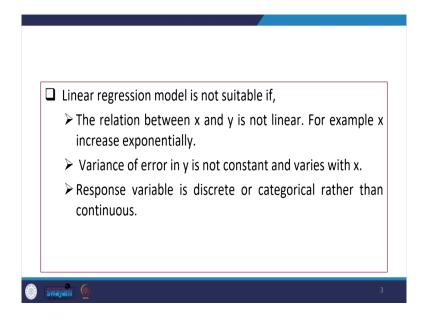
## Exploring Survey Data on Health Care Prof. Pratap C. Mohanty Department of Humanities and Social Sciences Indian Institute of Technology, Roorkee

## Lecture - 29 Generalized Linear Model

Welcome friends once again to my NPTEL MOOC module on Exploring Healthcare Survey Data. We are in the last week of the module. We have been discussing about health care data and in this particular lecture we are trying to understand generalized linear model. This GLM is used when your linear model has certain limitations and all the assumptions are not fulfilled.

Then we go for some generalizations of that linear model. Now I am just going to discuss about who has developed it? & how it has emerged? So, GLM is an advanced statistical modelling technique that generalizes the linear model. It was developed by John Nelder and Robert Wedderburn in 1972. It is an umbrella term that encompasses many other models which allows the response variable i.e., Y, to have an error distribution other than a normal distribution

(Refer Slide Time: 01:34)

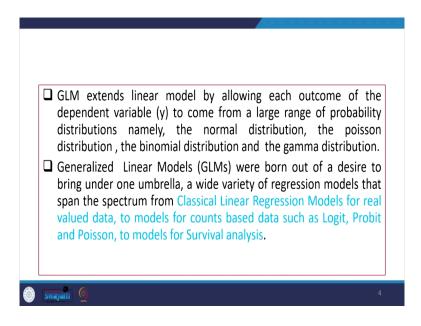


Linear regression model is not suitable if the relations between X and Y is not linear. For example, X increases exponentially, if there is exponential distribution like cumulative density function which is not linear. In fact, in that case linear regression is not suitable.

Similarly, another one is called variance of error in the Y i.e., Y is not constant and varies with X. So, the variance, if it is not constant throughout then in that case we already know that linear regression model is not fitted and we go for some form of transformation.

If Response variable is discrete or categorical rather than a continuous one then in that case linear regression is also not applicable. Therefore, we think of some possibility of some generalization and try to use generalized least square method or generalized least square model or generalized linear model.

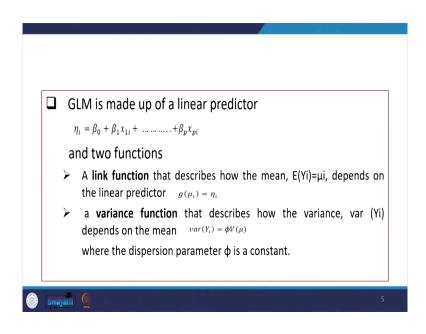
(Refer Slide Time: 02:41)



GLM extends linear model by allowing each outcome of the dependent variable to come from a large range of probability distribution. Namely, the normal distribution, the Poisson distribution, the binomial distribution and the gamma distribution. You might have heard about this distribution. It has certain specifications or certain features of Poisson. We know that it has certain count data of binomial, we know that it follows Bernoulli distribution with binary numbers. So, gamma functions are accordingly defined normal distribution where the data is of bell kind shape, symmetry is usually identified. The generalized linear models were born out of a desire to bring under one umbrella, a wide variety of regression models that span the spectrum from classical linear regression models for real valued data to models that are counts based data such as logit, or probit or Poisson, to models for survival analysis etc. just to give another example.

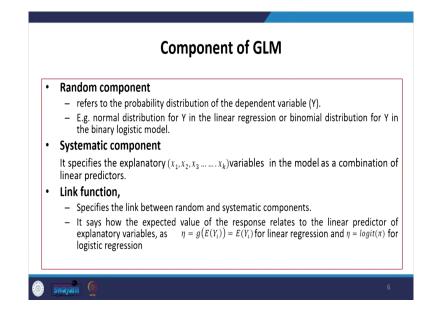
Poisson is like if you have certain count data such as number of consultations of doctors, number of visits to the doctors etc., those are simply count and not continuous series. The number of consultations cannot be just 1.8, 1.9 or 1.2. There is no fraction possibility. It has to be a count or absolute number and therefore, they are called Poisson kind of data.

