# Poisson and Negative Binomial Regression

Prediction models. But not for continuous or dichotomous outcomes. In these analyses, your dependent variable is a frequency of some event. How many times did some specific thing happen over a specified period of time? Relatively small counts. How many points were scored in the hockey game? How many visits to the hospital in a year? How many times did you go to the gym this month? That sort of thing.



# Before running your Poisson:

Linear regressions are way easier and faster to run. And they give you similar information. So it's not a bad idea to explore a phenomenon first by using linear regression (designing, defining, and *refining your model*) before throwing away those outputs and running that same model as a Poisson or negative binomial regression... you know, so it's *accurate*.

Once you have your model worked out, you should examine the dependent variable a bit.



# Evaluating your dependent variable:

This example will use hospital data, evaluating the frequency of older patients being admitted for fall-related injuries. The dependent variable will be the number of falls experienced over a specified period of time.

It is important to analyze your dependent variable ahead of time so you know which model (Poisson or negative binomial) is most appropriate.



SPSS Sta	tistics	File	Edit	View	Data	Transform	Analyze	Graphs	Utilities	Extensions	Window Help	
							Compar General	<mark>tive Statis</mark> e Means Linear Mo	del		III    Frequencies      Image: Construction of the second seco	
In SPSS then the option,	Desc	ript	ive S	Statis	stics		Mixed N Correlat Regress Loglinea Classify Dimens Scale Nonpara Forecas Survival	te sion ar ion Reduct ametric Te sting Response ilation	ion sts		<ul> <li>Crosstabs</li> <li>TURF Analysis</li> <li>Ratio</li> <li>P-P Plots</li> <li>Q-Q Plots</li> </ul>	

ROC Curve...

Spatial and Temporal Modeling...



# This box comes up.

Every variable in your database is in that left column.



Pacific Paper Mill.com	
------------------------	--

	Explore	
	Dependent List:	Charlinting
Subject_Number		Statistics
🖁 Admission_Date		Plots
Avg temp between 6:54am and 6:54pm h		
Light_snow_today	•	Options
Heavy_Snow_today		
Light_Snow_last_48hr		
Heavy_Snow_last_48hr		
🗞 Heavy_Snow_yesterdayNotToday		
🗞 Any_snow_yesterdayNotToday	Factor List:	
🗞 Rain_today		
Avg wind speed between 6:54am and 6:5		
Admission_Month		
Admission_Season	<b>•</b>	
🖀 Discharge_Date		
Length of Stay [LOS]		
Where the patient came from [Where_patie		
Did the patient come from home? [Came_f		
🗞 Sex [Sex]	Label Cases by:	
🖗 Age	Image: Second	
Display		
O Both ○ Statistics ○ Plots		
? Reset Paste		Cancel OK

Drag any dependent variables you might use in the regression models into the Dependent List.

Each regression analysis will only have one dependent variable. But you might run multiple analyses. If so, you'll be "Explore"ing more than one variable here.

Once selected, click OK.



Pacific Paper Mill.com
------------------------

		Explore	
		Dependent List:	Statistics
Subject_Number		Num_of_Previous_Falls	Statistics
Admission_Date		Num_Return_ER_Visits	Plots
Avg temp between 6:54am and 6:54pm h			
Light_snow_today	•		Options
Heavy_Snow_today			
Light_Snow_last_48hr			
Heavy_Snow_last_48hr			
Heavy_Snow_yesterdayNotToday			
Any_snow_yesterdayNotToday		Factor List:	
Rain_today			
Avg wind speed between 6:54am and 6:5			
Admission_Month			
Admission_Season	•		
Discharge_Date			
Length of Stay [LOS]			
Where the patient came from [Where_patie			
Did the patient come from home? [Came_f			
Sex [Sex]		Label Cases by:	
Age	•		
Display			
• Both 🔿 Statistics 🔷 Plots			
2 Decet			Canaal
? Reset Paste			Cancel OK
	_		

You'll get an output that looks like this:

If the mean and the variance are roughly equal (the mean and spread of the number of hospital visits) are equal, you're probably fine to use Poisson. If the variance is larger than the mean, that's called "overdispersion" and you have to use negative binomial model.

It's possible to have "underdispersion" (as seen in the top variable), but it won't be by much.

You'd like those numbers to be equal, but close to equal is good enough.

			Statistic	Std. Erro
Num_of_Previous_Falls	Mean		1.92	.053
	95% Confidence Interval	Lower Bound	1.82	
	for Mean	Upper Bound	2.02	
	5% Trimmed Mean		1.78	
	Median		1.00	
	Variance		1.713	
	Std. Deviation		1.309	
	Minimum		0	
	Maximum	11		
	Range	11		
	Interquartile Range	2		
	Skewness	2.048	.099	
	Kurtosis	7.708	.197	
Num_Return_ER_Visits	Mean	.53	.038	
	95% Confidence Interval	Lower Bound	.45	
	for Mean	Upper Bound	.60	
	5% Trimmed Mean		.39	
	Median		.00	
	Variance		.891	
	Std. Deviation		.944	
	Minimum		0	
	Maximum		9	
	Range		9	
	Interquartile Range		1	
	Skewness		3.020	.099
	Kurtosis		15.463	.197

Descriptives



<b>É SPSS Statistics</b> File Edit View Data Transform	Analyze Graphs Utilities Extensions Window Help
	Reports     Image: Compare Means
	General Linear Model
	Generalized Linear Models 🕨 🕅 Generalized Linear Models
Next, run your Poisson or negative binomial regression. Finding them in the menu:	Mixed Models Correlate Regression Loglinear Classify Dimension Reduction Scale Nonparametric Tests Forecasting Survival Multiple Response Simulation Quality Control ROC Curve
	Spatial and Temporal Modeling

With linear and logistic regression, you'll find them in the "Regression" tab. You won't find Poisson or negative binomial regression there. They're in the "Generalized Linear Models" tab. Select that option.



There are several options here.

If you're going to do linear or logistic, just do those the normal way.

Poisson, negative binomial, or custom (customizing your negative binomial) can be done here.

Poisson is selected.

We'll start there.



Pacific Paper Mill.com
------------------------

ei								
Choose one of the mode	I types listed	below or spe	cify a cust	tom combinat	ion of distrib	ution and link	functio	n.
🔗 Scale Response 👘				📲 Ordinal R	esponse			
🔵 Linear				🔵 Ordina	logistic			
🔵 Gamma with log lin	nk			🔵 Ordina	l probit			
Counts				🗢 Binary Res	ponse or Eve	nts/Trials Dat		
💿 Poisson loglinear				🔵 Binary	logistic			
O Negative binomial	with log link			🔵 Binary	probit			
Mixture				🔵 Interva	l censored su	rvival		
O Tweedie with log li	nk							
Tweedie with ident	ity link							
6 /								
Custom								
Custom								
Distribution:	Normal	\$	Link fu	nction:	dentity			0
Parameter			D					
<ul> <li>Specify va</li> </ul>	lue		Po	wer:				
Value: 1								
🔘 Estimate v	alua							
Estimate	arue							
	-						_	
? Reset	Paste						Canc	el OK

Click on the Response tab at the top.

From the Variables list on the left, select your dependent (outcome) variable.



