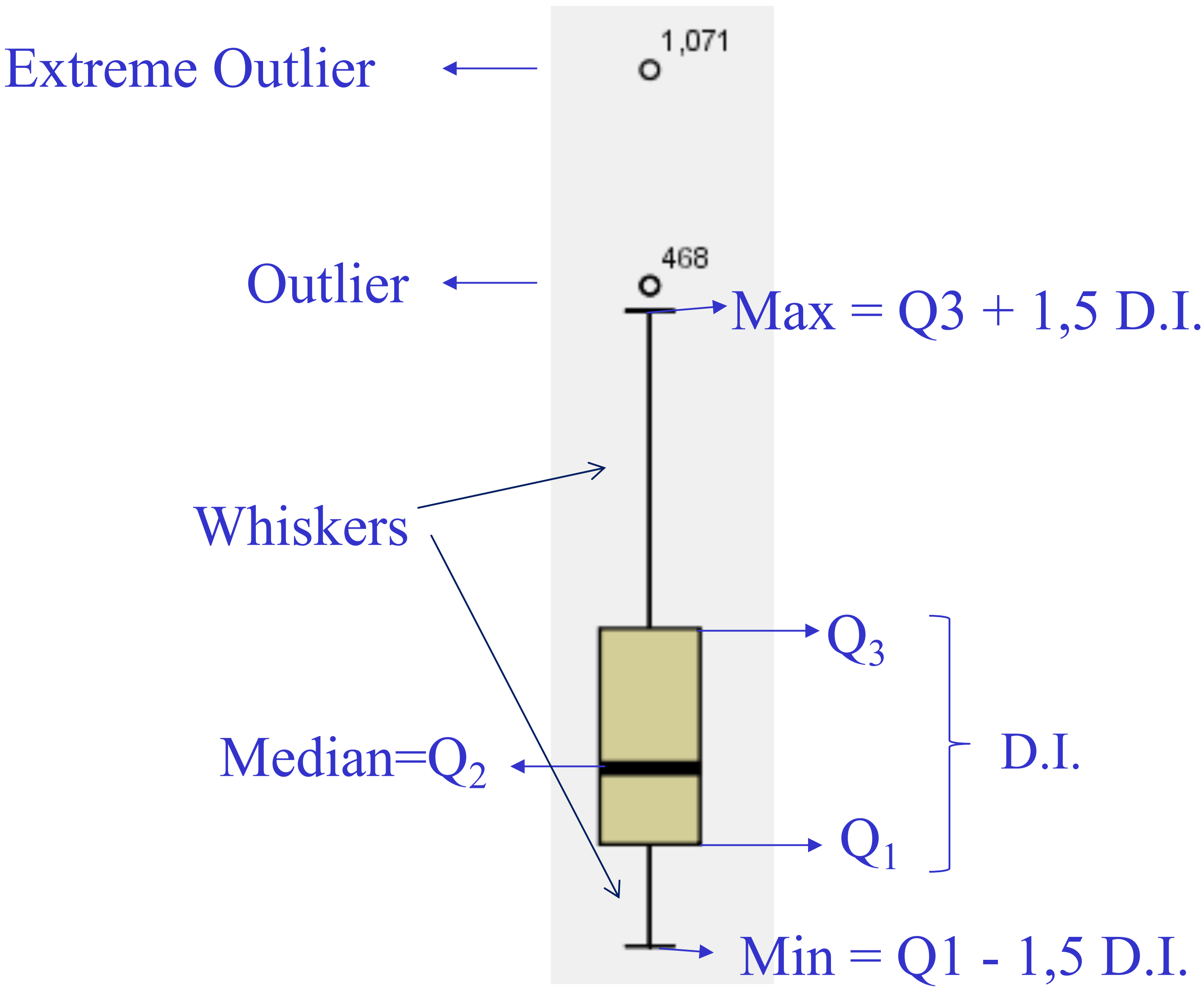
**Tests of Normality**

| | Kolmogorov-Smirnov[a] | | | Shapiro-Wilk | | |
|---|---|---|---|---|---|---|
| | Statistic | df | Sig. | Statistic | df | Sig. |
| residui | ,253 | 48 | ,000 | ,811 | 48 | ,000 |

a. Lilliefors Significance Correction

Not Normal!

