SYLLABUS

Monday, May 6, 2024 2:40 PM

NORMALITY TESTING:

1. **SHAPIRO WILK TEST:**

   shapiro.test(data)

   - The null hypothesis of the Shapiro-Wilk test is that the data are normally distributed. When p>0.05

   - The alternative hypothesis is that the data are not normally distributed. When p<0.05

2. **LILLE TEST:**

3. **AD TEST: ANDERSON DARLING**
   ad.test(data)

- **SKEWNESS = 0 and KURTOSIS=3 FOR NORMALLY DISTRIBUTED DATA**

LINEARITY:

- Check linearity using plot(scatterplot)

ONE WAY ANOVA: *ANALYSIS OF VARIANCE*

- It is a set of technique for studying the cause and effect of one or more factors on a single dependent variable

- Here only one independent variable is studied. That's why it is called One –Way ANOVA.

- Dependent variable- sales, performance, opinion, etc.

- Independent variable - education , gender , city, etc.

- **Assumptions:**

    - Independent should be Categorical
    - Dependent should be Continuous
    - Data should be normally distributed
    - Independence: The data should be independent of each other i.e. the data of one group doesn't influence the other group
    - Homogeneity of variance: variance of all groups should be equal
    - Group sizes should be same: each group should have same number of respondents
    - Residuals should be normally distributed

- Here, Pr(>F) is the P-value
- If p-value >0.05 then Null is accepted -> No relation b/n IV and DV
- If p-value < 0.05 then Alternate is accepted -> Relation between IV and DV

TWO WAY ANNOVA:

- Two way ANOVA is similar to one way ANOVA in all the aspects except that in this case additional independent variable is introduced.

- Each independent variable includes two or more variants(levels).
- **ASSUMPTIONS:**
    - Population normality : Data is numerical data representing samples from normally distributed populations.
    - Homogeneity of Variance: the variances of the groups are "similar"
    - The sizes of the groups are "similar"
    - The groups should be independent.
    - The residuals are normally distributed

CORRELATION:

1. PEARSON:

    - cor(x,y,method="pearson") or cor.test(x,y,method="pearson")

2. SPEARMAN:
    - cor(x,y,method="spearman")

3. KENDALL:
    - cor(x,y,method="spearman")

COV():

- In R, cov() is a function used to compute the covariance matrix of a set of variables.
- Covariance measures the degree to which two variables change together.
- A positive covariance indicates that as one variable increases, the other variable tends to increase as well, while a negative covariance indicates that as one variable increases, the other variable tends to decrease.

REGRESSION:

1. SIMPLE LINEAR REGRESSION:

- Regression Analysis uses data to identify relationship among variables and use the relationship to make predictions.

- In correlation two variables are treated as equals. In regression one variable is treated as independent (predictor=X) variable and the other variable is dependent (outcome=Y) variable.

- There will be only two variables in the study in which one is independent and other is dependent.

2. MULTIPLE LINEAR REGRESSION:

- We find the impact of two or more independent variables on a dependent variable

- *Scale of measurement:* **Data should be in interval or ratio scale** for all independent variables and dependent variable.

- *Linearity:* **There must be linear relationship between variables.**

- *Normality of residuals:* **Multiple regression assumes that the residuals are normally distributed.**

- *Multicollinearity:* **Independent variables should not be highly correlated with each other.** This assumption is tested using Variance Inflation Factor (VIF) values(should not be more than 10).

- *y intercept or constant* :

  - Even if the value of x is 0, then also the y will have some value which is the constant (b0).
  - b1, b2, b3…..= beta or slope: If xi value increases by 1 point , then y will increase by bi.

- *R square* : xi (independent variables) explains….% of the y (dependent variable)

- *Significant Value:* p –value if $p<0.05$ at **5% level of significance.** H0 is rejected.

- *Residual :*

  - Difference between actual (Observed) value and explained (predicted) value.

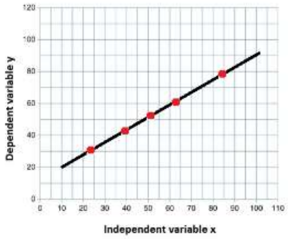
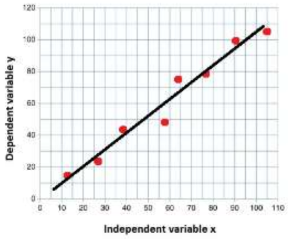
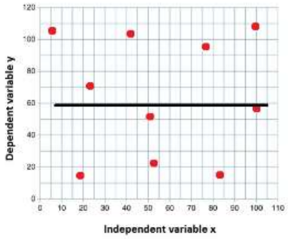
    Residual = Observed Value- Predicted Value

- *Standard Error* : Variance of residuals.

- *Adjusted R2* :

  - **The coefficient of determination, or R2 is a measure that provides information about the goodness of fit of a model**.

  - Adding more independent variables or predictors to a regression model tends to increase the R-squared value, which tempts makers of the model to add even more variables.

R^2 Values	Interpretation	Graph
$R^2 = 1$	All the variation in the y values is accounted for by the x values	
$R^2 = 0.83$	83% of the variation in the y values is accounted for by the x values	
$R^2 = 0$	None of the variation in the y values is accounted for by the x values	

$$y = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + \ldots\ldots + \varepsilon_i$$

y= Dependent Variable

x_i = Independent Variables, i=1,2,3…..

b_0 = y-intercept (Constant)

b_i = Slope, i= 1,2 ,3……

PARAMETRIC

TESTS	IVs	DVs
1 SAMPLE T-TEST	Test-Value	Continuous
2 SAMPLE T-TEST	Categorical	Continuous
PAIRED T-TEST	Continuous	Continuous
1 WAY ANOVA	Categorical	Continuous
2 WAY ANOVA	Both	Continuous
CORRELATION	Continuous	Continuous

SLR	Continuous	Continuous
MLR	Continuous	Continuous
LOGISTIC REGRESSION	Both	Categorical
DISCRIMINANT	Continuous with Normality	Categorical
DECISION TREE	Continuous or Mixed	Categorical for Classification ,
		Continuous for Regression
RANDOM TEST	Continuous or Mixed	Categorical
NAÏVE BAIYES	Continuous or Mixed	Categorical

NON-PARAMETRIC	
TESTS	VARIABLES
EFA	Continuous
CLUSTER	Categorical OR Continuous

TESTS	IVs	DVs
CHI-SQUARE TEST	Categorical	Categorical
SPEARMANS RANK	Ordinal	Continuous but no Normality OR Ordinal
WILCOXON ONE SAMPLE TEST	Test-Value	Continuous but no Normality OR Ordinal
MANN WHITNEY V-TEST	Categorical	Continuous but no Normality OR Ordinal
WILCOXON SINGLE RANK TEST	Ordinal	Continuous but no Normality OR Ordinal
KRUSHAL WALLIS TEST	3 or Categorical	Continuous but no Normality OR Ordinal