Pacific Paper Mill.com

Subject, Number         Admission_Date         Avg temp between 6:54am and 6:54         Light_snow_today         Light_snow_today         Heavy_Snow_tast_48hr         Heavy_Snow_yesterdayNotToday         Any snow-yesterdayNotToday         Any snow-yesterdayNotToday         Any snow-yesterdayNotToday         Admission_Month         Admission_Season         Discharge_Date         Length of Stay [LOS]         Where the patient come from home? [Ca         Did the patient come from home? [Ca         Sex [Sex]         Age 80 or older [Age_Threshold_80]         Hendrich Score (High Risk Patients) [         Hendrich Score (High Risk Patients) [         Hendrich Score (High Risk Patients) [         Hendrich_Score_Without_Sex         Hendrich_Score_Without_Sex	Type of Model Response Pre	dictors	Model	Estimation	Statistics	EM Means	Save	Export
<ul> <li>Admission_Date</li> <li>Arwission_Date</li> <li>Avg temp between 6:54am and 6:54</li> <li>Light_snow_today</li> <li>Light_snow_today</li> <li>Light_snow_last_48hr</li> <li>Heavy_Snow_vesterdayNotToday</li> <li>Arw, snow_yesterdayNotToday</li> <li>Arw, snow_yesterdayNotToday</li> <li>Arw, snow_yesterdayNotToday</li> <li>Arwission_Season</li> <li>Dicharge_Date</li> <li>Length of Stay (LOS)</li> <li>Where the patient came from (Where</li> <li>Dickarge_Date</li> <li>Length of Stay (LOS)</li> <li>Where the patient came from (Where</li> <li>Dickarge_Date</li> <li>Length of Stay (LOS)</li> <li>Where the patient came from (Where</li> <li>Dickarge_Date</li> <li>Length of Stay (LOS)</li> <li>Where the patient came from (Where</li> <li>Dickarge_Date</li> <li>Length of Stay (LOS)</li> <li>Where the patient came from (Where</li> <li>Dickarge_Date</li> <li>Length of Stay (LOS)</li> <li>Where the patient came from (Where</li> <li>Dickarge_Date</li> <li>Length of Stay (LOS)</li> <li>Where the patient came from (Where</li> <li>Sex [Sex]</li> <li>Age 80 or older (Age_Threshold_80)</li> <li>HENDRICH Fall Risk Score (HENDRIC</li> <li>Hendrich Score (High Risk Patients) [</li> <li>Hendrich Score (High Risk Patients) [</li> <li>Hendrich_Score_Plus_age80</li> </ul>	/ariables:	Depend	dent Varia	able				
<ul> <li>Ardmission_Date</li> <li>Ary temp between 6:54am and 6:54</li> <li>Light_snow_today</li> <li>Light_snow_today</li> <li>Light_snow_today</li> <li>Light_snow_tast_48hr</li> <li>Heavy_Snow_yesterdayNotToday</li> <li>Any_snow_yesterdayNotToday</li> <li>Any_snow_yesterdayNotToday</li> <li>Ary snow_yesterdayNotToday</li> <li>Ary snow_yesterdayNotToday</li> <li>Ary snow_yesterdayNotToday</li> <li>Ary snow_yesterdayNotToday</li> <li>Ary snow_seterdayNotToday</li> <li>Ary snow_seterdayNotToday</li> <li>Ary snow_yesterdayNotToday</li> <li>Ary snow_yesterdayNotToday</li> <li>Ary snow_yesterdayNotToday</li> <li>Ary snow_yesterdayNotToday</li> <li>Ary snow_seterdayNotToday</li> <li>Ary state in the st</li></ul>	🗗 Subject_Number		Depender	nt Variable:				
<ul> <li>Light_Show_today</li> <li>Light_Show_last_48hr</li> <li>Heavy_Snow_yesterdayNotToday</li> <li>Any_snow_yesterdayNotToday</li> <li>Rain_today</li> <li>Any_snow_yesterdayNotToday</li> <li>Rain_today</li> <li>Admission_Month</li> <li>Admission_Season</li> <li>Discharge_Date</li> <li>Length of Stay (LOS)</li> <li>Where the patient came from [Where</li> <li>Did the patient came from home? [Ca</li> <li>Sex [Sex]</li> <li>Age bracket: every 5 years [Every_5y</li> <li>Age bracket: every 5 years [Every_5y</li> <li>Age bracket: every 5 years [Every_5y</li> <li>Hendrich Score (High Risk Patients) [</li> <li>Hendrich Score (High Risk Patients) [</li> <li>Hendrich_Score_Plus_age80</li> </ul>	🔏 Admission_Date							
<ul> <li>Light_Show_today</li> <li>Light_Show_last_48hr</li> <li>Heavy_Snow_yesterdayNotToday</li> <li>Any_snow_yesterdayNotToday</li> <li>Rain_today</li> <li>Any_snow_yesterdayNotToday</li> <li>Rain_today</li> <li>Admission_Month</li> <li>Admission_Season</li> <li>Discharge_Date</li> <li>Length of Stay (LOS)</li> <li>Where the patient came from [Where</li> <li>Did the patient came from home? [Ca</li> <li>Sex [Sex]</li> <li>Age bracket: every 5 years [Every_5y</li> <li>Age bracket: every 5 years [Every_5y</li> <li>Age bracket: every 5 years [Every_5y</li> <li>Hendrich Score (High Risk Patients) [</li> <li>Hendrich Score (High Risk Patients) [</li> <li>Hendrich_Score_Plus_age80</li> </ul>								
<ul> <li>Light_Silow_last_Holin</li> <li>Heavy_Snow_last_48hr</li> <li>Heavy_Snow_yesterdayNotToday</li> <li>Any_snow_yesterdayNotToday</li> <li>Rain_today</li> <li>Avg wind speed between 6:54am an</li> <li>Admission_Month</li> <li>Admission_Season</li> <li>Discharge_Date</li> <li>Variable</li> <li>Obicharge_Date</li> <li>Vere the patient came from [Where</li> <li>Did the patient come from home? [Ca</li> <li>Sex [Sex]</li> <li>Age bracket: every 5 years [Every_5y</li> <li>Age 80 or older [Age_Threshold_80]</li> <li>Hendrich Score [HENDRIC</li> <li>Hendrich Score [HENDRICH</li> <li>Hendrich Score [HENDRICH</li> <li>Hendrich Score [HENDRICH</li> <li>Hendrich Score Plus_age80</li> </ul>	💑 Light_snow_today	Categ	ory orde	r (multinomial	only):	Asce	nding	0
<ul> <li>Heavy_Snow_last_48hr</li> <li>Heavy_Snow_yesterdayNotToday</li> <li>Any_snow_yesterdayNotToday</li> <li>Any_snow_yesterdayNotToday</li> <li>Rain_today</li> <li>Avg wind speed between 6:54am an</li> <li>Admission_Month</li> <li>Admission_Season</li> <li>Discharge_Date</li> <li>Variable</li> <li>O Variable</li> <li>Variable</li> <li>Variable</li> <li>Fixed value</li> <li>Number of Trials:</li> <li>Scale Weight Variable:</li> <li>Mendrich Score (High Risk Patients) [</li> <li>Hendrich Score_Plus_age80</li> </ul>	💑 Heavy_Snow_today	-	6.0		(D)			
<ul> <li>Heavy_Snow_yesterdayNotToday</li> <li>Any_snow_yesterdayNotToday</li> <li>Any_snow_yesterdayNotToday</li> <li>Rain_today</li> <li>Avg wind speed between 6:54am an</li> <li>Admission_Month</li> <li>Admission_Season</li> <li>Discharge_Date</li> <li>Variable</li> <li>Scale Weight Variable:</li> <li>Scale Weight Variable:</li> <li>Scale Weight Variable:</li> <li>Variable</li> <li< td=""><td>💑 Light_Snow_last_48hr</td><td></td><td></td><td>dent Variable</td><td>(Binomial Dis</td><td>tribution Only</td><td>/)</td><td></td></li<></ul>	💑 Light_Snow_last_48hr			dent Variable	(Binomial Dis	tribution Only	/)	
<ul> <li>Any_snow_yesterdayNotToday</li> <li>Rain_today</li> <li>Avg wind speed between 6:54am an</li> <li>Admission_Month</li> <li>Admission_Season</li> <li>Discharge_Date</li> <li>Variable</li> <li>Variable</li> <li>Trials Variable:</li> <li>Trials Variable:</li> <li>Fixed value</li> <li>Number of Trials:</li> <li>Sex [Sex]</li> <li>Age bracket: every 5 years [Every_5y</li> <li>Hendrich Score (High Risk Patients) [</li> <li>Hendrich NEW Categories [HENDRICH</li> <li>Hendrich Score_Without_Sex</li> <li>Hendrich_Score_Plus_age80</li> </ul>		• B	inary					
<ul> <li>Any_snow_yesterdayNotToday</li> <li>Rain_today</li> <li>Avg wind speed between 6:54am an</li> <li>Admission_Month</li> <li>Admission_Season</li> <li>Discharge_Date</li> <li>Length of Stay [LOS]</li> <li>Where the patient came from [Where</li> <li>Did the patient came from home? [Ca</li> <li>Sex [Sex]</li> <li>Age</li> <li>Decade of life [DecadeOfLife]</li> <li>Age boar older [Age_Threshold_80]</li> <li>HENDRICH Fall Risk Score [HENDRIC</li> <li>Hendrich Score (High Risk Patients) [</li> <li>Hendrich Score (High Risk Patients) [</li> <li>Hendrich_Score_Without_Sex</li> <li>Hendrich_Score_Plus_age80</li> </ul>			Refere	nce Category.				
<ul> <li>Avg wind speed between 6:54am an</li> <li>Admission_Month</li> <li>Admission_Season</li> <li>Discharge_Date</li> <li>Length of Stay [LOS]</li> <li>Where the patient came from [Where</li> <li>Did the patient come from home? [Ca</li> <li>Sex [Sex]</li> <li>Age</li> <li>Decade of life [DecadeOfLife]</li> <li>Age 80 or older [Age_Threshold_80]</li> <li>HENDRICH Fall Risk Score [HENDRIC</li> <li>Hendrich Score (High Risk Patients) [</li> <li>Hendrich NEW Categories [HENDRICH</li> <li>Hendrich Score_Without_Sex</li> <li>Hendrich_Score_Plus_age80</li> </ul>								
<ul> <li>Admission_Month</li> <li>Admission_Season</li> <li>Discharge_Date</li> <li>Length of Stay [LOS]</li> <li>Where the patient came from [Where</li> <li>Did the patient come from home? [Ca</li> <li>Sex [Sex]</li> <li>Age</li> <li>Decade of life [DecadeOfLife]</li> <li>Age s0 or older [Age_Threshold_80]</li> <li>HENDRICH Fall Risk Score [HENDRIC</li> <li>Hendrich Score (High Risk Patients) [</li> <li>Hendrich NEW Categories [HENDRICH</li> <li>Hendrich Score_Without_Sex</li> <li>Hendrich_Score_Plus_age80</li> </ul>		0 N	umber of	events occur	ring in a set o	of trials		
<ul> <li>Admission_Season</li> <li>Discharge_Date</li> <li>Length of Stay [LOS]</li> <li>Where the patient came from [Where</li> <li>Did the patient come from home? [Ca</li> <li>Sex [Sex]</li> <li>Age</li> <li>Decade of life [DecadeOfLife]</li> <li>Age 80 or older [Age_Threshold_80]</li> <li>HENDRICH Fall Risk Score [HENDRIC</li> <li>Hendrich Score (High Risk Patients) [</li> <li>Hendrich Score_Without_Sex</li> <li>Hendrich_Score_Plus_age80</li> </ul>		ſ	Trials —					
<ul> <li>Admission_Season</li> <li>Discharge_Date</li> <li>Length of Stay [LOS]</li> <li>Where the patient came from [Where</li> <li>Did the patient come from home? [Ca</li> <li>Did the patient come from home? [Ca</li> <li>Sex [Sex]</li> <li>Age</li> <li>Decade of life [DecadeOfLife]</li> <li>Age so or older [Age_Threshold_80]</li> <li>HENDRICH Fall Risk Score [HENDRIC</li> <li>Hendrich Score (High Risk Patients) [</li> <li>Hendrich Score_Without_Sex</li> <li>Hendrich_Score_Plus_age80</li> </ul>			Varia	able				
<ul> <li>Discharge_Date</li> <li>Length of Stay [LOS]</li> <li>Where the patient came from [Where</li> <li>Did the patient come from home? [Ca</li> <li>Did the patient come from home? [Ca</li> <li>Sex [Sex]</li> <li>Age</li> <li>Decade of life [DecadeOfLife]</li> <li>Age bracket: every 5 years [Every_5y</li> <li>Age 80 or older [Age_Threshold_80]</li> <li>HENDRICH Fall Risk Score [HENDRIC</li> <li>Hendrich Score (High Risk Patients) [</li> <li>Hendrich Score_Without_Sex</li> <li>Hendrich_Score_Plus_age80</li> </ul>	•••							
<ul> <li>Where the patient came from [Where</li> <li>Did the patient come from home? [Ca</li> <li>Sex [Sex]</li> <li>Age</li> <li>Decade of life [DecadeOfLife]</li> <li>Age bracket: every 5 years [Every_5y</li> <li>Age 80 or older [Age_Threshold_80]</li> <li>HENDRICH Fall Risk Score [HENDRIC</li> <li>Hendrich Score (High Risk Patients) [</li> <li>Hendrich Score_Without_Sex</li> <li>Hendrich_Score_Plus_age80</li> </ul>			- 🗣 T					
<ul> <li>Did the patient come from home? [Ca</li> <li>Sex [Sex]</li> <li>Age</li> <li>Decade of life [DecadeOfLife]</li> <li>Age s0 or older [Age_Threshold_80]</li> <li>HENDRICH Fall Risk Score [HENDRIC</li> <li>Hendrich Score (High Risk Patients) [</li> <li>Hendrich NEW Categories [HENDRICH</li> <li>Hendrich_Score_Without_Sex</li> <li>Hendrich_Score_Plus_age80</li> </ul>								
<ul> <li>Sex [Sex]</li> <li>Age</li> <li>Decade of life [DecadeOfLife]</li> <li>Age bracket: every 5 years [Every_5y</li> <li>Age 80 or older [Age_Threshold_80]</li> <li>HENDRICH Fall Risk Score [HENDRIC</li> <li>Hendrich Score (High Risk Patients) [</li> <li>Hendrich NEW Categories [HENDRICH</li> <li>Hendrich_Score_Without_Sex</li> <li>Hendrich_Score_Plus_age80</li> </ul>								
Age     Scale Weight     Scale Weight     Scale Weight     Scale Weight     Scale Weight Variable:			1	Number of Tri	als:			
Scale Weight          Scale Weight         Age bracket: every 5 years [Every_5y         Age 80 or older [Age_Threshold_80]         HENDRICH Fall Risk Score [HENDRIC         Hendrich Score (High Risk Patients) [         Hendrich NEW Categories [HENDRICH         Hendrich_Score_Without_Sex         Hendrich_Score_Plus_age80								
Age bracket: every 5 years [Every_5y Age 80 or older [Age_Threshold_80] HENDRICH Fall Risk Score [HENDRIC Hendrich Score (High Risk Patients) [ Hendrich_Score_Without_Sex Hendrich_Score_Plus_age80								
<ul> <li>Age 80 or older [Age_Threshold_80]</li> <li>HENDRICH Fall Risk Score [HENDRIC</li> <li>Hendrich Score (High Risk Patients) [</li> <li>Hendrich NEW Categories [HENDRICH</li> <li>Hendrich_Score_Without_Sex</li> <li>Hendrich_Score_Plus_age80</li> </ul>		Scale W	/eight —					
<ul> <li>HENDRICH Fall Risk Score [HENDRIC</li> <li>Hendrich Score (High Risk Patients) [</li> <li>Hendrich NEW Categories [HENDRICH</li> <li>Hendrich_Score_Without_Sex</li> <li>Hendrich_Score_Plus_age80</li> </ul>			Scale Wei	ght Variable:				
<ul> <li>Hendrich Score (High Risk Patients) [</li> <li>Hendrich NEW Categories [HENDRICH</li> <li>Hendrich_Score_Without_Sex</li> <li>Hendrich_Score_Plus_age80</li> </ul>				-				
Hendrich NEW Categories [HENDRICH Hendrich_Score_Without_Sex Hendrich_Score_Plus_age80								
Hendrich_Score_Without_Sex Hendrich_Score_Plus_age80								
Hendrich_Score_Plus_age80								
<pre> // Hendrich_Score_WithoutSexPlusAge //  //  //  //  //  //  //  //  //</pre>								
	Hendrich_Score_WithoutSex_PlusAge							