# Normality: Box-plot



# BOX PLOT:
# OUTLIERS

*OUTLIERS*° → 

$$Q_3 + STEP \leq \quad e_i \leq Q_3 + 2STEP$$

*or*

$$Q_1 - 2STEP \leq \quad e_i \leq Q_1 - STEP$$

**Where STEP=1.5*(Q3-Q1) = 1.5*D.I.**

*EXTREMEOULIERS* * →

$$e_i < Q_1 - 2STEP$$

*or*

$$e_i > Q_3 + 2STEP$$

| Residui | ni | Ni |
|---|---|---|
| -442.12 | 1 | 1 |
| -403.05 | 1 | 2 |
| -201.87 | 1 | 3 |
| -132.65 | 1 | 4 |
| -132.37 | 1 | 5 |
| -121.37 | 1 | 6 |
| -58.65 | 1 | 7 |
| -47.62 | 1 | 8 |
| -45.26 | 1 | 9 |
| -31.75 | 1 | 10 |
| -28.50 | 1 | 11 |
| -26.59 | 1 | 12 |
| -15.31 | 1 | 13 |
| 8.21 | 1 | 14 |
| 8.85 | 1 | 15 |
| 10.80 | 1 | 16 |
| 18.64 | 1 | 17 |
| 51.38 | 1 | 18 |
| 104.35 | 1 | 19 |
| 164.89 | 1 | 20 |
| 182.04 | 1 | 21 |
| 218.80 | 1 | 22 |
| 263.04 | 1 | 23 |
| 294.88 | 1 | 24 |
| 365.28 | 1 | 25 |

# EXAMPLE

RANK RESIDUALS!!!