(Refer Slide Time: 04:24)



So, GLM is made of a linear predictor of course, that is mentioned with a function called beta naught ( $\beta^{0}$ ) plus beta1\_X1I plus up to beta\_p Xi pi. And two functions are usually discussed in the case of GLM set up. One is called a link function, which describes how you get the mean or the expected mean with the population mean depends on the linear predictor i.e., the predictor value of the mean. And second one is the variance function, that describes how the variance that is the var of Yi depends on the mean value of it. And that can be of this value as per the equation whether that dispersion parameter theta is considered to be constant.

(Refer Slide Time: 05:20)



So, some of the components those are usually discussed in GLM are called random component, systematic component and link function. Random component is something that refers to the probability distribution of the dependent variable. For example, normal distribution for Y in the linear regression or binomial distribution for Y in the binary logistic function. In case of systematic component, it specifies the explanatory variable such as X1 to Xk variables in the model as a combination of linear predictors.

Link function specifies the link between random and systematic components. It says how the expected value of the response relates to the linear predictor of explanatory variables. And these are given in this equation with the function expected value of Y, how the function further transformation of that function is also related to the expected value of Y for linear regression. And that can be also applied through the logit function of the pi that is discussed accordingly we shared it as logistic regression

(Refer Slide Time: 06:34)

Model	Random	Link	Systematic
Linear Regression	Normal	Identity	Continuous
ANOVA	Normal	Identity	Categorical
ANCOVA	Normal	Identity	Mixed
Logistic Regression	Binomial	Logit	Mixed
Loglinear	Poisson	log	Categorical
Poisson Regression	Poisson	log	Mixed
Multinomial response	Multinomial	Generalized Logit	Mixed

So here is a summary of GLM following Agresti paper of chapter 4 published in 2013. It has given the GLM functions or models and how it is applied. As we know, the linear regression is applied when the data is continuous.

But an ANOVA or ANCOVA analysis of covariance or variance are also discussed in the context of categorical and mixed data. The link function gives the identity of it, the identity of the variable that has discussed the randomness, is in fact defined to be normal. The distribution is defined to be normal. So, in case of the specific GLM model like logistic loglinear, Poisson and multinomial response.

The random number table or randomness of this distribution is no more normal, they are either binomial Poisson or multinomial and according the link function it is defined as logit log function or generalized logit function. And therefore, it has mixed responses.

## (Refer Slide Time: 07:50)

<ul> <li>In linear regression model the dependent variable 'y' is expressed as a linear function of all the predictors.</li> <li>The underlying relationship between the response and predictors is linear.</li> <li>Also the error distribution should be normally distributed.</li> <li>GLM allows us to build a linear relationship between response and predictors, even though their underlying relationship is not linear.</li> <li>This is made possible by using link function, which links the response variable to linear model.</li> <li>Unlike linear regression models, the error distribution of the response variable need not to be normally distributed.</li> <li>The errors in the response variable are assumed to follow an exponential family of distribution (i.e. normal, binomial, Poisson, or gamma distributions).</li> </ul>	Linear Model	Generalized linear model
	<ul> <li>dependent variable 'y' is expressed as a linear function of all the predictors.</li> <li>The underlying relationship between the response and predictors is linear.</li> <li>Also the error distribution should be</li> </ul>	<ul> <li>relationship between response and predictors, even though their underlying relationship is not linear.</li> <li>This is made possible by using link function, which links the response variable to linear model.</li> <li>Unlike linear regression models, the error distribution of the response variable need not to be normally distributed.</li> <li>The errors in the response variable are assumed to follow an exponential family of distribution (i.e. normal, binomial,</li> </ul>

The comparison between linear model to GLM is mentioned here. In linear regression model, the dependent variable Y is expressed as a linear function of all the predictors. The underlining relationship between the response and the predictor predictors is linear. Also, the error distribution should be normally distributed.

Coming to the context of GLM, this allows us to build a linear relationship between the response and predictors, even though their underlining relationship is not linear but still it actually emerged with linear relationship with certain transformation. This is made possible by using the log link function. The way we have discussed in the previous slide which links the response variable to a linear model unlike linear regression model.

The error distribution of the response variable need not necessarily to be defined as normally distributed, the errors in the response variables are assumed to follow an exponential family of distribution.

Therefore, we say that- though the distribution are accordingly defined as may be of binomial, Poisson or gamma, but that are part of the exponential family of distribution.