Put that variable in the Dependent Variable box.

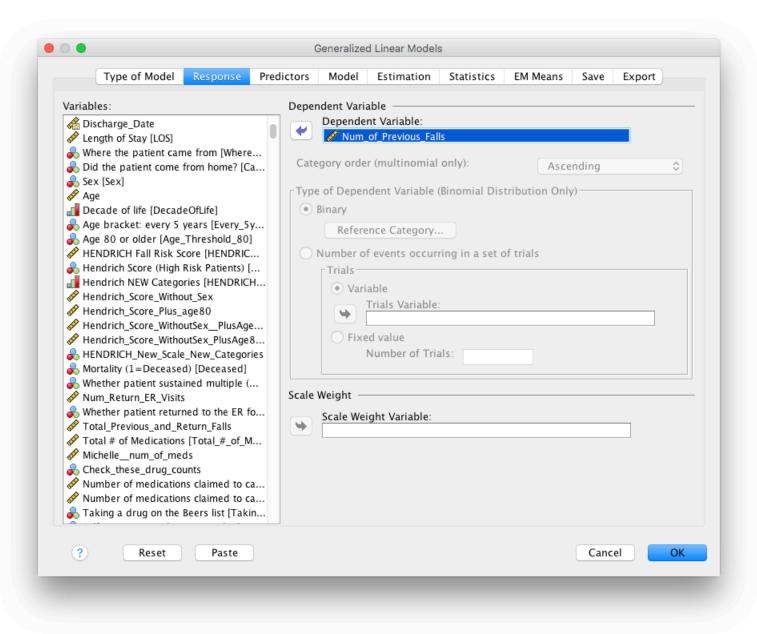
It has to be a count variable. Don't put a continuous or dichotomous variable there.

What I've placed in the dependent box is the number of times a patient has been admitted to a hospital for a fall-related injury.

Don't hit "OK" yet. First, go through all of the tabs at the top.



