# P02 Instructor & Reference Guide – Box Plot, Regression, T‑Test & Chi‑Square

## Theory Brief

A box plot summarises the distribution through quartiles and outliers. Linear regression models the relationship between two continuous variables. The t‑test compares means of two independent samples. The chi‑square test assesses whether observed frequencies differ from expected frequencies.

## Worked Example

Below is a snapshot of the first few rows of the synthetic dataset and summary statistics:

X Y Group Response Category  
4.697 10.761 B 54.52 Red  
4.477 14.534 A 49.18 Red  
1.241 1.053 A 34.95 Blue  
7.548 21.347 A 40.24 Yellow  
9.730 30.398 A 58.08 Green

### Basic Statistics

* X\_mean: 4.74
* X\_median: 4.46
* X\_mode: 0.09
* X\_var: 7.83
* X\_std: 2.80
* Y\_mean: 14.29
* Y\_median: 13.67
* Y\_mode: -0.82
* Y\_var: 74.28
* Y\_std: 8.62
* Response\_mean: 53.56
* Response\_median: 53.53
* Response\_mode: 45.53
* Response\_var: 73.66
* Response\_std: 8.58

### Correlation Matrix

X Y Response  
X 1.000000 0.973354 -0.063655  
Y 0.973354 1.000000 -0.071508  
Response -0.063655 -0.071508 1.000000

### Visualisations

An example plot is saved in the results folder as P02\_plot1.png.

## Evaluation Rubric

* Correct creation and interpretation of box plot (20%)- Accuracy of regression parameters (25%)- Proper execution of t‑test and reporting of statistics (30%)- Correct chi‑square calculation (15%)- Clarity of documentation (10%)

## Common Pitfalls

Mixing up dependent and independent variables in regression, forgetting to check assumptions for t‑test, or misinterpreting chi‑square results.