31/07/2024

Scales of Measurement

1) Nominal scale
2) Ordinal scale
3) Ratio scale
4) Interval scale

1) Nominal scale

This is the simplest level of measurement, where data is classified into mutually exclusive groups.

Ex:- Gender (male, Female)
Eye colour (blue, brown, green)

2) Ordinal scale

It is the 2^{nd} level of measurement that reports the ranking and ordering of the data without actually establishing the degree of variation b/w them.

Ex:- A number such as 1^{st}, 2^{nd}, 3^{rd}, 4^{th} that shows the position of something in a list of things.

3) Ratio scale

It is a type of variable measurement scale which is quantitative in nature, where there is a true zero and equal intervals between the measurements.

Ex:- Length, area and population

4) Interval scale:-

The difference b/w the two values is meaningful.

Ex:- Tempelature in Fahrenheit, celcius.

Computational statistics.

→ UNIT - I

Multivariate Normal Distribution:-

Multivariate Normal distribution functions, conditional distribution and its relation to regression model, estimation of parameters.

→ UNIT - II

Multiple linear regression Model:-

Standard multiple linear regression Models with emphasis on detection of collinearity, outliers, non-normality and auto correlation, validation of model assumptions.

→ UNIT - III

Multivariate Regression:-

Assumptions of Multivariate regression models, parameter estimation, Multivariate analysis of variance and co-variance.

→ UNIT - IV

Discriminant Analysis:-

Part - I
Statistical background, linear discriminant function analysis, estimating linear discriminant functions and their properties.

Part - II
Principal component Analysis:-

principal components, Algorithm for conducting principal component analysis, deciding on how

many principal components to retain, H-plot.

→ UNIT-V

Factor Analysis:-

part-I Factor analysis model, Extracting common factors, determining number of factors, Transformation of factor analysis solutions, factor scores, part-II Cluster Analysis:-Introduction, Types of clustering, correlations and distances, clustering by partitioning methods, hierarchical clustering, overlapping clustering, K-Means clustering-profiling and Interpreting clusters.

Text book:

→ An Introduction to Multivariate Statistical Analysis, T. W. Anderson.

→ ~~App~~ programming Python, Mark

UNIT-II

Multiple Linear Regression Model:-

~~Multiple linear~~

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon;$$

We consider the problem of Regression when ~~the~~ ~~stud~~ ~~variable~~ the study variable depends on more than one explanatory or independent variables, called a Multiple Linear Regression Model.

The Linear Model:-

Let y denotes the dependent (or study) variable i.e., linearly related to k independent (or Explanatory) variables $x_1, x_2, \ldots, x_k$ throughout the parameters $\beta_1, \beta_2, \ldots \beta_k$ and we write it,

response var. ↑

predictor var. ←

$$y = x_1\beta_1 + x_2\beta_2 + \cdots + x_k\beta_k + \varepsilon$$

This is called the multiple linear regression model

- The parameters $\beta_1, \beta_2, \ldots, \beta_k$ are the regression co-efficients associated with $x_1, x_2, \ldots, x_k$ respectively and 'ε' is the random error component reflecting the difference b/w the observed and fitted linear relationships.

Note:-

Note that 'j^{th}' regression coefficient 'β_j' represents the expected change in y per unit change in the j^{th} independent variable x_j.

- Assuming expectation, $E(\varepsilon)=0$

- A model is said to be linear when it is linear in parameters.

(i) $y=\beta_0+\beta_1 x$ is a linear model as it is linear in the parameters.

(ii) $y=\beta_0+\beta_1 x+\beta_2 x^2$ is linear in parameters β_0, β_1 and β_2 but it is non-linear in variables x. So it is a linear model.

Ex:- The income and education of a person are related. It is expected that, on average a higher level of education provides higher income.

- So a simple linear regression model can be expressed as,

$$\text{Income} = \beta_0 + \beta_1 \cdot \text{Education} + \varepsilon$$

Note that:- β_1 reflects a change in income w.r.t per unit change in education and β_0 reflects the income when education is zero as it is expressed (expected) that even and illiterate person can also have some income.

$$\text{Income} = \beta_0 + \beta_1 \cdot \text{education} + \varepsilon$$

$$\text{Income} = \beta_0 + \beta_1 \cdot \text{education} + \beta_2 \cdot \text{Age} + \varepsilon$$

Often it is observed that the income tends to rise less rapidly in the later earning years than is early years.