Pacific Pay	perMill.com
-------------	-------------



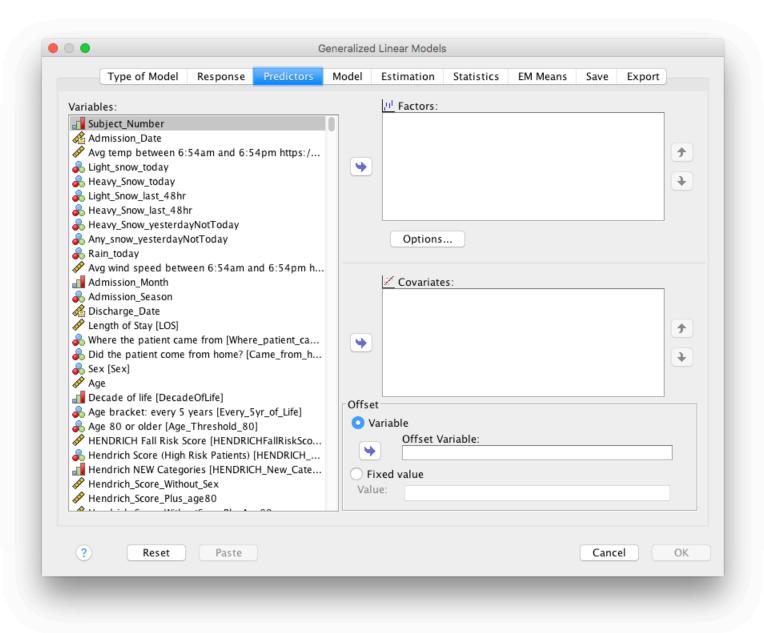
#### Predictors tab.

Now you select your predictors (sometimes called "independent variables", but that's a bit of a misnomer as these variables aren't often *independent of each other*; if they're strongly – or even mediumly – correlated with each other, they aren't "independents" but they're still predictors).

If you have fewer than 30 total subjects, it's a little bit irresponsible to include multiple predictors. Aim for at least 20 subjects per variable. 15 is a little low. 10 is really pushing it. 50 is great.



Pacific Paper Mill.com	УК
------------------------	----



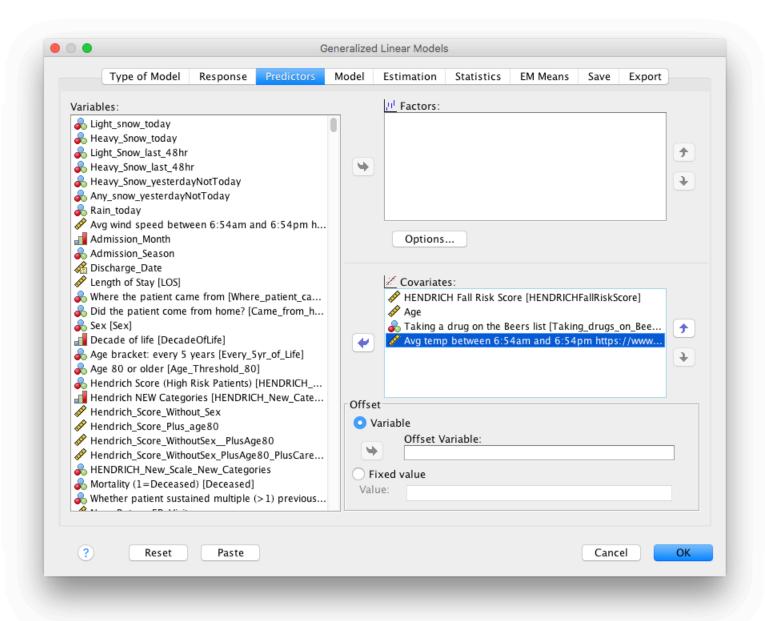
Put your predictors in the covariates box. There are conditions in which you might use "Factors" and "Offset" (but not now).

**Factors:** It's okay to just put them in the Covariates box.

**Offset:** Only use this if there's a problem with the specified time period, as in: some subjects have a longer period than others. For example: What predicts the number of tournaments an athlete wins? If one has competed for 2 years and another for 20 years, then the subjects have differing levels of exposure. Create a variable that reflects that and put it in the "Offset" box.

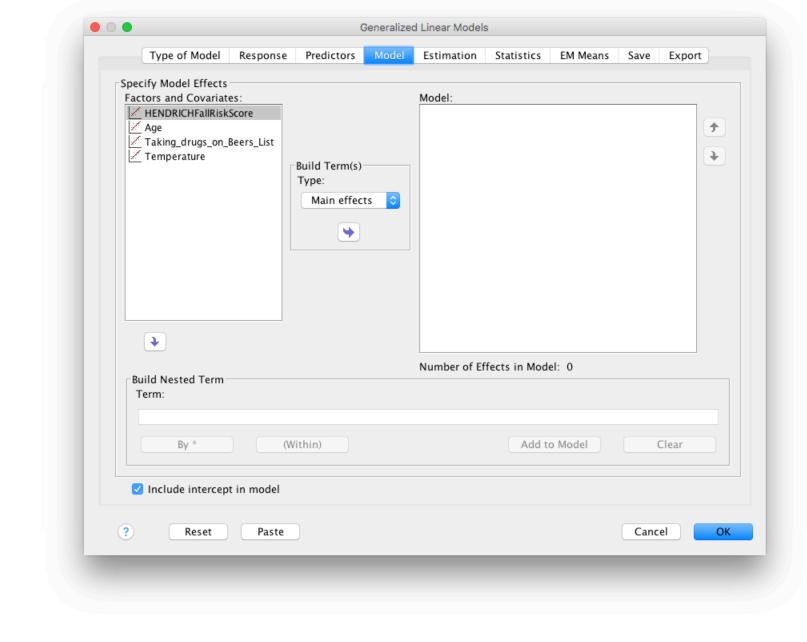


Pacific Pay	erMill.com
-------------	------------



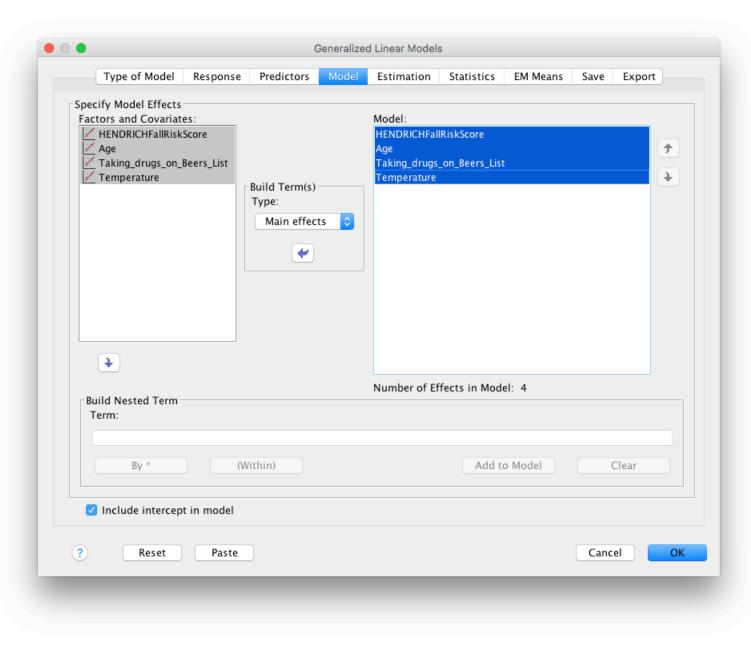
Model tab.

See all your predictors listed in the "Factors and Covariates" box?





Drag all of them over to the "Model" box.





Estimation tab.

Do nothing here. You don't even have to click on this one. If you do, feel free to browse for a minute, and then click on the next one (Statistics).



ific Paper Mill.com
---------------------

	Predictors	Model	Estimation	Statistics	EM Means	Save	Export
arameter Estimation			Covariance Ma	atrix			
Method: Hybrid		٥	O Model-ba		r		
Maximum Fisher Scoring Iter	ations:		O Robust es	timator			
Scale Parameter Method: Fixed v	alue	٢		alues for para rom a dataset			
Value:	1		Initial Va	lues			
terations							
Maximum Iterations: 100		Check	for separation	n of data poin	ts		
Maximum Step-Halving: 5		Sta	arting Iteration:	20			
Convergence Criteria							
At least one convergence criterion m	ust be specifi	ed with a	a minimum are	eater than 0.			
	Minimum:		Type:				
🗹 Change in parameter estimates	1E-006		Absolute	<b>\$</b>			
			Absolute	0			
Change in log-likelihood			Absolute	0			
<ul> <li>Change in log-likelihood</li> <li>Hessian convergence</li> </ul>							
_	3						

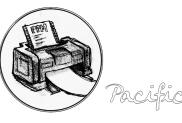
Statistics tab.

There's one option to select here.

