# Chow Test

**Dataset Background:**

This dataset contains information on the sales price of houses and the interest levied between the years 1996 to 2017. The dataset witnesses a global boom in between the years 2007 and 2008 due to change in interest levied.

**Dataset Glimpse:**

| year | interest | sales |
|---|---|---|
| 1996 | 57.5 | 831 |
| 1997 | 65.4 | 893.5 |
| 1998 | 59.7 | 980.5 |
| 1999 | 86.1 | 1098.7 |
| 2000 | 93.4 | 1205.7 |
| 2002 | 100.3 | 1307.3 |
| 2002 | 93 | 1546.3 |
| 2003 | 87.9 | 1801.4 |
| 2004 | 107.8 | 2006.6 |
| 2005 | 123.3 | 2467.1 |
| 2006 | 153.8 | 2790.4 |
| 2007 | 191.8 | 3012.5 |
| 2008 | 199.5 | 2890.9 |
| 2009 | 168.7 | 2902.5 |
| 2010 | 222 | 2954.6 |
| 2011 | 189.3 | 3279.8 |
| 2012 | 187.5 | 3590.4 |
| 2013 | 142 | 3802 |
| 2014 | 155.7 | 4075.9 |
| 2015 | 152.1 | 4380.3 |
| 2016 | 175.6 | 4664.2 |
| 2017 | 199.6 | 4828.3 |

Total Number of Rows: 22

Total Number of Columns: 3

Column Details:

- year: the year the data was recorded.
- interest: the interest earned through sale of houses (in billions of dollars).
- sales: total money earned through sale of houses (in billions of dollars).

Main dependent variable: sales.

Using SPSS Software, we have analyzed the data:

**Descriptive Statistics:**

| | INTEREST | SALES |
|---|---|---|
| Mean | 136.9091 | 2604.995 |
| Median | 147.0500 | 2840.650 |
| Maximum | 222.0000 | 4828.300 |
| Minimum | 57.50000 | 831.0000 |
| Std. Dev. | 51.34667 | 1283.980 |
| Skewness | -0.060239 | 0.149583 |
| Kurtosis | 1.676762 | 1.823706 |
| Jarque-Bera | 1.618350 | 1.350404 |
| Probability | 0.445225 | 0.509054 |
| Sum | 3012.000 | 57309.90 |
| Sum Sq. Dev. | 55366.10 | 34620708 |
| Observations | 22 | 22 |

Inferences:

- The variable sales is slightly right skewed, ranging between 831 to 4828.3 billion dollars.
- The variable interest is slightly left skewed, ranging between 57.5 to 222 billion dollats.
- There is no missing data.

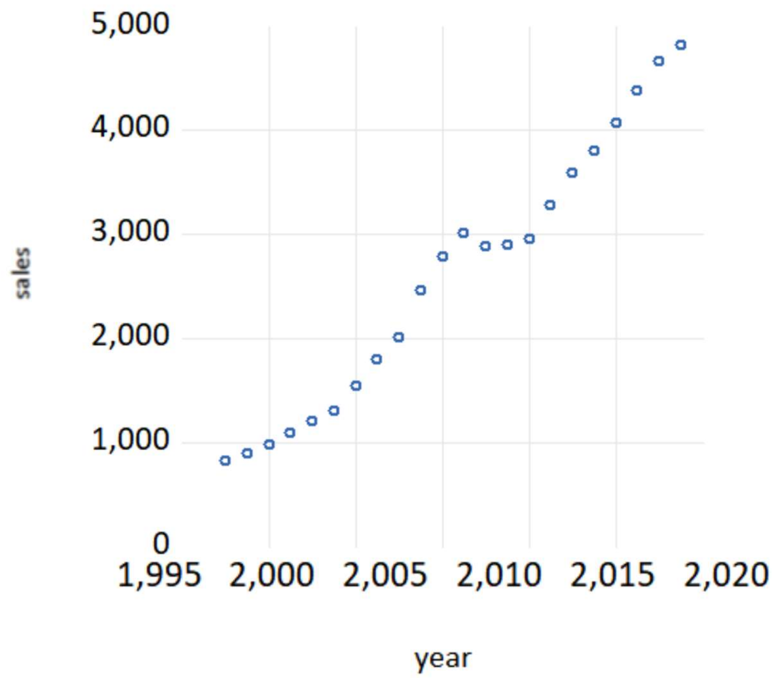
**Correlation Analysis:**

| | SALES | INTEREST |
|---|---|---|
| SALES | 1.000000 | 0.813538 |
| INTE... | 0.813538 | 1.000000 |

Inferences:

- The variables sales and interest have a high degree of positive linear correlation.

**Detecting Structural Changes in Dependent Variable:**




Inference: as we can see, there is a structural change between the year 2007 to 2008.

We will do further analysis to confirm the structural change by performing the chow test.