Model set-up for Multiple linear regression model

Let an experiment be conducted 'n' times and the data is obtained as follows:

observation no.	Response (y)	Explanatory var. ($x_1, x_2, x_3 \dots x_k$)
1	y_1	x_{11} $x_{12} \dots x_{1k}$
2	y_2	x_{21} $x_{22} \dots x_{2k}$
⋮	⋮	⋮ ⋮ ⋮
n	y_n	x_{n1} $x_{n2} \dots x_{nk}$

Assuming that, the model is original model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon$$

be n-tuples of observations are also assumed to follow the same model.:

Thus, they satisfy

$$y_1 = \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \dots + \beta_k x_{1k} + \varepsilon_1$$

$$y_2 = \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \dots + \beta_k x_{2k} + \varepsilon_2$$

$$\vdots$$

$$y_n = \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \dots + \beta_n x_{nk} + \varepsilon_n$$

These n-equations can be written as

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & x_{12} \dots x_{1k} \\ 1 & x_{21} & x_{22} \dots x_{2k} \\ \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} \dots x_{nk} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

or

$$\boxed{y = x\beta + \varepsilon}$$

In general, the model with k-explanatory variables can be expressed as

$$\boxed{y = x\beta + \varepsilon}$$

where, $y = (y_1\ y_2 \dots y_n)'$ is a $(n*1)$ vector of n observations on steady variable

when $x = \begin{pmatrix} 1 & x_{11} & x_{12} \dots x_{1k} \\ 1 & x_{21} & x_{22} \dots x_{2k} \\ \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} \dots x_{nk} \end{pmatrix}$

is a $(n*k)$ matrix of n observations of each of the k explanatory variables.

$\beta = (\beta_1, \beta_2, \ldots, \beta_k)'$ is a $(k*1)$ vector of regression co-efficients and

$\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots \varepsilon_n)'$ is a $(n*1)$ vector of random error components or disturbance term.

- If intercept term is present take first column of X to be $(1,1,\ldots 1)'$.

Assumptions of Multiple Linear Regression *

1) Linearity :- There is a linear relationship b/w the dependent variable & each of the independent variable.

Dependent - also called study variable.

Independent - also called as explanatory variable.

2) Independence of errors :- The residuals (errors) of the model are independent of each other.

3) Homoscadaltacity :- The variance of the errors is constant across all levels of the independent variables.

4) No multicollinearity :- 2 or more independent variables are collinear. The independent variables are not highly correlated.

5) Normality of Errors:- The residuals are of the model are normally distributed.

6) No Auto-correlation:- The residuals are not auto-correlated.

(or) Outliers & its Detection Methods

In simple terms, an outlier is extremely high or extremely low data points relative to the nearest data points & the rest of the neighbouring co-existing values in a data graph or data set (or) outlier are extreme values that stand out greatly from the overall patterns in a data set (or) a graph.

Eg:-

Height (in inches)	60	62	64	66	68	70	72	75
Weight (in pounds)	110	115	120	125	130	135	140	220

Leverage

Leverage measures the distance b/w the observations predictor values & mean of the predictor values (or) points with high leverage have predictor values that are far from the mean & therefore have a greater potential to influence the regression model.

Jacknife Residuals

Jacknife Residuals are derived by systematically excluding each observation from the dataset & recalculating the regression model without the observation. The difference b/w the actual observed value & the predicted value from this new model (excluding the observation) gives the Jacknife Residual.

Calculations:-

$$y_i^{(Jack)} = \frac{(y_i - \hat{y}_{(i)})}{\sqrt{1-h_i}}$$

y_i = observed value

$\hat{y}_i$ = predicted value

h_i = Leverage observation

Cook's Distance

Cook's distance is a statistical measure used on regression analysis to identify observations that have a significant influence on the estimated co-efficient of the model. It combines the information about both the leverage of the (practised datapoint variable, how far it is from the devarage of the practised variable) and the residual (the difference b/w observed & predicted).

Calculation:-

For an observation (i) the cook's distance (D_i) is as follows.

$$D_i = \frac{(r_i)^2}{P.(MSE)} \cdot \frac{h_i}{(1-h_i)^2}$$

where,
- r_i is the residual for observation (i) i.e., $y_i - \hat{y_i}$
- P is the no. of predictors in the model including the intercept.
- MSE is the mean squared error of the regression model.
- h_i is the leverage of observation which measures how far the observations predicted values are from the mean of the predicted values.