Look at the bottom left row of checked boxes (beneath the word "Print").

The last checked box is for "Parameter estimates". Beneath that box, indented like a subheading, is an unchecked box with this option: "Include exponential parameter estimates".

Click it.



hcificPaperMill.com	N
---------------------	---

	Model	Estimation	Statistics	EM Means	Save	Export
odel Effects Analysis Type: Type III	•	Confiden	ice Interval Le	evel (%): 95		
Турст	<b></b>					
Chi-square Statistics			e Interval Ty	pe		
• Wald		💽 Wald				
C Likelihood ratio		O Profile	e likelihood			
		Toler	rance level:	0001		
Log-Likelihood Function: Full		L				
int						
Case processing summary	Co	ntrast coefficie	ent (L) matric	es		
Descriptive statistics	Ger	neral estimable	e functions			
Model information		ration history				
🗹 Goodness of fit statistics	Pri	nt Interval: 1				
🗹 Model summary statistics		ırange multipli				
🗹 Parameter estimates		ameter or nega illary paramete		.1		
Include exponential parameter estimates		, ,				
Covariance matrix for parameter estimates						
Correlation matrix for parameter estimates						

The exponential parameter estimates box is clicked.

This will give you your "incidence rate ratio" (IRR), which is the statistic you actually report.

Type of Model Response Predictors	s Model Estimation Statistics EM Means Save Export
lodel Effects	
Analysis Type: Type III	Confidence Interval Level (%): 95
Chi-square Statistics	Confidence Interval Type
<ul> <li>Wald</li> </ul>	Wald
🔿 Likelihood ratio	O Profile likelihood
	Tolerance level: .0001
Log-Likelihood Function: Full	
rint	
Case processing summary	Contrast coefficient (L) matrices
Descriptive statistics	General estimable functions
🗹 Model information	Iteration history
Goodness of fit statistics	Print Interval: 1
<b>— ••</b> • • •	Lagrange multiplier test of scale
Model summary statistics	parameter or negative binomial ancillary parameter
<ul> <li>Model summary statistics</li> <li>Parameter estimates</li> </ul>	ancinary parameter
	ancinary parameter
✓ Parameter estimates	
<ul> <li>Parameter estimates</li> <li>Include exponential parameter estimates</li> </ul>	
<ul> <li>Parameter estimates</li> <li>Include exponential parameter estimates</li> <li>Covariance matrix for parameter estimates</li> </ul>	
<ul> <li>Parameter estimates</li> <li>Include exponential parameter estimates</li> <li>Covariance matrix for parameter estimates</li> </ul>	
<ul> <li>Parameter estimates</li> <li>Include exponential parameter estimates</li> <li>Covariance matrix for parameter estimates</li> </ul>	
<ul> <li>Parameter estimates</li> <li>Include exponential parameter estimates</li> <li>Covariance matrix for parameter estimates</li> </ul>	

EM Means tab.

Do nothing here. Unless you placed one of your predictors in the "Factor" (rather than "Covariate") box.

Then you'd just drag that predictor over from the left to the right.

But, in general, do nothing here.



acific Paper Mill.com	ĸ
-----------------------	---

	Type of Model	Response	Predictors	Model	Estimation	Statistics	EM Means	Save	Export
	rs and Interactions				Display Me				
М		Term			Term	Co	ntrast	Refe	rence Category
- Scal				₩ By *					
0	Compute means fo	r response							
$\bigcirc$	Compute means fo	r linear predi	ctor						
Adjus	tment for Multiple	Comparisons							
	st significant diffe		٥		🗌 Display	y overall esti	mated mean		
?	Reset	Paste						Canc	el Oł

Save tab.

Do nothing.

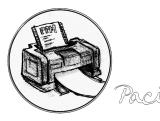


Pacific Paper Mill.com

	Type of Model	Response	Predictors	Model	Estimation	Statistics	EM Means	Save	Export
	Item to Save				Variable Nam	a or Root	Categories to S	-	
ave.	Predicted value of	mean of respo	nse		MeanPredicte		categories to 5	1	
	Lower bound of co			response	CIMeanPredic			-	
1					CIMeanPredic			-	
1					PredictedValu			-	
Ť	Predicted value of		r		XBPredicted	nc -		-	
Ť	Estimated standar			linear pre	XBStandardEr	ror		-	
Ť	Cook's distance	a ciror or prea	letted value of	inical press	CooksDistance			-	
Ť	Leverage value				Leverage	-		-	
Ť	Residual				Residual			-	
Ť	Pearson residual				PearsonResid	ual		-	
Ť	Standardized Pear	son residual			StdPearsonRe			-	
Ť	Deviance residual	Sonresidua			DevianceResi			-	
Ť	Standardized devi	ance residual			StdDevianceR			-	
1	Likelihood residua				LikelihoodRes			_	
0	ting Variable with Add suffix to nam Replace existing v Jf you provide yo	e of new varia ariable (applie	s to both def	ault and us	er-provided i		th evicting varia	alos in t	the active

Export tab.

Do nothing. Except for hit "OK"... which then runs the model you just created.



Pacific Paper Mill.com	5
------------------------	---

				Seneralize	d Linear Model	5			
	Type of Model	Response	Predictors	Model	Estimation	Statistics	EM Means	Save	Export
Ex	port model as dat	ta							
	nation								
	ataset								
Na	ime:								
O D	ata file								
	Browse								
xpor	rt as Data								
• Pa	arameter estimate	es and covaria	ance matrix						
O Pa	arameter estimate	es and correla	ation matrix						
• Pa	rt as XML arameter estimate arameter estimate		ance matrix						

The "Output" screen will appear.

# And gibberish that looks like this will appear at the top of it.

Ignore it (but do notice that the "Probability Distribution" is listed as Poisson).

But some of the stuff that follows is important.

\* Generalized Linear Models.

GENLIN Num\_of\_Previous\_Falls WITH HENDRICHFallRiskScore Age Taking\_drugs\_on\_Beers\_List Temperature /MODEL HENDRICHFallRiskScore Age Taking\_drugs\_on\_Beers\_List Temperature INTERCEPT=YES DISTRIBUTION=POISSON LINK=LOG

/CRITERIA METHOD=FISHER(1) SCALE=1 COVB=MODEL MAXITERATIONS=100 MAXSTEPHALVING=5
PCONVERGE=1E-006(ABSOLUTE) SINGULAR=1E-012 ANALYSISTYPE=3(WALD) CILEVEL=95 CITYPE=WALD
LIKELIH00D=FULL
/MISSING CLASSMISSING=EXCLUDE
(DDINT ODE DESCRIPTIVES MODELINES ET CUMMARY COLUTION (EVENING ATTACE))

/PRINT CPS DESCRIPTIVES MODELINFO FIT SUMMARY SOLUTION (EXPONENTIATED).

#### **Generalized Linear Models**

#### **Model Information**

Dependent Variable	Num_of_Previous_Falls
Probability Distribution	Poisson
Link Function	Log

Case	Processing
Su	immary

	· · · · · · · · · · · · · · · · · · ·	
	Ν	Percent
Included	593	96.4%
Excluded	22	3.6%
Total	615	100.0%





This is just means and standard deviations. Good to know, but you should already know all of this from running your descriptives (whole sample) and t-tests (subsamples).

		Ν	Minimum	Maximum	Mean	Std. Deviation
Dependent Variable	Num_of_Previous_Falls	593	0	11	1.91	1.313
Covariate	HENDRICH Fall Risk Score	593	0	14	2.46	2.715
	Age	593	65	101	79.95	9.082
	Taking a drug on the Beers list	593	0	1	.49	.500
	Avg temp between 6: 54am and 6:54pm https://www. timeanddate. com/weather/usa/india napolis/historic? month=1&year=2015	593	3.0	84.4	49.799	20.2321

#### **Continuous Variable Information**

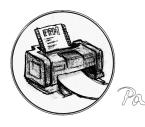


# Goodness of Fit box.

Goodness of fit means the values you observe in your sample match the values you'd expect to find in the population.

You want the Deviance row (last column: deviance/degrees of freedom) to be around 1. If it's too big, switch to a negative binomial regression.

Also, you want the AIC and BIC to be low. These values appear huge, but they're not. "Huge" is relative. This is okay.



Goodness of Fit<sup>a</sup>

	Value	df	Value/df
Deviance	303.484	588	.516
Scaled Deviance	303.484	588	
Pearson Chi-Square	344.452	588	.586
Scaled Pearson Chi- Square	344.452	588	
Log Likelihood <sup>b</sup>	-869.642		
Akaike's Information Criterion (AIC)	1749.283		
Finite Sample Corrected AIC (AICC)	1749.385		
Bayesian Information Criterion (BIC)	1771.209		
Consistent AIC (CAIC)	1776.209		

Dependent Variable: Num\_of\_Previous\_Falls Model: (Intercept), HENDRICH Fall Risk Score, Age, Taking a drug on the Beers list, Avg temp between 6:54am and 6: 54pm https://www.timeanddate. com/weather/usa/indianapolis/historic? month=1&year=2015

- a. Information criteria are in smaller-is-better form.
- b. The full log likelihood function is displayed and used in computing information criteria.

