Universiti Teknologi MARA
Fakulti Sains Komputer dan Matematik
UiTM Cawangan Negeri Sembilan
Kampus Kuala Pilah

What's *what* FSKM

eISSN: 2756-7729

VOLUME 1/2021

EDITORIAL BOARD

fskm_kp@uitm.edu.my
FSKM UITM Kuala Pilah
FSKM UITM Kuala Pilah

Regression Model: Choose the Right One!!!

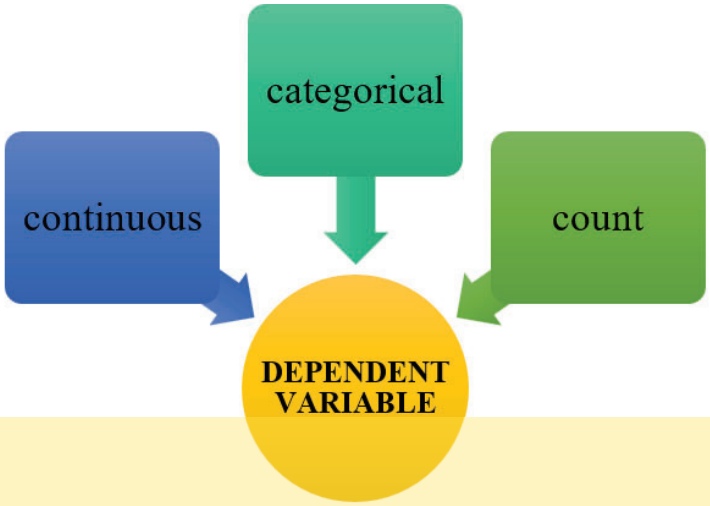
Prepared by Zuraida Jaafar

Regression analysis is a set of statistical methods used for the estimation of relationships between a dependent variable and one or more independent variables. It can be utilized to assess the strength of the relationship between variables and for modelling the future relationship between them.

Regression analysis can be used to answer those questions such as:
Do socio-economic status and race affect educational achievement?
- Do education and IQ affect earnings?
- Do exercise habits and diet affect weight?
- Are drinking coffee and smoking cigarettes related to mortality risk?

There are numerous types of regression models that can be used. This choice often depends on the kind of data for the **dependent variable** and the type of model that provides the best fit.

Researchers may choose appropriate regression models based on the type of dependent variables such as continuous, categorical or count data.



References:

1. David A. Freedman (2009). Statistical Models: Theory and Practice. Cambridge University Press. p. 26.
2. Bates, D. M., & Watts, D. G. (1988). Nonlinear regression analysis and its applications. New York: John Wiley & Sons.
3. Darroch, J.N. & Ratcliff, D. (1972). "Generalized iterative scaling for log-linear models". The Annals of Mathematical Statistics. 43 (5): 1470–1480.
4. Menard, Scott (2002). Applied Logistic Regression Analysis. SAGE. p. 91
5. Long, J. S. and Freese, J. (2006). Regression Models for Categorical Dependent Variables Using Stata, Second Edition. College Station, TX: Stata Press.

Table below is the list of several common regression model according to different types of dependent variables.

Type of Dependent Variable	Regression Model	Description
Continuous ▪Numerical data which arises from measuring process ▪Example: weight, time, and length.	Linear Example of dependent variable: academic performance (CGPA=3.85)	Linear approach to model the relationship between dependent variable and one or more independent variables. $y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \varepsilon_i$
	Partial Least Square (PLS) Example of dependent variable: Customer satisfaction score	Used to find the relationships when there are very few observations compared to the number of independent variables or when independent variables are highly correlated. $y_{nj} = \sum_{i=0}^{k} \beta_i x_{ni} + \varepsilon_{nj}$
	Non-Linear Example of dependent variable: Population growth	Model in which observational data are modelled by a function which is a nonlinear combination of the model parameters. $y \sim f(x, \beta)$
Categorical ▪ Values that can be put into a countable number of distinct groups based on a characteristic. ▪Categorical data might not have a logical order.	Binary Logistic Example of dependent variable: Decision to vaccinate (yes, no)	Used to predict the probability that an observation falls into one of two categories of a dichotomous dependent variable based on one or more independent variables that can be either continuous or categorical. $p(x) = \sigma(t) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$
	Ordinal Logistic Example of dependent variable: Internet addiction level (none, mild, severe)	Used to predict an ordinal dependent variable given one or more independent variables. $\sigma(\theta_i - \mathrm{w} \cdot \mathrm{x}) = \frac{1}{1 + e^{-(\theta_i - \mathrm{w} \cdot \mathrm{x})}}$
	Nominal Logistic Example of dependent variable: Profession (surgeon, doctor, nurse, dentist, therapist)	Used when the dependent variable in question is nominal (*categorical*) and for which there are more than two categories. $\Pr(Y_i = c) = \frac{e^{\beta_c} \cdot \mathrm{X}_i}{\sum_{k=1}^{K} e^{\beta_{ic} \cdot \mathrm{X}_i}}$
Count ▪Count of items, events, results, or activities. ▪Numerical data which arises from counting process.	Poisson Example of dependent variable: Number of children in a family	Used to predict a dependent variable that consists of count data given one or more independent variables. $\log\big(\mathrm{E}(Y \mid \mathrm{x})\big) = \alpha + \beta' \mathrm{x})$
	Negative Binomial Example of dependent variable: Number of accidents in Negeri Sembilan.	Used to predict a dependent variable that consists of count data. It is known as NB2, which can be more appropriate when the variance of count data is greater than the mean count (overdispersion).