$$position \quad of \ Q1 = 25/4 = 6.25$$
$$Q1 = -121.37$$

$$position \ of \quad Q3 = 25*3/4 = 18.75$$
$$Q3 = 104.35$$

$$STEP = 1.5*(Q3-Q1) = 1.5*140.12 = 210.18$$

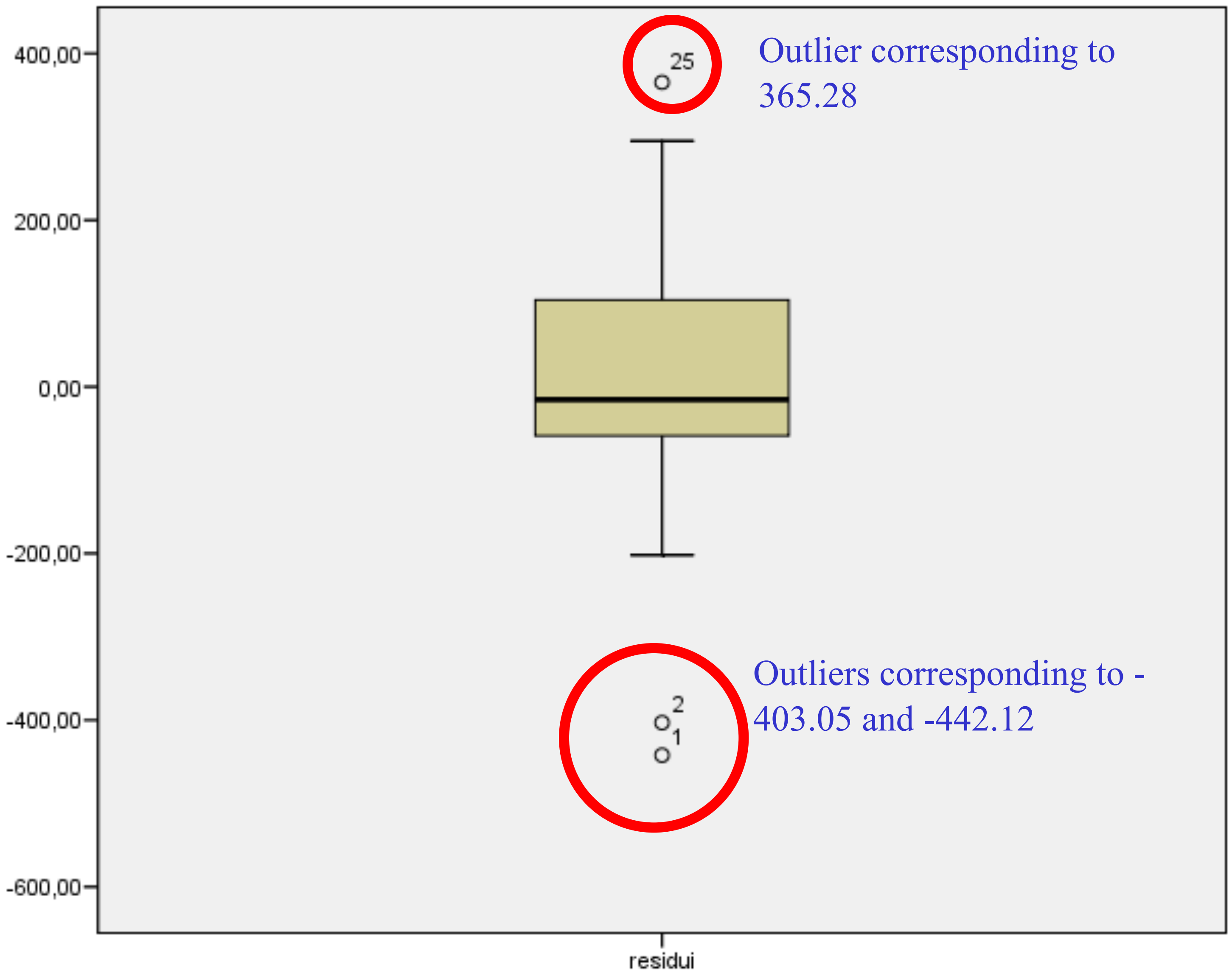
## Outlier

$$Q_3 + STEP \leq \quad e_i \leq Q_3 + 2STEP \rightarrow 104.35 + 210.18 \leq \quad e_i \leq \ 104.35 + 420.36$$
$$314.53 \leq \ e_i \leq \ 524.71$$
$$or$$
$$Q_1 - 2STEP \leq \ e_i \leq Q_1 - STEP \rightarrow -121.37 - 420.36 \leq \ e_i \leq \ -121.37 - 210.18$$
$$-541.73 \leq \ e_i \leq \ -331.55$$