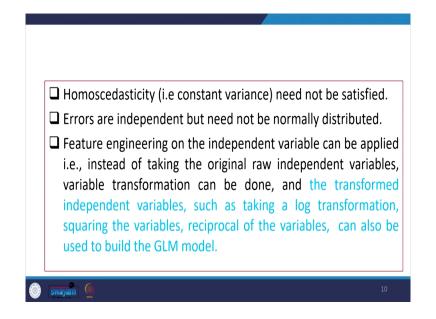
(Refer Slide Time: 09:12)

Data should be independent and random.
The response variable (Yi) does not need to be normally distributed, but the distribution is from an exponential family (e.g. binomial, Poisson, multinomial, normal).
The original response variable need not have a linear relationship with the independent variables, but the transformed response variable (through the link function) is linearly dependent on the independent variables.

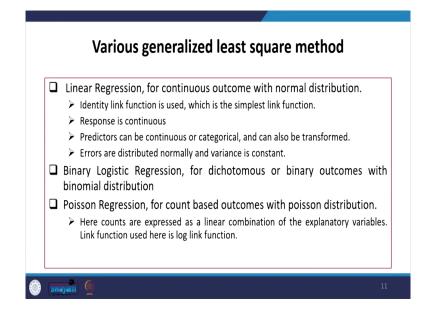
There are certain assumptions of GLM, the first one is that the data should be independent and random. The response variable does not need to be normally distributed. Response variable i.e., Yi need not be normally distributed, but the distribution is from an exponential family nor from a normally distributed family.

So, binomial count can be poisson or multinomial or may be normal, but that should still follow a exponential family. The original response variable need not have a linear relationship with the independent variables, but the transform response variable through the link function is linearly dependent on the independent variables that we have just discussed.

(Refer Slide Time: 10:00)



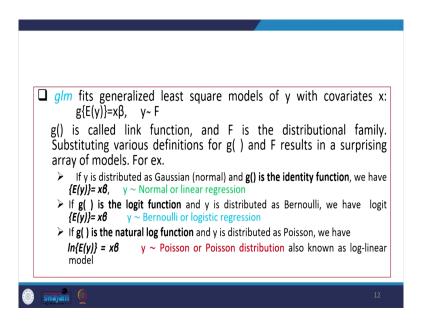
Another assumption is that, the homoscedasticity need not be satisfied since our distribution is following a different exponential format. So, its need not required to be satisfied. So, errors are independent, but need not be normally distributed. Feature engineering on the independent variable can be applied i.e., instead of taking the original raw independent variables, variable transformation can be done, and the transformed independent variable, such as taking a log transformation, squaring the variables, like reciprocal of the variables, can also be used to build the GLM model. So, there are various forms of transformation we do to develop a GLM model. (Refer Slide Time: 10:52)



Then there are various generalized least square methods like linear regression for continuous outcome with normal distribution. In that case, identity link function is used which is the simplest link function and the responses is continuous, predictors can be continuous or categorical or can be transformed, errors are distributed normally and variance is constant in case of linear regression. But we use binary logistic regression for dichotomous or binary outcomes with binomial distribution.

There are another type of generalized least square method. Poisson regression is used for count based outcomes with Poisson distribution. Here that counts are expressed as a linear combination of the explanatory variables. So, the link function that is used here is usually called the log link function.

(Refer Slide Time: 11:53)



So, we follow the command that is a GLM that feeds generalized least square models of Y with covariates of x. So, further g is taken as the transformation of the expected value of Y that should follow a normal distribution with X beta.

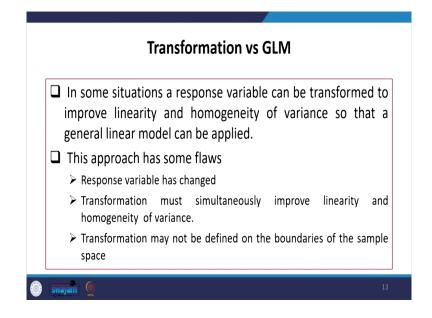
And Y follows the density function where F stands for the density function and g is called the link function because of some transformation and F is the distribution family or usually we present in the CDF type of functions. Substituting various definitions for g function and F, results in a surprising array of models.

So, if Y is distributed as Gaussian normal distribution and g is the identity function. We have the expected Y is equal to nothing but the normal distribution or normal or linear regression.

