General Linear Model and Optimisation

Topics covered

- About General Linear Model
- Independent sample t-test
- One-way ANOVA
- Regression Analysis
- Optimisation

About General Linear Model

- The **general linear model (**GLM) is a statistical model that provides a general framework to explain or predict a **continuous dependent variable** by a set of **independent variables that can be categorical or continuous**.

- **Y = a + bx,** In this equation, the components are:

- y = the y-axis variable, the outcome or posttest
- x = the x-axis variable, the pretest
- a = the intercept (value of y when x=0)
- b = the slope of the line

➢ The GLM covers Various Statistical Models/Techniques. **The key ones are:**

- **Independent sample T-test**
- **Analysis of variance (ANOVA) –One-way, two-way**
- **MANOVA**
- **Linear regression**
- **And many more**

Types of data

- **Nominal or categorical** data ***is data that comprises of categories that cannot be rank ordered – each category is just different.***

- **Ordinal data** is **data that comprises of categories that *can* be ranked in an order.**

- **Interval data** is an Ordinal data but with constant differences between observations… There is NO natural zero point.

- **Ratio data** measured on a ***continuous* scale** and ***does* have a natural zero point.**

Independent-samples T-test

Independent-samples T-test

- The independent-samples t-test *compares the means between two unrelated groups on the same continuous, dependent variable.*

- It is used to understand whether there is a significant difference in Dependent variable (like annual purchase etc.) based on Independent variable (like gender).

Assumptions for Independent-samples T-test

- **Assumption #1:** Your **dependent variable** should be measured on a **interval** or **ratio** scale.
- **Assumption #2:** Your **independent variable** should consist of **two categorical**, **independent groups**.
- **Assumption #3:** You should have **independence of observations**, which means that there is no relationship between the observations in each group or between the groups themselves.
- **Assumption #4:** There should be **no significant outliers**.
- **Assumption #5:** Your dependent variable should be **approximately normally distributed** for each group of the independent variable.
- **Assumption #6:** There needs to be **homogeneity of variances**.

Analysis of Variance (ANOVA)

ANOVA

- *ANOVA is used to determine whether there are any statistically significant differences between the means of different groups.*

- **ANOVA** uses the **F**-test to determine whether the variability between group means is larger than the variability of the observations within the groups. If that ratio is sufficiently large, you can conclude that not all the means are equal.

- A **N-way** (or factorial) analysis of variance, can examine data that are classified on multiple independent variables.

One-way ANOVA/ Single factor ANOVA

- The one-way analysis of variance (ANOVA) is used to determine whether there are any statistically significant differences between the means of two or more independent (unrelated) groups (although you tend to only see it used when there are a minimum of three, rather than two groups).

- For example, you could use a one-way ANOVA to understand whether exam performance differed based on test anxiety levels amongst students, dividing students into three independent groups (e.g., low, medium and high-stressed students).