Multicollinearity

It refers to the situation in statistical modeling, particularly a regression analysis, where two or more predictor variables are highly correlated. This correlation means that as predictor variable can be linearly predictor from the others with a substaincial degree of accuracy.

When multicollinearity is present, it undetermines the statistical significance of the predictor variables, making it difficult to determine the individual effect of each predictor on the dependent variable.

*** Detection Methods of Multicollinearity:

1) Correlation Matrix

A simple way to detect multicollinearity is by examining the correlation matrix of the predictor variables. If the correlation coefficient b/w any two variables is high (above 0.8 or below -0.8) this indicates potential multicollinearity.

2) Variance inflation factor (VIF)

The VIF quantifies how much the variance of a regression co-efficient is inflated due to multi-collinearity. VIF values above 10 indicates high multi-collinearity.

Formula:

$$VIF = \frac{1}{1 - R_i^2}$$

where, R_i^2 is the value obtained by regressing the i^{th} predictor on all other predictors.

3) Tolerance

Tolerance is the inverse of VIF. It is another measure used to detect multi-collinearity low tolerance value (below 0.1) indicates high multi-collinearity.

Formula:

$$\text{Tolerance} = \left(\frac{1}{VIF}\right) = 1/\left(1/1-R_i^2\right) = 1-R_i^2$$

4) Condition Index (C.I)

Condition Index is derived from eigen values of the scaled, centered matrix of the predictors. High condition index values (above 30) suggest multi-collinearity.

- It is computed as the square-root of the ratio of largest eigen values to each value.

5) Eigen values and Eigen vectors.

By examining the eigen values of the correlation matrix of the predictors, one can detect multi-collinearity. Near zero eigen values indicate near linear dependency among the predictors.

Sources of Multi-collinearity

1) Method of data collection:-

It is expected that the data is collected over the whole class section of variables. It may happen that the data is collected over a sub-space of the explanatory variables

where the variables are linearly dependent.

2) Model and population constraints

There may exists some constraints on the model or on the population where the sample is drawn. The sample may be generated from that part of the population having linear ~~populations~~ combinations.

3) Existence of Identities

There may exist some relationship among the variables which maybe due to the defination of the variables or any identity relation among them.

For example, if data is collected on the variables like income, saving and expenditure then, ~~income equal to sav~~

$$\text{Income} = \text{Saving} + \text{expenditure.}$$

Such relationship will not change even the sample size increases.

4) Imprecise formulation of the Model.

The formulation of the model where unnecessari be complicated.

For example, the quadratic or polynomial terms or cross-product terms may appear as explanatory variables.

5) ~~An over-determining~~

5) An over-determined Model

Sometimes due to over enthusiasm, a large number of variables are included in the model to make it more realistic. Consequently the number of observations (n) becomes smaller than the number of explanatory variables (k). Such a situation can arise in medical research where the number of patients maybe small, but the information is collected on a large number of variables.

Auto-correlation

One common way for the ~~Independence~~ "Independence" condition in a multiple regression model to fail when the sample data have been collected over-time and the regression model fails to effectively capture any time trends. In such a circumstance the random errors in the model are often positively correlated over time, so that each random error is more likely to be similar to the

previous random error that would be if a random process for independent of one another. The phenomenon is known as auto-correlation (or serial-correlation), and can sometimes can be detected by plotting the model residuals v/s time.

Durbin-Watson Statistic

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

where, $e_i = (y_i - \hat{y}_i)$ are the residuals.

n = no. of observations / elements in the sample

k = no. of independent variables.

- d takes on values b/w 0 and 4.

A value of $d = 2$ means there is no auto-correlation

A value substentially below 2 (and especially a values less than 1) mean that the data is positively correlated, i.e., on average, a data element is close to the subsequent data element. A value of 'd' substentially above 2 means that negatively auto-correlated, i.e., on average

a data element is far from the subsequent data element.