If g is a function of the variables with the transformation called g as the logit function. And Y is distributed as Bernoulli kind, we have the logit distribution that is otherwise called Bernoulli or logistic regression that follows a sigmoid kind of function where the concentration is usually at the extreme ends.

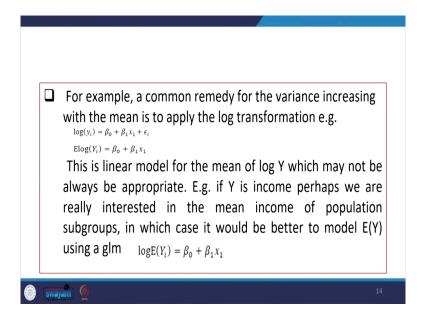
And, if g function is the natural log function. In that case, usually we go for the Poisson kind of distribution. So, Poisson distribution are for the count data. And log ln is natural log and expected value will convert the data to a linear function therefore, it is called log linear functions.

(Refer Slide Time: 13:45)



So, now we are just comparing between transformation and GLM. In some situations, a response variable can be transformed to improve linearity and homogeneity of variance. So, that a general linear model can be applied. This approach has some flaws like- response variable has changed, transformation must simultaneously improve linearity and homogeneity of variance, transformation may not be defined on the boundaries of the sample space that has been considered. As an example, a common remedy for the variance increasing with the mean is to apply the log transformation. If variance is there, simply taking the log transformation that will reduce the variance. So, there are other ways by which we can also do some expected value of the variance, log function can also be taken.

(Refer Slide Time: 14:47)



Here, this is a linear model for the mean of log Y which cannot be or may not be always appropriate. For example, if Y is income, perhaps we are really interested in the mean income of population subgroups, in which case it would be better to model expected Y using the GLM. GLM of this function i.e., log expected Yi is actually converting to a linear function.

(Refer Slide Time: 15:12)

	Advantage of GLM over simple linear regression
	□ We do not need to transform the response Y to have a normal distribution
	□ The choice of link is separate from the choice of random component thus have more flexibility in modeling
	□ If the link produces additive effects, then we do not need constant variance.
	The models are fitted via Maximum Likelihood estimation; thus optimal properties of the estimators.
9	swayaan @15

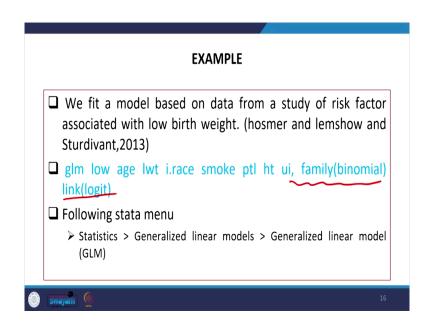
So, some advantages of GLM over the simple linear regression which are counted and necessary to define are- First one is that we do not need to transform the response Y to have a

normal distribution, the GLM command itself will automatically deal with this transformation.

Then, the choice of link is separate from the choice of random component and thus have more flexibility in modelling. So, the link function that we usually carry from the beginning of the command may be like log transformation or logit. At the beginning we need to think very carefully, but in this case with simple GLM command at the end we can specify the link function and derive the result.

Third, if the link produces additive effects, then we do not need constant variance. And fourth, the models are fitted via the MLE i.e., maximum likelihood estimation, thus optimal properties of the estimators are also defined.

(Refer Slide Time: 16:20)



One example we will also show you and on the basis of that you can just think of how you would be able to derive the result and be confident about your result. If we fit a model that is based on the data from a study of- risk factor associated with low birth weight as suggested in a paper by Hosmer and Lemshow and Sturdivant in 2013, where we say that the command is equal to glm then at first "dependent variable" i.e., low birth weight, and then age and then other variables, which are reference categorical variable that we can also mention.

Now, after all such variables we can give a comma and define which family they belong to. And if suppose we identify that they the data belongs to binomial family then your link command will be of link logit link(within bracket logit) and then that will be giving you the result.

So, here are other way of deriving the command- go to statistics, then GLM generalized, linear models then generalized linear model GLM will be displayed on the screen. I will also show that to you.