Assumptions for One-way ANOVA test

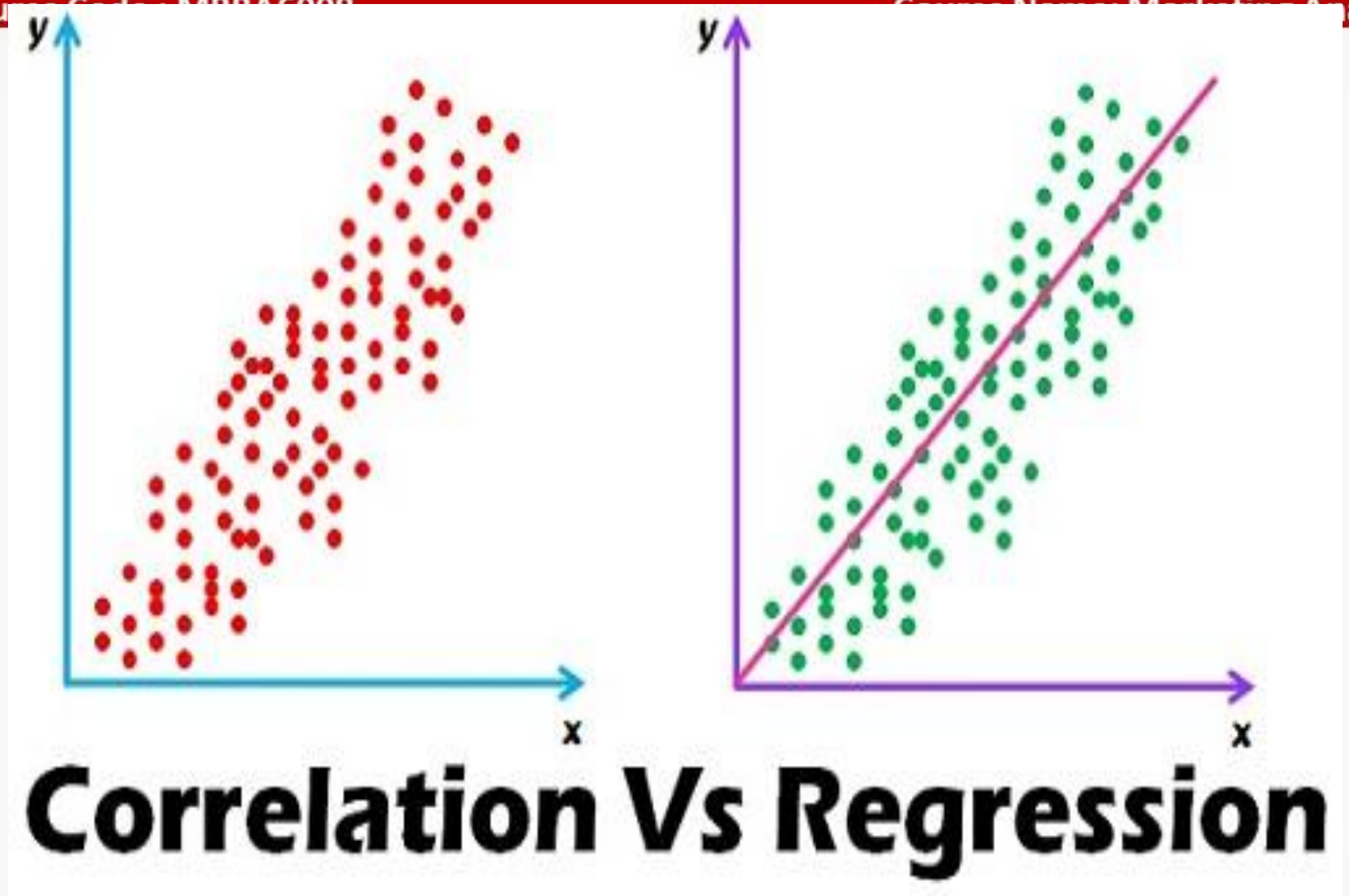
- **Assumption #1:** Your **dependent variable** should be measured on a **interval** or **ratio** scale.
- **Assumption #2:** Your **independent variable** should consist of **THREE OR MORE categorical**, **independent groups**. Can be used for two var.
- **Assumption #3:** You should have **independence of observations**, which means that there is no relationship between the observations in each group or between the groups themselves.
- **Assumption #4:** There should be **no significant outliers**.
- **Assumption #5:** Your dependent variable should be **approximately normally distributed** for each group of the independent variable.
- **Assumption #6:** There needs to be **homogeneity of variances**.

Regression

Correlation and Regression

- **Correlation** is a *statistical measure which determines co-relationship or association of two variables.*

- **Regression** *describes how an independent variable is numerically related to the dependent variable.* ... ***Regression*** *indicates the impact of a unit change in the known variable (x) on the estimated variable (y).*