Fitting of a Multiple Linear Regression Model:-

Example:-

1) The owner of a chain of 10 stores wishes to forecast net profit with the help of next year's projected sales of food and non-food items. The data about current year's sales of food items, sales of non-food items as also profit for all the 10 stores available as follows:

Super Market No.	1	2	3	4	5	6	7	8	9	10
Net profit (cr) (y)	5.6	4.7	5.4	5.5	5.1	6.8	5.8	6.2	5.8	6.2
Sales of food items (cr) (x_1)	20	15	18	20	16	25	22	30	24	25
Sales of non-food items (cr) (x_2)	5	5	6	5	6	6	4	7	3	4

Sol In this case the relationship is expressed by:

$$y = b_0 + b_1 x_1 + b_2 x_2$$

where, y denotes net profit
x_1 denotes sales of food items
x_2 denotes sales of non-food items

b_0, b_1, b_2 are regression co-efficients or constants. Their values are obtained by the formulae derived from the "principles of least squares"

Super market no.	y	x_1	x_2
1			

The required calculations can be made with the help of following table.

Super-Market	y_i	x_1 (x_{1i})	x_2 (x_{2i})	x_{1i}^2	$x_1 y_i$	y_i^2	$x_2 y_i$	x_{2i}^2	$x_{1i} \cdot x_{2i}$
1	5.6	20	5	400	112	31.36	28	25	100
2	4.7	15	5	225	70.5	22.09	23.5	25	75
3	5.4	18	6	324	97.2	29.16	32.4	36	108
4	5.5	20	5	400	110	30.25	27.5	25	100
5	5.1	16	6	256	81.6	26.01	30.6	36	96
6	6.8	25	6	625	170	46.24	40.8	36	150
7	5.8	22	4	484	127.6	33.64	23.2	16	88
8	8.2	30	7	900	246	67.24	57.4	49	210
9	5.8	24	3	576	139.2	33.64	17.4	9	72
10	6.2	25	4	625	155	38.44	24.8	16	100
Sum	59.1	215	51	4815	1309.1	356.07	305.6	273	1099
average	5.91	21.5	5.1	481.5	130.91	35.607	30.56	27.3	109.9

$$y = b_0 + b_1 x_1 + b_2 x_2$$

$$b_1 = \frac{(\Sigma y_i x_{1i} - n\bar{y}\bar{x}_1)(\Sigma x_{2i}^2 - n\bar{x}_2^2) - (\Sigma y_i x_{2i} - n\bar{y}\bar{x}_2)(\Sigma x_{1i} x_{2i} - n\bar{x}_1\bar{x}_2)}{(\Sigma x_{1i}^2 - n\bar{x}_1^2)(\Sigma x_{2i}^2 - n\bar{x}_2^2) - (\Sigma x_{1i} x_{2i} - n\bar{x}_1\bar{x}_2)^2}$$

$$b_1 = 0.196$$

$$b_2 = \frac{\Sigma(y_i x_{2i} - n\bar{y}\bar{x}_2)(\Sigma x_{1i}^2 - n\bar{x}_1^2) - (\Sigma y_i x_{1i} - n\bar{y}\bar{x}_1)(\Sigma x_{1i} x_{2i} - n\bar{x}_1\bar{x}_2)}{(\Sigma x_{1i}^2 - n\bar{x}_1^2)(\Sigma x_{2i}^2 - n\bar{x}_2^2)(\Sigma x_{1i} \cdot x_{2i} - n\bar{x}_1\bar{x}_2)^2}$$

$$b_2 = 0.287$$

$$b_0 = (\bar{y} - b_1\bar{x}_1 - b_2\bar{x}) \Rightarrow b_0 = 0.233 \Rightarrow \boxed{y = 0.233 + 0.196x_1 + 0.287x_2}$$

This equation is known as multiple regression eq^n of y on x_1 and x_2 and it indicates as to how 'y' changes w.r.t changes in x_1 & x_2.

The interpretation of the values of the coefficient of x_1 b_1 i.e., 0.196 is held constant then for every addition crore of sales of food items, the net profit is increased by 0.196 crore i.e., 19.6 lakhs. Similarly, the interpretation of the value of coefficient of x_2 b_2 i.e., 0.287 is that if x_1 held constant the sales of non-food items increased by 1 crore. The net profit is increased by 0.287 crore i.e., Rs. 28.76 lakh.