(Refer Slide Time: 17:46)

Stata Command for GLM	
Model of y as a function of x when y is a proportion glm y x, family(binomial)	
Logit model of y events occurring in 15 trials as a function of x glm y x, family(binomial 15) link(logit)	
Probit model of y events as a function of x using grouped data with group sizes n glm y x, family(binomial n) link(probit)	
Model of discrete y with user-defined family myfamily and link mylink glm y x, family(myfamily) link(mylink)	
Bootstrap standard errors in a model of y as a function of x with a gamma family and l glm y x, family(gamma) link(log) vce(bootstrap)	og link
🖲 swayani 🧕	17

Stata command for GLM- there are different command schedule. I am just going to show you that how you can operate just for your reference.

(Refer Slide Time: 17:56)

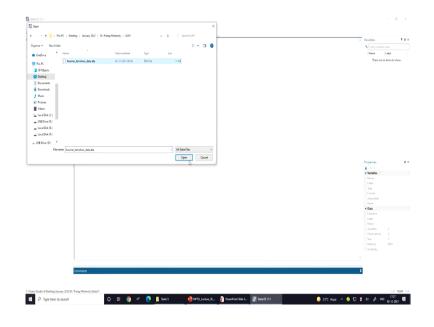
🔯 Stata/SI 15.1							0 X
File Edit Data Graphics Statistics	User Window Help						
🗑 🛛 🕲 🗍 🖼 - 🖬 - 📓 - 📓 - 📓 -	3 🗂 O - O						
Review T # X						<ul> <li>Variables</li> </ul>	τu×
A Liter commands have 0						A little constitue for	
# Command ,rc	5/5/5 ***	Copyright 1905-2017 StateCorp LLC				Name	bel
There are no items to show.	Statistics/Data Analysis	StateCorp				There are no it	work of arms
	Special Edition	4905 Lakeway Drive College Station, Texas 77045 USA					
	apaciar buicton	000-STATA-RC http://www.stata	. 608				
		979-696-4600 stata@stata.com					
		979-696-460) (fex)					
	from aggre and and a						
	Notes:						
	<ol> <li>Unicode is supported; see b</li> </ol>						
	2. Maximum number of variables	is set to 5000; see help set_maxrar.					
						Properties	# ×
						8.11	
	4					* Variables	
	-					Name	
						Label Type	
						Format	
						Value label	
						Notes	
						* Data	
						Literame	
						Variables	
						Memory Sorted by	ым
	Command						
C. Usen/Studio A/Desktop/January 2022/Dr.	Pratap Mohanty/Stata15						OF NUM OR
	o # 🧔	🛷 👩 🚺 Satulis  🏨	NPTIL Lecture, 30, 👔 PowerPoint Side	e S	22% Hum A	o 17 🛿 🕸 🖧 196 ,	17.27
P Type here to search	V H 🔮		reserved as	14 march 14	JSC Have A	o to a de de teo	18-12-2021

Here is the data. Now we want to operate it once GLM is here.

(Refer Slide Time: 18:05)

	Nis PC > Desitop > January 2022 > Di. Pratap	Mohanty >		< 0 P 1	earch Dr. Pratap Mohanty	-		<ul> <li>Variables</li> </ul>
ise • New fals	der				E + 0 0			1 filter variables here
A A	Name	Date modified	lype	Size				Name Label
	(RM)	68-12-2021 16-52	Filefulder					There are no items to sho
No PC	Lec (1907	18-15-2021 17:24	Lie folder					
10 Objects	Lec 10	22-10-2021 16-18	File folder					
Dektop	Lec 11	15-11-2021 12:12	File Folder					
Documents	Lec 12	27-10-2021 17:17	Filefolder					
Downloads	lec 1)	29-10-2021 17:47	f de tolder					
Music	Lecture_17_syntax_and_dofile	08-11-2021 16:21	File folder					
Pictures	Lecture, 18, dictionary/lie	01-11-2021 12:43	Filetolder					
Yidees	Lecture_30_Independence_of_indevent_a		File folder					
Local Disk (C)	Lecture, 37, Propensity, Score, Matching	26-11-2021 16:58	l de folder					
	Lecture26_ regression_model on healtra		File folder					
USB Drive (D.)	Lecture27_Qualitative_Dependent_variabl		l de folder					
Local Disk (E:)	Lecture31,balanced,and,unbalanced,he.		File Folder					
Local Disk (F.)	NPTEL_Lecture_24_Composite_Index	01-12-2021 17:01	File folder					
58 Drive (D.)	NPTEL_Lecture_25_measure_of_goodness	20-11-2021 16:53	Fiefolder					
	name			A11 (a.	ta files 🔍 🗸			
1.6	rane .							
				0	Ipen Cancel			
						)		Properties
								<ul> <li>Warlables</li> <li>Name</li> </ul>
								- Name Label Type
								- Name Label Totel Format
								Name Label Type Kornat Value label
								Name Label Type Kornat Value lubel Notes
								Nume Exhiri Tgar Komut Valac tabal Hister <b>4 Data</b>
								Name Labai Igen Lorust Value label Notee <b>Data</b> Laname
								Nume Label Losi Koma Value tabel Notes <b>- Duta</b> Frances Label
								Norme Labert Lyne Konnel Voron Konnen Labet Norien
								Nume Label Losi Koma Value tabel Notes <b>- Duta</b> Frances Label
								Nurre Label Loren Forman Work Indel Hotor Littel Hotor Woodba D
								Nove Likel Sori Fornat Usa kad Nore Edital Nore Usadan D Disenter D
								Nores Likel - Vert Source Source Likel Nores Likel Nores Disservators 0 Sos 0
								Nores Lines Lons Lons Notes - Date - Date - Lines - Li
								Nores Nores Variante Variante Nores - Disc - Dis
								Nores Nores Variante Variante Nores - Disc - Dis
								Nores Nores Variante Variante Nores - Disc - Dis
	Konneld							Nores Lines Lons Lons Notes - Date - Date - Lines - Li
	Connand							Nores Lines Lons Lons Notes - Date - Date - Lines - Li