#### Omnibus Test box.

This is another "Goodness of fit" statistic.

Does our model (with all of our predictors in it) predict our outcome better than a model with no predictors in it (a "null model" or "intercept-only model")?

The omnibus test just says whether this model (with all of our predictors in it) is better than the null model.

In this case (p<0.001), it's way better than the null model. So we have an improvement in "fitness". We have "goodness of fit". It's significant. That's all you care about.

#### Omnibus Test<sup>a</sup>

Likelihood Ratio Chi– Square	df	Sig.
135.405	4	.000
Dependent Var Num_of_Previou Model: (Intercep Risk Score, Age the Beers list, A 54am and 6:54 timeanddate. com/weather/u ric?month=1&ye	us_Falls pt), HENDRI , Taking a c vg temp be 1pm https:/ usa/indianap	drug on tween 6: /www.
a. Compares against the model.	the fitted r e intercept–	



#### **Parameter Estimates**

The B on the left. This is the unstandardized regression coefficient. It represents the predicted change in expected *log counts* of the dependent variable (number of falls).

			95% Wald Conf	idence Interval	Нурс	othesis Test			95% Wald Confi for Ex	
Parameter	В	Std. Error	Lower	Upper	Wald Chi– Square	df	Sig.	Exp(B)	Lower	Upper
(Intercept)	136	.2763	677	.406	.242	1	.623	.873	.508	1.500
HENDRICH Fall Risk Score	.097	.0099	.077	.116	94.798	1	.000	1.102	1.080	1.123
Age	.008	.0034	.002	.015	6.158	1	.013	1.008	1.002	1.015
Taking a drug on the Beers list	.115	.0601	003	.233	3.660	1	.056	1.122	.997	1.262
Avg temp between 6: 54am and 6:54pm https://www. timeanddate. com/weather/usa/india napolis/historic? month=1&year=2015	005	.0015	008	002	10.257	1	.001	.995	.992	.998
(Scale)	1 <sup>a</sup>									

Dependent Variable: Num\_of\_Previous\_Falls

Model: (Intercept), HENDRICH Fall Risk Score, Age, Taking a drug on the Beers list, Avg temp between 6:54am and 6:54pm https://www.timeanddate.com/weather/usa/indianapolis/historic?month=1&year=2015

a. Fixed at the displayed value.

For each one unit increase in the predictor variable (e.g., Hendrich score, age, etc.), you get the B value of predicted change in the expected log count of your dependent variable. That's sort of gibberish. Just think of it this way: if the number is positive, then the relationship is positive. Predictor goes up, count goes up. Predictor goes down, count goes down. A negative value is the inverse of that.



#### **Parameter Estimates**

The *real* finding is Exp(B), i.e., exponentiation of beta. But you don't report it as Exp(B); report it as IRR, i.e., Incidence Rate Ratio.

Just as the Exp(B) in a logistic regression gets reported as the odds ratio (OR), this is reported as the IRR.

			95% Wald Conf	idence Interval	Hypothesis Test				95% Wald Confi for Ex	
Parameter	В	Std. Error	Lower	Upper	Wald Chi- Square	df	Sig.	Exp(B)	Lower	Upper
(Intercept)	136	.2763	677	.406	.242	1	.623	.873	.508	1.500
HENDRICH Fall Risk Score	.097	.0099	.077	.116	94.798	1	.000	1.102	1.080	1.123
Age	.008	.0034	.002	.015	6.158	1	.013	1.008	1.002	1.015
Taking a drug on the Beers list	.115	.0601	003	.233	3.660	1	.056	1.122	.997	1.262
Avg temp between 6: 54am and 6:54pm https://www. timeanddate. com/weather/usa/india napolis/historic? month=1&year=2015	005	.0015	008	002	10.257	1	.001	.995	.992	.998
(Scale)	1 <sup>a</sup>									

Dependent Variable: Num\_of\_Previous\_Falls

Model: (Intercept), HENDRICH Fall Risk Score, Age, Taking a drug on the Beers list, Avg temp between 6:54am and 6:54pm https://www.timeanddate. com/weather/usa/indianapolis/historic?month=1&year=2015

a. Fixed at the displayed value.

Take your dependent variable (in this case the number of hospital admissions owing to falls) and multiply that number by the IRR. If the IRR is less than 1, that would be a reduction in the frequency of occurrences. If it's positive, that would increase the frequency of occurrences.





#### **Parameter Estimates**

What you actually report depends on what variable you care about.

In this list, there are <u>four</u> <u>predictors</u>. These \_\_\_\_\_

				95% Wald Conf	idence Interval Hypothesis Test			95% Wald Confi for Ex			
	Parameter	В	Std. Error	Lower	Upper	Wald Chi- Square	df	Sig.	Exp(B)	Lower	Upper
	(Intercept)	136	.2763	677	.406	.242	1	.623	.873	.508	1.500
	HENDRICH Fall Risk Score	.097	.0099	.077	.116	94.798	1	.000	1.102	1.080	1.123
1	Age	.008	.0034	.002	.015	6.158	1	.013	1.008	1.002	1.015
3	Taking a drug on the Beers list	.115	.0601	003	.233	3.660	1	.056	1.122	.997	1.262
L	Avg temp between 6: 54am and 6:54pm https://www. timeanddate. com/weather/usa/india napolis/historic? month=1&year=2015	005	.0015	008	002	10.257	1	.001	.995	.992	.998
	(Scale)	1 <sup>a</sup>									

Dependent Variable: Num\_of\_Previous\_Falls

Model: (Intercept), HENDRICH Fall Risk Score, Age, Taking a drug on the Beers list, Avg temp between 6:54am and 6:54pm https://www.timeanddate. com/weather/usa/indianapolis/historic?month=1&year=2015

a. Fixed at the displayed value.

We want to know the effect of each of those predictors (in isolation of the other three) on the dependent variable's frequency of occurrences (number of falls).



Parameter Estimates

95% Wald Confidence Interval

Ignore these six columns:

			95% Wald Confid	lence Interval	Нуро	thesis Test			for Ex	(p(B)
Parameter	в	Std. Error	Lower	Upper	Wald Chi– Square	df	Sig.	Exp(B)	Lower	Upper
(Intercept)	136	.2763	677	.406	.247	1	.623	.873	.508	1.500
HENDRICH Fall Risk Score	.097	.0099	.07	.116	94.78		.000	1.102	1.080	1.123
Age	.008	0034	.002	.015	6,158		.013	1.008	1.002	1.015
Taking a drug on the Beers list	<b>V</b> 15	.0601	-003	233	<b>A</b> .660		.056	1.122	.997	1.262
Avg temp between 6: 54am and 6:54pm https://www. timeanddate. com/weather/usa/india napolis/historic? month=1&year=2015	4.005	.0015	008	.002	10.257		.001	.995	.992	.998
(Scale)	$1^{h}$	1	$\mathbf{V} = \mathbf{N}$	/ \	/ \	$I \land$				
Dependent Variable: Num	of Previou	s Falls	7 7	1						

Dependent Variable: Num\_of\_Previous\_Falls Model: (Intercept), HENDRICH Fall Risk Score, Age, Taking a drug on the Beers list, Avg temp between 6:54am and 6:54pm https://www.timeanddate. com/weather/usa/indianapolis/historic?month=1&year=2015

a. Fixed at the displayed value.

Those columns offer some information, but nothing you will report.



**Parameter Estimates** 

95% Wald Confidence Interval

The last four columns provide
the data you will report:

			95% Wald Confidence Interval		Hypothesis Test			for Ex	for Exp(B)	
Parameter	В	Std. Error	Lower	Upper	Wald Chi– Square	df	Sig.	Exp(B)	Lower	Upper
(Intercept)	136	.2763	677	.406	.242	1	.623	.873	.508	1.500
HENDRICH Fall Risk Score	.097	.0099	.077	.116	94.798	1	.000	1.102	1.080	1.123
Age	.008	.0034	.002	.015	6.158	1	.013	1.008	1.002	1.015
Taking a drug on the Beers list	.115	.0601	003	.233	3.660	1	.056	1.122	.997	1.262
Avg temp between 6: 54am and 6:54pm https://www. timeanddate. com/weather/usa/india napolis/historic? month=1&year=2015	005	.0015	008	002	10.257	1	.001	.995	.992	.998
(Scale)	1 <sup>a</sup>									

Dependent Variable: Num\_of\_Previous\_Falls

Model: (Intercept), HENDRICH Fall Risk Score, Age, Taking a drug on the Beers list, Avg temp between 6:54am and 6:54pm https://www.timeanddate. com/weather/usa/indianapolis/historic?month=1&year=2015

a. Fixed at the displayed value.