Multiple Linear Regression in Matrix form:-

$$y = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + e_i$$

$$y = x\beta + e.$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{21} \\ 1 & x_{12} & x_{22} \\ \vdots & \vdots & \vdots \\ 1 & x_{1k} & x_{2k} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_k \end{bmatrix}$$

Ex:-

1) Find the coefficient of Regression in Matrix form from the given data.

y	9	10	13	14	16
x_1	1	3	4	6	7
x_2	10	14	15	18	20

Sol

$$y = \begin{bmatrix} 9 \\ 10 \\ 13 \\ 14 \\ 16 \end{bmatrix}, \quad X = \begin{matrix} x_1 & x_2 \\ \begin{bmatrix} 1 & 1 & 10 \\ 1 & 3 & 14 \\ 1 & 4 & 15 \\ 1 & 6 & 18 \\ 1 & 7 & 20 \end{bmatrix} \end{matrix}$$

$$\beta = (x'x)^{-1} x'y = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}$$

$$x' = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 3 & 4 & 6 & 7 \\ 10 & 14 & 15 & 18 & 20 \end{bmatrix}$$

$$x'x = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 3 & 4 & 6 & 7 \\ 10 & 14 & 15 & 18 & 20 \end{bmatrix} \begin{bmatrix} 1 & 1 & 10 \\ 1 & 3 & 14 \\ 1 & 4 & 15 \\ 1 & 6 & 18 \\ 1 & 7 & 20 \end{bmatrix}$$

$$= \begin{bmatrix} (1+1+1+1+1) & (1+3+4+6+7) & (10+14+15+18+20) \\ (1+3+4+6+7) & (1+9+16+36+49) & (10+42+60+105+140) \\ (10+14+15+18+20) & (10+42+60+105+140) & (100+196+225+324+400) \end{bmatrix}$$

$$= \begin{bmatrix} 5 & 21 & 77 \\ 21 & 111 & 360 \\ 77 & 360 & 1245 \end{bmatrix} = x'x$$

$$x'y = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 3 & 4 & 6 & 7 \\ 10 & 14 & 15 & 18 & 20 \end{bmatrix} \begin{bmatrix} 9 \\ 10 \\ 13 \\ 14 \\ 16 \end{bmatrix} = \begin{bmatrix} 9+10+13+14+16 \\ 9+30+52+84+112 \\ 90+140+195+252+320 \end{bmatrix} = \begin{bmatrix} 62 \\ 287 \\ 997 \end{bmatrix}$$

$$x'x = \begin{bmatrix} 5 & 21 & 77 \\ 21 & 111 & 360 \\ 77 & 360 & 1245 \end{bmatrix}$$

$$(x'x)^{-1} = \frac{1}{\det(x'x)} \cdot \text{Adj of } x'x$$

$$\text{Adj. of } x'x = (\text{cofactor of } x'x)'$$

UNIT-I

Multivariate Normal - Distribution

Introduction to the Multivariate Normal Distribution

The probability density function of the Univariate normal distribution ($p=1$ variables):

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-1}{2}\left[\frac{x-\mu}{\sigma}\right]^2\right], \text{ for } -\infty < x < \infty$$

- The parameters that completely characterize the distribution:

$$\mu = E(x) = \text{mean}$$
$$\sigma^2 = \text{var}(x) = \text{variance}$$

Generalization

$$\left(\frac{x-\mu}{\sigma}\right)^2 = (x-\mu)(\sigma^2)^{-1}(x-\mu)$$

$-\infty < x_i < \infty$ for $i = 1, \ldots p$.

- This is a scalar and reduces to what's at the top for $p=1$.
- It is a squared statistical distance of x to μ (if Σ^{-1} exists). It takes into consideration both variability and covariability.
- Integrating

$$\int_{x_1} \cdots \int_{x_p} \exp\left(\frac{-1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right) = (2\pi)^{p/2}|\Sigma|^{1/2}$$

- Since the sum of probabilities over all possible values must add up to 1, we need to divide by $(2\pi)^{p/2}|\Sigma|^{1/2}$ to get a "proper" density function.

Multivariate Normal Density Function:

$$f(x) = \frac{1}{(2\pi)^{p/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right)$$

where $-\infty < x_i < \infty$ for $i = 1, \ldots, p$

To denote this, we use

$$N_p(\mu, \Sigma)$$

For $p = 1$, this reduces to univariate

$p = 2$, bivariate.