(Refer Slide Time: 18:06)

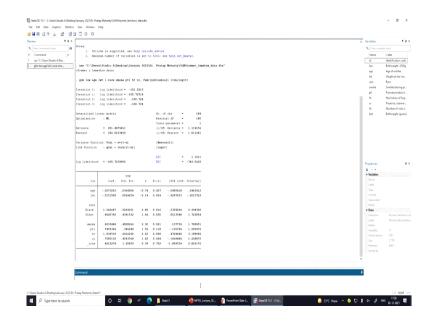


(Refer Slide Time: 18:09)

	stor User Hindow Help 2 🔄 🔲 🖸 - O			
Review 1 #			<ul> <li>Variables</li> </ul>	τü×
A Liter correrands been	0 (1)		A Liter venal	
# Command _rc		Copyright 1995-2017 StateCorp LLC	Nave	Label
1 une "Clibben/Studio A/Des	Statistics/Data Analysis	StateCorp	id	Identification code
		4903 Lakeway Drive	low	Bethweight - 200g
	Special Edition	College Station, Texas 77045 USA 800-STATA-PC http://www.stata.com	491	Age of mother
		979-696-4600 stateBstate.com	het rete	Weight at leat ree Race
		979+696-4601 (fax)	race .	Face Smoked during pr
			pti	Premature labor h
	A second the second second second		N	Has history of hyp
	succession of the second			Presence, uterine i
			ftv	Number of visits L
			bwt	Bithweight (grams)
	Notes:			
	<ol> <li>Unicode is supported; see</li> </ol>	help unicode_admice. # 1# met to 5000; mee help met maximut.		
	<ol> <li>Realmon number of veriable</li> </ol>	a 18 set to 10001 see nelp set maximal.		
	. use "C:\Users\Studio A\Desktop\Jan (Hosmer & Lemeshow data)	ary 2022Ur. Protog Monary/UDDoeser_leastics_dist.dix"		
			Properties	9 x
			Properties	ų x
			Properties B - + + Variables	¥ x
			Variables	ų x
			Variables	U X
			<ul> <li>Weinheiter</li> <li>Marine</li> <li>Naren</li> <li>Label</li> <li>Type</li> </ul>	ų x
			<ul> <li>Watables</li> <li>Name</li> <li>Ushel</li> <li>Sprit</li> <li>Formation</li> </ul>	ų x
			In the second	ų x
			<ul> <li>Watables</li> <li>Name</li> <li>Ushel</li> <li>Sprit</li> <li>Formation</li> </ul>	0 x
			Workables     Norma	
			Vice and Antiperiod States and Antiperi	Rosmer (arrobus d
			Variable	hoome iamshow di Hourse & Lempho
			<ul> <li>Variable</li> </ul>	Rome lember d Horne & Lember
			the statement of t	Korone (encolere de Horner & Lenephor 11
			A Statistic Andrew Label Gyst Konnell Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker Maker	Roome Semalou d Horne & Lenado 11
	Ģ		the statement of t	Norma Jamahara, di Hisana di Lenashar 11 13 2 521
	Ŀ		totalation     total     total	Annon Jenske d Henra & Lenske 11 13 2 Jak
	Ŀ		totalation     totalation     totalation     total     total	Annon Jenske d Henra & Lenske 11 13 2 Jak
			totalation     totalation     totalation     total     total	Annon Jenske d Henra & Lenske 11 13 2 Jak
	Command		totalation     totalation     totalation     total     total	Annon Jenske d Henra & Lenske 11 13 2 Jak
		. fautydownal Inkilopu	totalation     totalation     totalation     total     total	Annon Jenske d Henra & Lenske 11 13 2 Jak
	Command	L. KALIY SHAMAD LIALISYU	totalation     totalation     totalation     total     total	Annon Jenske d Henra & Lenske 11 13 2 Jak
Citiza Guda K Batagiana 20	Connand gim low age let 1.race maske pil ht o	1. Gally@lows40 11480990	totalation     totalation     totalation     total     total	Annon Jenske d Henra & Lenske 11 13 2 Jak