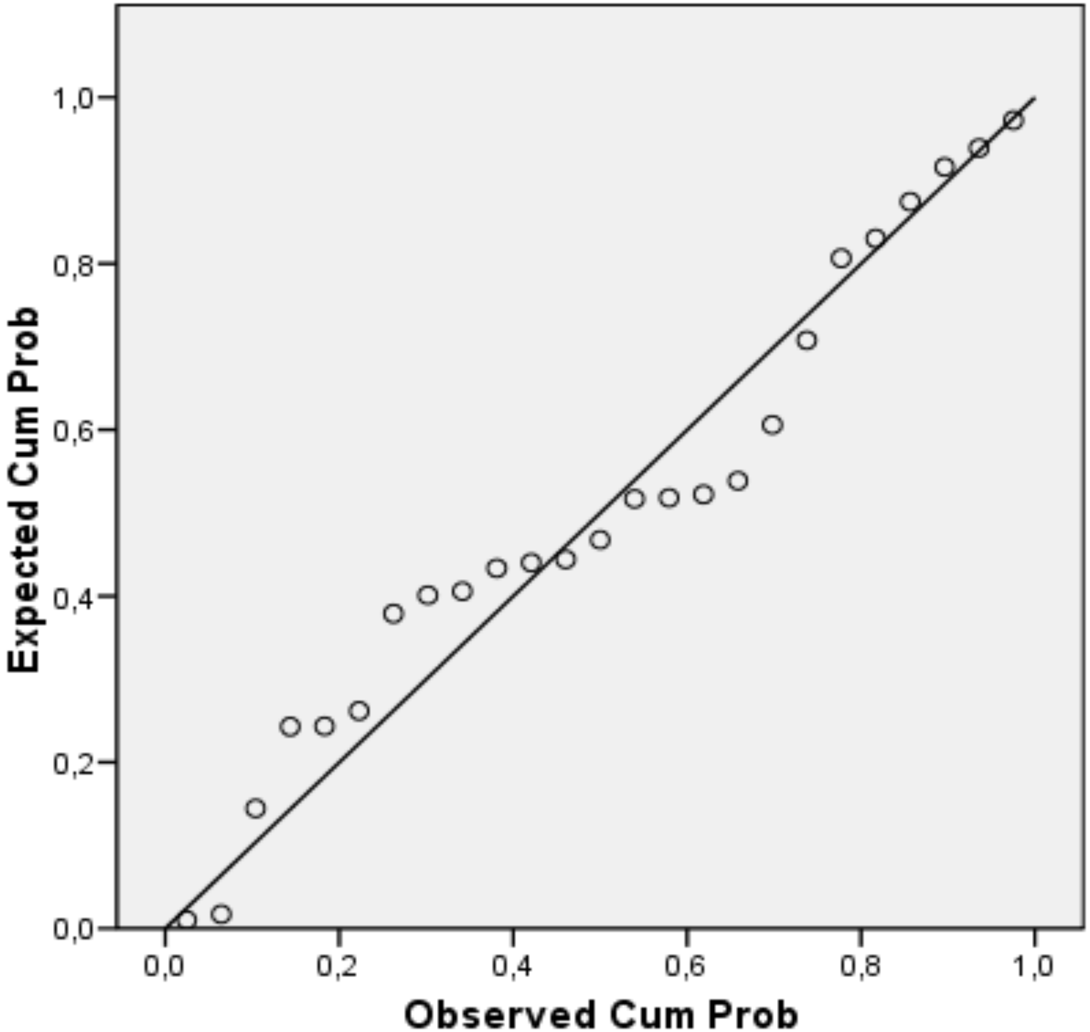
## Limits

$$e_i < Q_1 - 2STEP \rightarrow e_i < -121.37 - 420.36 \rightarrow e_i < -541.73$$
$$or \qquad or \qquad or$$
$$e_i > Q_3 + 2STEP \rightarrow e_i > 104.35 + 420.36 \rightarrow e_i > 524.71$$