Bivariate Normal: $p = 2$

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \quad E(x) = \begin{bmatrix} E(x_1) \\ E(x_2) \end{bmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} = \mu$$

$$\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}$$

and $$\Sigma^{-1} = \frac{1}{\sigma_{11}\sigma_{22} - \sigma_{21}\sigma_{12}} \begin{bmatrix} \sigma_{22} & -\sigma_{21} \\ -\sigma_{12} & \sigma_{11} \end{bmatrix}$$

replace σ_{12} by $\rho_{12}\sqrt{\sigma_{11}\sigma_{12}}$, we get

$$\Sigma^{-1} = \frac{1}{\sigma_{11}\sigma_{22}(1-\rho_{12}^2)} \begin{bmatrix} \sigma_{22} & -\rho_{12}\sqrt{\sigma_{11}\sigma_{22}} \\ -\rho_{12}\sqrt{\sigma_{11}\sigma_{22}} & \sigma_{11} \end{bmatrix}$$

Bivariate Normal

$$(x-\mu)'\Sigma^{-1}(x-\mu)$$

$$= ((x_1-\mu_1), (x_2-\mu_2)) \left(\frac{1}{\sigma_{11}\sigma_{22}(1-\rho_{12}^2)}\right) \times$$

$$\begin{pmatrix} \sigma_{22} & -\rho_{12}\sqrt{\sigma_{11}\sigma_{22}} \\ -\rho_{12}\sqrt{\sigma_{11}\sigma_{22}} & \sigma_{11} \end{pmatrix} \begin{pmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{pmatrix}$$

$$= \frac{1}{1-\rho_{12}^2}\left\{\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)^2+\left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)^2-2\rho_{12}\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)\left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)\right\}$$

$$= \frac{1}{1-\rho_{12}^2}\left\{z_1^2+z_2^2-2\rho_{12}z_1z_2\right\}.$$

Bivariate Normal & Independence.

$$f(x) = \frac{1}{2\pi\sqrt{\sigma_{11}\sigma_{22}}}\exp\left[\frac{-1}{2(1-\rho_{12}^2)}\left\{\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)^2+\left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)^2 - 2\rho_{12}\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)\left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)\right]\right.$$

If $\sigma_{12}=0$ or equivalently $\rho_{12}=0$, then X_1 and X_2 are uncorrelated. For bivariate normal, $\sigma_{12}=0$ implies that X_1 and X_2 are statistically independent, because the density factors.

$$f(x) = \frac{1}{2\pi\sqrt{\sigma_{11}\sigma_{22}}}\exp\left[\frac{-1}{2}\left\{\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)^2+\left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)^2\right\}\right]$$

$$= \frac{1}{\sqrt{2\pi\sigma_{11}}}\exp\left[\frac{-1}{2}\left(\frac{x_1-\mu_1}{\sqrt{\sigma_{11}}}\right)^2\right]\frac{1}{\sqrt{2\pi\sigma_{22}}}\exp\left[\frac{-1}{2}\left(\frac{x_2-\mu_2}{\sqrt{\sigma_{22}}}\right)^2\right]$$

$$= f_1(x_1)\, f_2(x_2).$$

Properties of Multivariate Normal distribution

If $X \sim N_p(\mu, \Sigma)$ then

- Linear combinations of components of X are (multivariate) normal.

- All sub-sets of the components of X are (multivariate) normal.
- Zero covariance implies that the corresponding components of X are statistical independent.
- The conditional distributions of the components of X are (multivariate) normal.

Linear combinations

If $X \sim N_p(\mu, \Sigma)$, then any linear combination

$$a'X = a_1X_1 + a_2X_2 + \cdots + a_pX_p$$

is distributed as

$$a'X \sim N_1(a'\mu, a'\Sigma a)$$

Also, if $a'X$ is normal $N(a'\mu, a'\Sigma a)$ for all possible a, then X must be $N_p(\mu, \Sigma)$.

Sub-sets of variables

If $X \sim N_p(\mu, \Sigma)$, then all sub-sets of X are (multivariate) normally distributed.