So, the sample data we are just operating here for your reference, we have all those data and we will also open the do file simultaneously or we can just directly run as per the command we have given here is the command section. So, we simply copy from do file and you can run for your own practice.

## (Refer Slide Time: 18:27)



So, here is the result on your screen. Now, this would have given you the idea that how the link function is defined and how the families of the data are also defined. Now, this result has given us the coefficients and their significant significance levels and on the basis of that we can take decisions of interpreting the results. So, it seems that these results are similar to that of the logit function.

A logit command that we have already specified can also be given in same way. Now you mark that, this is also called the coefficient or the odd ratio. So, this has a combined operation and we need not go for logit separately. Similarly, there are other Stata commands for GLM. Once the link function and the family of it is known that your family is binomial, you can give this command.

When your family is binomial, but you are also very clearly about the function to be applied you can give that. So, GLM with your dependent variable, independent variable should be specified when the is families of binomial type. But your distribution is more probabilistic type or distribution follows normal distribution, error term might have followed a normal distributive structure then in that case you can apply accordingly.

Similarly, other approaches are there that you can easily explore. Gamma function- if it is there then bootstrapping through this particular command is going to be helpful for you. Another approach that you can use is through the window through is statistics.

(Refer Slide Time: 20:17)

1 3 · 1 · 6	Summales tables and tests   Linear models and related	<ul> <li>Veriables</li> </ul>	1.
	Bray success	A Liter varia	
nd	Ordinal outcomes + ported; see help unicode_advice.	Nava	Label
uen/Studio A/Des	Categorial outcomes , of variables is set to 5000; see help set maxmar.	id a	Identification code
het ivace smo.	Court outcomes 'Desktop)January 2022\Dr. Pratap Mohanty\GIM\hosmer_lemshow_data.dta"	low	Bithweight-200g
	Fractional outcomes	age	Age of mother
	Generalized Innear model  Generalized Innear model (GUM) Time server (GMTer the binomed Rendy A (Logi 1)	het	Weight at last ere.
	The sets of second and se	race produc	Race Smoked during pr
	construction of the state of th	pt)	Premature labor h
		N	Has history of hug
	to gran a provide and the second s		Presence, uterine is
	Multilevel mixed effects models	fiv	Number of visits L
	Sunnal analysis  Bayesian generalized linear models (GLM)  - 389	but	Bethweight (grans
	Epidemology and related + Exyrcura multificati GLM = 180		
	Endoprious covariation in Endoprious Covariante State Provide Tarring Provide		
	Sample selection models ' 0233425 (1/df) Pearson = 1.011241		
	Featwent effects		
	S(M (dructural equation modeling) + * * (1-*) (Bernos111)		
	(CA (Intent class analysis) = Im (m/(1-m)) (Log(12)		
	(MM (inte miture model) + AIC + 1.1611		
	IRI (tem impone theny) 7239956 82C = -742.0665	Properties	
	Survey data analysis +	8 · · ·	
	Multiple imputation 01H	+ Variables	
	Nonperametric analysis ' Std. Ezr. z P>(z) (954 Conf. Interval)	Name	
	Multivariate analysis +	Label	
	Eract statistics , 0 .0364504 -0.74 0.4570985418 .0443412	1,04 Format	
	Reserving , 0.0049259 -2.19 0.02902872530015763	Formal Value label	
	Power and sample size	Noter	
	Bayesian analysis * 7 .5264101 2.40 0.016 .2309024 2.294392	+ Data	
	Pedestination 2 .4391532 1.96 0.050 .0013548 1.722804	Liename	hoome, levels
	Other , 4008266 2.30 0.021 .137739 1.709951	Likel	Hosmer & Lev
-	pt1 .5418366 .346249 1.56 0.118136799 1.220472	Noter Verables	
	ht 1.032510 .6916292 2.65 0.000 .4769494 3.100006	Venables Observation	
	ul .7583135 .4393768 1.65 0.0991418484 1.638875	Sat	2.776
	_cons .4612239 1.20459 0.30 0.702 -1.099729 2.022176	Memory	6454
		Sorted by	
	P		
	Command		