# BOX-PLOT

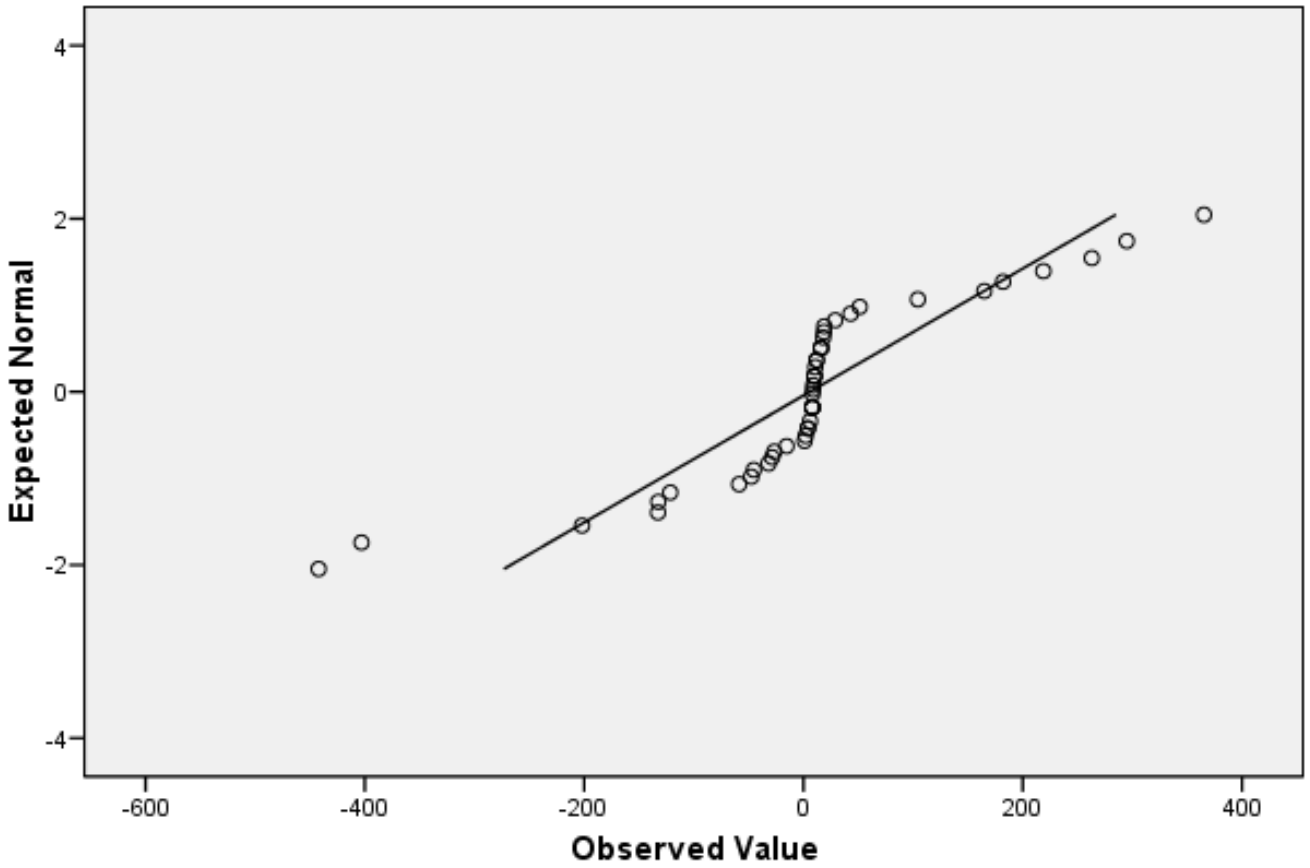

400,00
200,00
0,00
-200,00
-400,00
-600,00
25
2
1
residui
Outlier corresponding to 365.28
Outliers corresponding to -403.05 and -442.12

# P-P PLOT



1,0
0,8
0,6
0,4
0,2
0,0
Expected Cum Prob
0,0
0,2
0,4
0,6
0,8
1,0
Observed Cum Prob