Example

Suppose $X = \begin{bmatrix} X_1 \\ X_2 \\ X_3 \end{bmatrix} \sim N_3(\mu, \Sigma)$

Due to the result on sub-sets of multivariate normals

$$X_1 \sim N(\mu_1, \sigma_{11})$$
$$X_2 \sim N(\mu_2, \sigma_{22})$$
$$X_3 \sim N(\mu_3, \sigma_{33})$$

Also,

$$\begin{bmatrix} X_2 \\ X_3 \end{bmatrix} \sim N\left(\begin{bmatrix} \mu_2 \\ \mu_3 \end{bmatrix}, \begin{bmatrix} \sigma_{22} & \sigma_{23} \\ \sigma_{32} & \sigma_{33} \end{bmatrix}\right)$$

Zero covariance & statistical independence

There are three parts to this one:

- If X_1 is $(q_1 \times 1)$ and X_2 is $(q_2 \times 1)$ are statistically independent, the $\text{cov}(X_1, X_2) = \Sigma_{12} = 0$

- If $$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N_{q_1+q_2}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \left(\begin{array}{c|c} \Sigma_{11} & \Sigma_{12} \\ \hline \Sigma_{12} & \Sigma_{22} \end{array}\right)\right)$$

  Then X_1 and X_2 are statistically independent if and only if $\Sigma_{12} = \Sigma'_{21} = 0$.

- If X_1 and X_2 are statistically independent and distributed as $N_{q_1}(\mu_1, \Sigma_{11})$ and $N_{q_2}(\mu_2, \Sigma_2)$, respectively, then

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim N_{q_1+q_2}\left(\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \left(\begin{array}{c|c} \Sigma_{11} & 0 \\ \hline 0 & \Sigma_{22} \end{array}\right)\right).$$

Conditional Distributions

Let $X' = (X'_{1(q_1 \times 1)}, X'_{2(q_2 \times 1)})$ be distributed as $N_{q_1+q_2}(\mu, \Sigma)$ with

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \text{ and } \Sigma = \left(\begin{array}{c|c} \Sigma_{11} & \Sigma_{12} \\ \hline \Sigma_{21} & \Sigma_{22} \end{array}\right)$$

and $|\Sigma| > 0$ (i.e., positive definite). Then the conditional distribution of X_1 given $X_2 = x_2$ is (multivariate) normal with mean and covariance matrix.

$\mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(x_2 - \mu_2)$ and $\Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$

Least Square Estimation

$Y = Z\beta + \epsilon$ where $E(\epsilon) = 0$ and $\text{cov}(\epsilon) = \sigma^2 I$

β and σ^2 are unknown parameters that need to be estimated from data.

Let $y_1, y_2, \ldots, y_n$ be a random samples with values $z_1, z_2, \ldots z_r$ on the explanatory variables. The least squares estimate of β is the vector b that minimizes

$$\sum_{j=1}^{n}(y_j - z_j' b)^2 = \sum_{j=1}^{n}(y_j - b_0 - b_1 z_{j1} - b_2 z_{j2} - \ldots - b_r z_{jr})^2$$

$$= (y - Zb)'(y - Zb)$$

$$= \epsilon'\epsilon$$

where z_j' is the j^{th} row of Z and $b = (b_0, b_1, \ldots b_r)'$

If Z has full rank (i.e., the rank of Z is $r+1 \leq n$) then the least squares estimate of β is

$$\hat{\beta} = (Z'Z)^{-1}Z'y$$

Estimation of μ and Σ

Suppose we have a p dimensional normal distribution with mean μ and covariance matrix Σ.

Take n observations $x_1, x_2, \ldots x_n$ (these are each $(p \times 1)$ vectors).

$x_j \sim N_p(\mu, \Sigma)$ $j = 1, 2, \ldots, n$ and independent

For $p = 1$, we know that the MLEs are

$$\hat{\mu} = \bar{x} = \frac{1}{n}\sum_{j=1}^{n} x_j \sim N\left(\mu, \frac{1}{n}\sigma^2\right)$$

And $n\hat{\sigma}^2 = \sum_{j=1}^{n} (x_j - \bar{x})^2$ and $\frac{1}{\sigma^2} \sum_{j=1}^{n} (x_j - \bar{x})^2 \sim \chi^2_{(n-1)}$

(or) $\hat{\sigma}^2 = \frac{1}{n} \sum_{j=1}^{n} (x_j - \bar{x})^2 \sim \sigma^2 \chi^2_{(n-1)}$

↳ chi-square

- A maximum likelihood estimation of μ is

$$\hat{\mu} = \bar{x} = \frac{1}{n} \sum_{j=1}^{n} x_j$$

and the ML estimator of Σ is

$$\hat{\Sigma} = \frac{n-1}{n} s^2 = S_n = \frac{1}{n} \sum_{j=1}^{n} (x_j - \hat{\mu})(x_j - \hat{\mu})'$$