Each of the four predictor variables has a different effect on the dependent variable. For example: Age. For each additional year of age (holding the other three predictors constant), we expect patients to experience a 0.8% increase in the number of falls. The "Sig." column provides a p-value, which is our confidence that this prediction applies to the larger population, and not just our sample. In this case, if the null hypothesis were true (i.e., age is totally unrelated to frequency of falling), there is a 1.3% chance we would have observed an effect on the IRR this large or larger.

## The last two columns provide the 95% confidence interval for the IRR.

			95% Wald Confi	dence Interval	Нуро	thesis Test			95% Wald Confi for Ex	
Parameter	В	Std. Error	Lower	Upper	Wald Chi– Square	df	Sig.	Exp(B)	Lower	Upper
(Intercept)	136	.2763	677	.406	.242	1	.623	.873	.508	1.500
HENDRICH Fall Risk Score	.097	.0099	.077	.116	94.798	1	.000	1.102	1.080	1.123
Age	.008	.0034	.002	.015	6.158	1	.013	1.008	1.002	1.015
Taking a drug on the Beers list	.115	.0601	003	.233	3.660	1	.056	1.122	.997	1.262
Avg temp between 6: 54am and 6:54pm https://www. timeanddate. com/weather/usa/india napolis/historic? month=1&year=2015	005	.0015	008	002	10.257	1	.001	.995	.992	.998
(Scale)	1 <sup>a</sup>									

Parameter Estimates

Dependent Variable: Num\_of\_Previous\_Falls

Model: (Intercept), HENDRICH Fall Risk Score, Age, Taking a drug on the Beers list, Avg temp between 6:54am and 6:54pm https://www.timeanddate. com/weather/usa/indianapolis/historic?month=1&year=2015

a. Fixed at the displayed value.

The 95% confidence interval means we're 95% sure that the true IRR in the actual population falls between 0.2% and 1.5%. So for each additional year of age, holding constant the other three predictors, we're pretty sure (95% sure to be exact) that patients in the larger population (not merely this sample) will experience between 0.2% more and 1.5% more falls.



Interpret the other predictors the same way.

			95% Wald Conf	idence Interval	Нурс	othesis Test			95% Wald Confi for Ex	
Parameter	В	Std. Error	Lower	Upper	Wald Chi– Square	df	Sig.	Exp(B)	Lower	Upper
(Intercept)	136	.2763	677	.406	.242	1	.623	.873	.508	1.500
HENDRICH Fall Risk Score	.097	.0099	.077	.116	94.798	1	.000	1.102	1.080	1.123
Age	.008	.0034	.002	.015	6.158	1	.013	1.008	1.002	1.015
Taking a drug on the Beers list	.115	.0601	003	.233	3.660	1	.056	1.122	.997	1.262
Avg temp between 6: 54am and 6:54pm https://www. timeanddate. com/weather/usa/india napolis/historic? month=1&year=2015	005	.0015	008	002	10.257	1	.001	.995	.992	.998
(Scale)	1 <sup>a</sup>									

**Parameter Estimates** 

Dependent Variable: Num\_of\_Previous\_Falls

Model: (Intercept), HENDRICH Fall Risk Score, Age, Taking a drug on the Beers list, Avg temp between 6:54am and 6:54pm https://www.timeanddate. com/weather/usa/indianapolis/historic?month=1&year=2015

a. Fixed at the displayed value.

If what you care about is the effect of drugs, then look at the drug row.

Holding the other three predictors constant, if a patient is taking a drug on that list (this is a list of drugs that might be inappropriate for older adults), that predicts an IRR% increase in the number of falls experienced over the specified period. Notice the p-value isn't *quite* "significant" (p<0.05) though. Rather, it's "trending" (p<0.08).



#### **Parameter Estimates**

95% Wald Confidence Interval

If the IRR is less than 1...

			95% Wald Conf	idence Interval	Нурс	thesis Test			for E	xp(B)
Parameter	В	Std. Error	Lower	Upper	Wald Chi– Square	df	Sig.	Exp(B)	Lower	Upper
(Intercept)	136	.2763	677	.406	.242	1	.623	.873	.508	1.500
HENDRICH Fall Risk Score	.097	.0099	.077	.116	94.798	1	.000	1.102	1.080	1.123
Age	.008	.0034	.002	.015	6.158	1	.013	1.008	1.002	1.015
Taking a drug on the Beers list	.115	.0601	003	.233	3.660	1	.056	1.122	.997	1.262
Avg temp between 6: 54am and 6:54pm https://www. timeanddate. com/weather/usa/india napolis/historic? month=1&year=2015	005	.0015	008	002	10.257	1	.001	.995	.992	.998
(Scale)	1 <sup>a</sup>									

Dependent Variable: Num\_of\_Previous\_Falls

Model: (Intercept), HENDRICH Fall Risk Score, Age, Taking a drug on the Beers list, Avg temp between 6:54am and 6:54pm https://www.timeanddate. com/weather/usa/indianapolis/historic?month=1&year=2015

a. Fixed at the displayed value.

Notice the IRR of ambient temperature is less than 1. That means the frequency that the dependent variable occurs goes *down*. If the IRR were .500, that would be a 50% reduction in the number of occurrences. If the IRR were .800, that would be a 20% reduction. If it were .990, that's a 1% reduction. You can't have an IRR less than 0. Anything between 0 and .999 means you're reducing the incidence. Anything over 1 means you're increasing the incidence.



Let's look at another set of outcomes.

We'll use the same database. And the same dependent variable. But use a different set of predictors.

So the regression model will be familiar, but entirely new.

So the following outputs are from an entirely new Poisson regression, run in the exact same way, just containing different predictors.



Here are the means and standard deviations for the dependent variable and the seven predictors used in this model:

		Ν	Minimum	Maximum	Mean	Std. Deviation
Dependent Variable	Num_of_Previous_Falls	593	0	11	1.91	1.313
Covariate	Admission_Month	593	1	12	6.50	3.919
	Age	593	65	101	79.95	9.082
	SelfReport_PoorBalance_ NotAsked	593	0	1	.03	.167
	LightheadednessDuring Appointment	593	0	1	.13	.342
	Cognitive Struggles (Coded: 0/1)	593	0	1	.25	.431
	Dementia	593	0	1	.12	.325
	DIABETES: Rapid-acting insulin (glulisine, aspart, humalog, novolog) (no=0; yes=1)	593	0	1	.04	.205

#### **Continuous Variable Information**



Here's the Goodness of Fit table.

Again, you're looking at that top right number (ideally it's around 1 as opposed to 15 or 20 or 150) and at AIC and BIC, making sure they're not huge.



Pacific Paper Mill.com	S
------------------------	---

#### Goodness of Fit<sup>a</sup>

	Value	df	Value/df
Deviance	344.043	585	.588
Scaled Deviance	344.043	585	
Pearson Chi-Square	392.561	585	.671
Scaled Pearson Chi- Square	392.561	585	
Log Likelihood <sup>b</sup>	-889.921		
Akaike's Information Criterion (AIC)	1795.841		
Finite Sample Corrected AIC (AICC)	1796.088		
Bayesian Information Criterion (BIC)	1830.923		
Consistent AIC (CAIC)	1838.923		

Dependent Variable: Num\_of\_Previous\_Falls Model: (Intercept), Admission\_Month, Age, Dementia, SelfReport\_PoorBalance\_NotAsked, LightheadednessDuringAppointment, Cognitive Struggles (Coded: 0/1), DIABETES: Rapid-acting insulin (glulisine, aspart, humalog, novolog) (no=0; yes=1)

- a. Information criteria are in smaller-is-better form.
- b. The full log likelihood function is displayed and used in computing information criteria.

Here's the Omnibus Test.

Because p<0.05 (its less than 0.001 here), this collection of predictors is better than the null model. There is improvement in "fitness".

#### Omnibus Test<sup>a</sup>

Likelihood Ratio Chi– Square	df	Sig.
94.847	7	.000

Dependent Variable: Num\_of\_Previous\_Falls Model: (Intercept), Admission\_Month, Age, Dementia, SelfReport\_PoorBalance\_NotAsked, LightheadednessDuringAppointment , Cognitive Struggles (Coded: 0/1), DIABETES: Rapid-acting insulin (glulisine, aspart, humalog, novolog) (no=0; yes=1)

a. Compares the fitted model against the intercept-only model.