# Q-Q PLOT

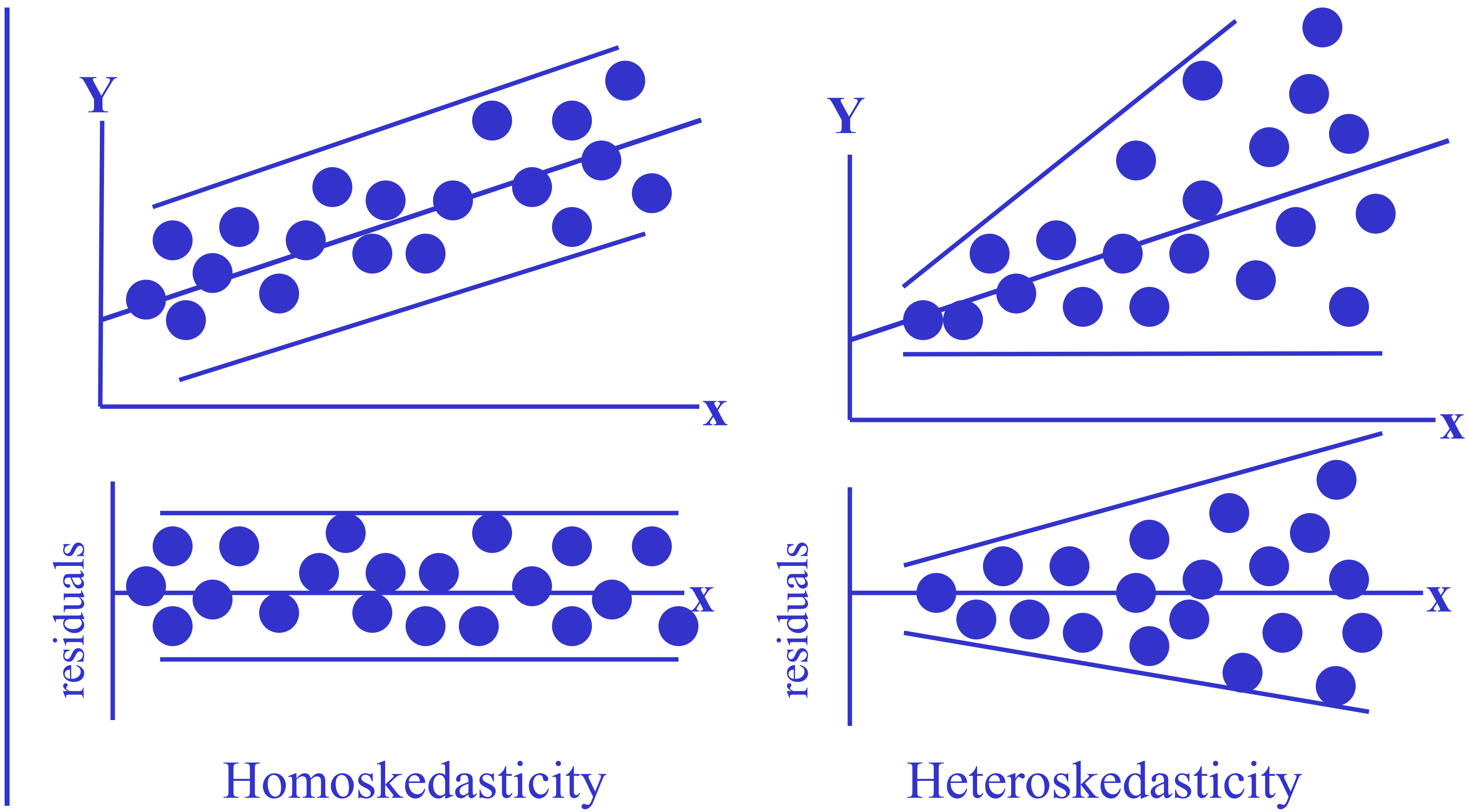


# HOMOSKEDASTICITY

FINITE AND CONSTANT VARIANCE OF RESIDUALS WITH RESPECT TO REGRESSORS

- Scatter-plot of standardized residuals and regressor (X)
- Goldfeld- Quandt test

# Scatter plot: residuals vs $x_i$



# HOMOSKEDASTICITY:
# Goldfeld- Quandt Test

In the case of heterosckedasticity:

- Rank Y with respect to X
- Omit "c" central observations
- Two regressions on the first and the last (n-c)/2 observations
- SSE1 and SSE2 are the sum of square residuals for the two groups and their ratio is $F_{(n-c-2k)/2}$ and we reject H0 (Homoskedasticity) if $F_{(n-c-2k)/2} > F_{(n-c-2k)/2,\ \alpha}$
- In this last case, we can adopt a linear transformation of the dependent variable (Y) or we can apply the method of weighted least squares where the weights are inversely proportional to the error variance