y
x
y
x
Correlation Vs Regression

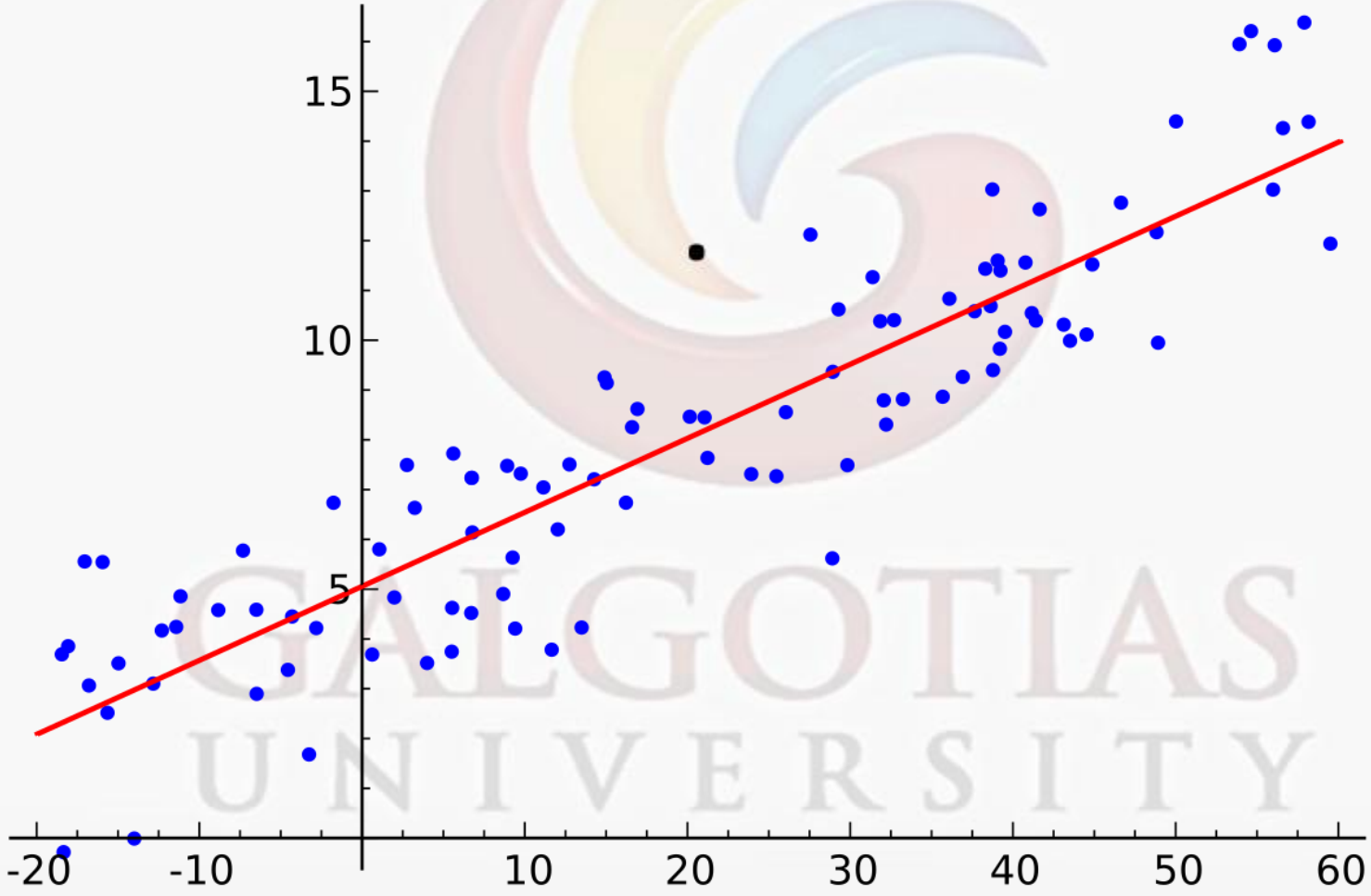
REGRESSION

- **Regression analysis** is a *set of statistical processes for estimating the relationships between a dependent variable (often called the 'outcome variable') and one or more independent variables (often called 'predictors').*

- The most common form of regression analysis is **linear regression,** *in which a researcher finds the line that most closely fits the data according to a specific mathematical criterion*… **y=a+bx**

- In statistics, **nonlinear regression** is a form of regression analysis in which the mathematical criterion is different from the above……
  . Y=a+(b/c+x)

Linear Regression



Linear REGRESSION

- *It is used when we want to predict the value of a variable based on the value of another variable*. The variable we want to predict is called the dependent variable (or sometimes, the outcome variable). The variable we are using to predict the other variable's value is called the independent variable (or sometimes, the predictor variable).

- If you have two or more independent variables, rather than just one, you need to **use multiple regression.**

Linear Regression Assumptions

- **Assumption #1:** Your two variables should be measured at the **continuous** level (i.e., they are either **interval** or **ratio** variables).
- **Assumption #2:** There needs to be a **linear relationship** between the two variables.
- **Assumption #3:** There should be **no significant outliers**. An outlier is an observed data point that has a dependent variable value that is very different to the value predicted by the regression equation. As such, an outlier will be a point on a scatterplot that is (vertically) far away from the regression line.
- **Assumption #4:** You should have **independence of observations**, which you can easily check using the Durbin-Watson statistic, which is a simple test to run using SPSS Statistics.
- **Assumption #5:** Your data needs to show **homoscedasticity**, which is where the variances along the line of best fit remain similar as you move along the line.
- **Assumption #6:** Finally, you need to check that the **residuals (errors)** of the regression line are **approximately normally distributed** (we explain these terms in our enhanced linear regression guide). Two common methods to check this assumption include using either a histogram

Optimization

- **Optimization** … the action of making the best or most effective use of a situation or resource.
- **Optimization** … Finding the best alternative with the highest achievable performance under the given constraints, by maximizing desired factors and minimizing undesired ones.
- **optimization problem** is the problem of finding the *best* solution from all feasible solutions. **LPP is an optimization problem.**
- **Linear programming** refers to choosing the best alternative from the available alternatives, whose objective function and constraint function can be expressed as linear mathematical functions.

- **Definition** of **objective function**: An equation to be optimized given certain constraints and with variables that need to be minimized or maximized.

➢ **Example of optimisation problem:**

✓ Minimise/Maximise f**(x) = 0.3x1+0.4x2**

Where **f(x) is Objective fn** and **x1, x2 are decision variables**

✓ **S**ubject to constraints x1+x2 <= OR >= 23, INEQUALITY CONSTRAINTS

x1=0 EQUALITY CONSTRAINTS

x1>=0, x2>=0 NON-NEGATIVE CONSTRAINTS

References:

- Rackley, Jerry (2015). Marketing Analytics Roadmap: Methods, Metrics, and Tools. Apress, 2015
- Winston, Wayne L (2014). Marketing Analytics: Data-Driven Techniques with Microsoft Excel. Hohn Wiley & Sons, 1st Ed.
- Grigsby, Michael (2015). Marketing Analytics: A practical guide to real marketing science. Kogan page.