You can go to statistics then it has what is called generalized linear models.

(Refer Slide Time: 20:29)

view T.B.X															<ul> <li>Variables</li> </ul>	та
l itser commands have	Notesi 1 Date	ode is suppo	rrad: and he	In union	de admire										🐴 litter variab	
Command _rc						e help set_max	THE.								Name	Label
use "Cilbios/Studio A/Des															id	Identification code
gim low age lut i.race smo	(Nomer & Lene		lesktop\Janua	ry 2022/	Dr. Prata	p Mohanty\GLM\8	isemer_les	abov_dat	ta.dta"						low	Bithweight - 2500g
	Universe a bene														age last	Age of mother Weight at last me
	. gim low age	lwt i.race s	moke pt1 ht	01. Capt	lyibinomi	<li>1) link(logit)</li>									rece	Rect
				ч											sweke	Smaked during pr.
	Iteration 0: Iteration 1:		005 = -101. 005 = -100.7												pti	Premature labor h.
	Iteration 2:		ood = -100			🗊 gin - Generalize							×	1	N	Has history of hyp.
	Iteration 3:	log likelih	ood = -100	.724		[]] (pri - cenerate	d trear mode						~		ui .	Presence, uterine i.
						Model Model 2	by/Kin W	eights 56.19	lobust Repr	oting Mai	imitation				Rv bwt	Number of visits L.
	Generalized linear models No. Optimization : ML Resi						ie inte	pendent var	abler						bwt	Birthweight (grams)
					Scale	1 1	1						v			
	Deviance Pearson	<ul> <li>201.44</li> <li>102.02</li> </ul>			(1/8f (1/8f								Land Con-			
	restron	- 102.02	0.0425		(1/01	link choices	Gaussian	Inverse Gaussian	Energial	Poisson	Negative	Gamma				
	Variance funct				(Bern											
	Link function	: ĝ(a) =	ln(0/(1-0))		(Log1)	log			0							
					AIC			-	0		0					
	Log likelihood	-100.72	39956		810	Logi Posh			0						Properties	
															B + +	
			OIN			C. log log		0			0	0			<ul> <li>Variables</li> </ul>	
	104	Coef.			P>121	Fore		0		0	0				Name	
						Odds pewe									Label	
	a ge	0271003	.0364504	-0.74	0.457	Neg binom					0				3ge Format	
	let	6121264	.0049259	.4.19	0.029	Log log									Value label	
	zace.					Log-comp			0						Notes	
	Black	1.262647	.5264101	2.40	0.016										# Data	
	Other	.8620192	.4391532	1.96	0.050					_					Liename	hosmer, lencho
	ancke	. 9233440	4038266	2.30	0.021	00					OK	Cancel	Submit		Label	Hosner & Leni
	pt1	.5410366	.346249	1.56	0.118		1.220472							,	Variables	
	ht ut	1.032510	.6916292	2.65	0.008		3.109086								Obsenations	100
	01	. 4612239	1.20459	0.30	0.702		2.022176								Sax	2.776
															Memory	64M
															Sorted by	
															×	
	Command															

Then here is your GLM now you can specify each of it, i.e., the dependent and independent variable and with that it is possible for you and some family name you can give it. So, with the family name like as I already discussed whether it is a Gaussian type, binomial, Poisson, gamma etc. and your result will be displayed on your screen the way we have shown it to you.

So, these is all in this lecture. I think at this moment we will come up with further details in our next episode and with this we are going to stop here. And I look forward to your comments or queries and we will be happy to address it.

Thank you.