Breaking the dataset into two halves – 1996 to 2007 and 2008 to 2017, we have:

| year | interest | sales |
|---|---|---|
| 1996 | 57.5 | 831 |
| 1997 | 65.4 | 893.5 |
| 1998 | 59.7 | 980.5 |
| 1999 | 86.1 | 1098.7 |
| 2000 | 93.4 | 1205.7 |
| 2002 | 100.3 | 1307.3 |
| 2002 | 93 | 1546.3 |
| 2003 | 87.9 | 1801.4 |
| 2004 | 107.8 | 2006.6 |
| 2005 | 123.3 | 2467.1 |
| 2006 | 153.8 | 2790.4 |
| 2007 | 191.8 | 3012.5 |

Total Number of Rows: 12

| year | interest | sales |
|---|---|---|
| 2008 | 199.5 | 2890.9 |
| 2009 | 168.7 | 2902.5 |
| 2010 | 222 | 2954.6 |
| 2011 | 189.3 | 3279.8 |
| 2012 | 187.5 | 3590.4 |
| 2013 | 142 | 3802 |
| 2014 | 155.7 | 4075.9 |
| 2015 | 152.1 | 4380.3 |
| 2016 | 175.6 | 4664.2 |
| 2017 | 199.6 | 4828.3 |

Total Number of Rows: 10

On performing regression on both the sets, we have:

Set 1: Year 1996 to 2007:

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C | -169.6446 | 234.4420 | -0.723610 | 0.4859 |
| INTEREST | 18.01372 | 2.162915 | 8.328446 | 0.0000 |

| | | | |
|---|---|---|---|
| R-squared | 0.873997 | Mean dependent var | 1661.750 |
| Adjusted R-squared | 0.861396 | S.D. dependent var | 756.4158 |
| S.E. of regression | 281.6099 | Akaike info criterion | 14.26993 |
| Sum squared resid | 793041.3 | Schwarz criterion | 14.35075 |
| Log likelihood | -83.61961 | Hannan-Quinn criter. | 14.24001 |
| F-statistic | 69.36301 | Durbin-Watson stat | 0.945915 |
| Prob(F-statistic) | 0.000008 | | |

Inference:

- Equation: Sales = -169.644 + (18.01)(Interest).
- The $R^2$ of the model is 0.87, which means that the model has a very good explanatory power.
- The independent variable interest is statistically significant, having p-value 0.000.
- The intercept is statistically insignificant, having p-value 0.4859.
- Sum squared resid = RSS1 = 793041.3.

Set 2: Year 2008 to 2017:

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C | 5540.958 | 1760.547 | 3.147292 | 0.0137 |
| INTEREST | -10.06734 | 9.739557 | -1.033655 | 0.3315 |

| | | | |
|---|---|---|---|
| R-squared | 0.117820 | Mean dependent var | 3736.890 |
| Adjusted R-squared | 0.007547 | S.D. dependent var | 733.2195 |
| S.E. of regression | 730.4473 | Akaike info criterion | 16.20205 |
| Sum squared resid | 4268426. | Schwarz criterion | 16.26256 |
| Log likelihood | -79.01024 | Hannan-Quinn criter. | 16.13566 |
| F-statistic | 1.068443 | Durbin-Watson stat | 0.294947 |
| Prob(F-statistic) | 0.331530 | | |

Inference:

- Equation: Sales = 5540.958 + (-10.067)(Interest).
- The $R^2$ of the model is 0.11, which means that the model has a poor explanatory power.

- The independent variable interest is statistically insignificant, having p-value 0.3315.
- The intercept is statistically significant, having p-value 0.00137.
- Sum squared resid = RSS2 = 4268426**.**

Set 3: Year 1996 to 2017:

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C | -180.2031 | 474.1095 | -0.380088 | 0.7079 |
| INTEREST | 20.34342 | 3.251543 | 6.256543 | 0.0000 |

| | | | |
|---|---|---|---|
| R-squared | 0.661844 | Mean dependent var | 2604.995 |
| Adjusted R-squared | 0.644936 | S.D. dependent var | 1283.980 |
| S.E. of regression | 765.0881 | Akaike info criterion | 16.20437 |
| Sum squared resid | 11707195 | Schwarz criterion | 16.30355 |
| Log likelihood | -176.2480 | Hannan-Quinn criter. | 16.22773 |
| F-statistic | 39.14433 | Durbin-Watson stat | 0.443451 |
| Prob(F-statistic) | 0.000004 | | |

Inference:

- Equation: Sales = -180.203 + (20.343)(Interest).
- The $R^2$ of the model is 0.66, which means that the model has moderate explanatory power.
- The independent variable interest is statistically significant, having p-value 0.000.
- The intercept is statistically insignificant, having p-value 0.7069.
- Sum squared resid = RSS3 = 11707195**.**

**Calculating,**

RSSur = RSS1 + RSS2 = 793041.3 + 4268426 = 5061467.3

RSSr = RSS3 = 11707195

K = 2

n1 = 12

n2 = 10

Using the formula:

$F_{cal} = (RSS_R - RSS_{UR})/k / (RSS_{UR})/(n_1 + n_2 - 2k) = 5.01$.

From the table:

$F_{tab} = 2.91$.

Since, $F_{cal} > F_{tab}$, we reject the null hypothesis and acknowledge the presence of parameter instability and structural break in the data.