# And here are the Parameter Estimates.

This is the table of values you report.

Remember: You only need to know the last four columns in this table:

"Sig." (i.e., p-value) "Exp(B)" (i.e., IRR) Bottom of the 95% CI Top of the 95% CI

			95% Wald Conf	idence Interval	Нуро	thesis Test			95% Wald Confid for Ex	
Parameter	В	Std. Error	Lower	Upper	Wald Chi- Square	df	Sig.	Exp(B)	Lower	Upper
(Intercept)	.051	.2823	503	.604	.032	1	.858	1.052	.605	1.829
Admission_Month	025	.0079	041	010	10.329	1	.001	.975	.960	.990
Age	.007	.0035	.000	.014	4.378	1	.036	1.007	1.000	1.014
Dementia	.232	.1049	.027	.438	4.912	1	.027	1.262	1.027	1.549
SelfReport_PoorBalance_ NotAsked	.286	.1415	.009	.564	4.095	1	.043	1.332	1.009	1.757
LightheadednessDuring Appointment	.198	.0821	.037	.359	5.820	1	.016	1.219	1.038	1.432
Cognitive Struggles (Coded: 0/1)	.237	.0874	.066	.409	7.372	1	.007	1.268	1.068	1.505
DIABETES: Rapid-acting insulin (glulisine, aspart, humalog, novolog) (no=0; yes=1)	.317	.1265	.069	.565	6.259	1	.012	1.372	1.071	1.759
(Scale)	1 <sup>a</sup>									

**Parameter Estimates** 

Dependent Variable: Num\_of\_Previous\_Falls

Model: (Intercept), Admission\_Month, Age, Dementia, SelfReport\_PoorBalance\_NotAsked, LightheadednessDuringAppointment, Cognitive Struggles (Coded: 0/1), DIABETES: Rapid-acting insulin (glulisine, aspart, humalog, novolog) (no=0; yes=1)

a. Fixed at the displayed value.

The first column (B) is your unstandardized regression coefficient. You can ignore it and nobody will ask you about it, but it's is the predicted change in your expected logged count of previous hospital admissions. Every one-unit increase in your predictor variable (age, whether you have dementia, etc.) predicts the logged count of your dependent (number of hospital visits) to change by that much. It's too complicated of a statistic. No point in reporting it. The numbers are effectively meaningless to everybody who doesn't immediately start playing around with a calculator. Report the IRR, the significance, and the confidence interval. That's all.



## And here are the Parameter Estimates. This is the table of values you report.

Interpretation of a single row (you can interpret every other row in the same way): Dementia. If dementia is the variable of interest, you report: Holding the other five predictors (age, poor balance, etc.) constant, having dementia predicts a 26.2% increase in the number of previous hospitalizations for falls (p=0.027; 95% CI of IRR: 1.027 to 1.549).

			95% Wald Conf	idence Interval	Нуро	thesis Test			95% Wald Confi for Ex	
Parameter	В	Std. Error	Lower	Upper	Wald Chi- Square	df	Sig.	Exp(B)	Lower	Upper
(Intercept)	.051	.2823	503	.604	.032	1	.858	1.052	.605	1.829
Admission_Month	025	.0079	041	010	10.329	1	.001	.975	.960	.990
Age	.007	.0035	.000	.014	4.378	1	.036	1.007	1.000	1.014
Dementia	.232	.1049	.027	.438	4.912	1	.027	1.262	1.027	1.549
SelfReport_PoorBalance_ NotAsked	.286	.1415	.009	.564	4.095	1	.043	1.332	1.009	1.757
LightheadednessDuring Appointment	.198	.0821	.037	.359	5.820	1	.016	1.219	1.038	1.432
Cognitive Struggles (Coded: 0/1)	.237	.0874	.066	.409	7.372	1	.007	1.268	1.068	1.505
DIABETES: Rapid-acting insulin (glulisine, aspart, humalog, novolog) (no=0; yes=1)	.317	.1265	.069	.565	6.259	1	.012	1.372	1.071	1.759
(Scale)	1 <sup>a</sup>									

#### Parameter Estimates

Dependent Variable: Num\_of\_Previous\_Falls

Model: (Intercept), Admission\_Month, Age, Dementia, SelfReport\_PoorBalance\_NotAsked, LightheadednessDuringAppointment, Cognitive Struggles (Coded: 0/1), DIABETES: Rapid-acting insulin (glulisine, aspart, humalog, novolog) (no=0; yes=1)

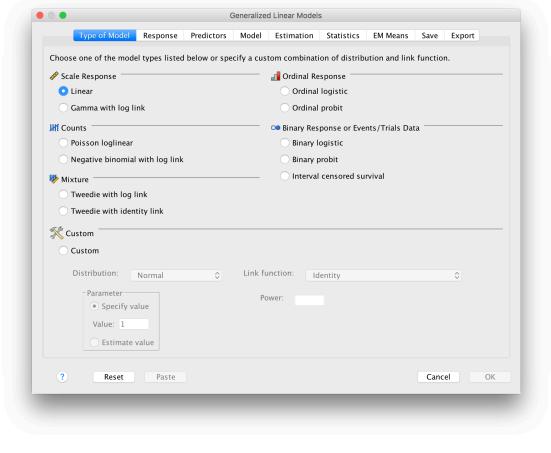
a. Fixed at the displayed value.

Remember: IRR = Incidence Response Rate. Multiply the number of expected counts (i.e., number of hockey goals in a game or number of workouts in a week or, in this case, the number of hospital admissions owing to fall-related injuries) by the IRR, which is written as Exp(B).



# What if there is overdispersion and a negative binomial regression (as opposed to Poisson) needs to be run?

Open up a Generalized Linear Model and go to the Type of Model tab. It'll look like this:





What if there is overdispersion and a negative binomial regression (as opposed to Poisson) needs to be run?

Select "Custom" in the Custom section.

Select "Negative Binomial" in the Distribution section.

And select "Log" in the Link function section.

The default setting for the dispersion parameter is 1. I'll leave it at 1; you can set it to 0 if that creates a more appropriate model.



Pacific Paper Mill.com

🔗 Scale Response	Grdinal Response
	Ordinal logistic
Gamma with log link	<ul> <li>Ordinal probit</li> </ul>
If Counts	• Binary Response or Events/Trials Data
O Poisson loglinear	O Binary logistic
Negative binomial with log link	O Binary probit
Wixture	Interval censored survival
Tweedie with log link	
Tweedie with identity link	
<ul> <li>Custom</li> <li>Distribution: Negative binomial </li> <li>Parameter</li> <li>Specify value</li> <li>Value: 1</li> <li>Estimate value</li> </ul>	Link function: Log

# Then fill out all the other tabs *exactly* as you did before and hit OK.

• We're using the same predictors as we did in our first Poisson regression



These are the four predictors. In regression models, "independents" are generally referred to as "predictors" owing to the fact that they might not be independent of each other. Predictor is a better term.

The only thing we're changing in this model is making it negative binomial as opposed to Poisson. The dependent variable and predictors are identical.



# Outputs:

• Your "Model Information" box (first box you encounter) will reflect a negative binomial regression

# **Generalized Linear Models**

#### **Model Information**

Dependent Variable	Num_of_Previous_Falls
Probability Distribution	Negative binomial (1)
Link Function	Log



# Outputs:

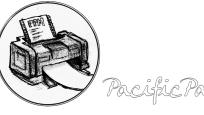
In the "Goodness of Fit" box, you'll have new fitness values: ٠

(Poisson) model:

Original

In this case, the Poisson is better than the negative binomial model.

Look at AIC, BIC, and ratio of deviance to degrees of freedom. All of those are better (AIC and BIC are lower and deviance/df is closer to 1) in the Poisson. Each of those values would be different if you set the dispersion parameter to 0 instead of 1.



acific Paper Mill.com	L
-----------------------	---

#### Goodness of Fit<sup>a</sup>

	Value	df	Value/df
Deviance	303.484	588	.516
Scaled Deviance	303.484	588	
Pearson Chi-Square	344.452	588	.586
Scaled Pearson Chi- Square	344.452	588	
Log Likelihood <sup>b</sup>	-869.642		
Akaike's Information Criterion (AIC)	1749.283		
Finite Sample Corrected AIC (AICC)	1749.385		
Bayesian Information Criterion (BIC)	1771.209		
Consistent AIC (CAIC)	1776.209		

Dependent Variable: Num\_of\_Previous\_Falls Model: (Intercept), HENDRICH Fall Risk Score, Age, Taking a drug on the Beers list, Avg temp between 6:54am and 6: 54pm https://www.timeanddate. com/weather/usa/indianapolis/historic? month=1&vear=2015

a. Information criteria are in smaller-is-better form.

b. The full log likelihood function is displayed and used in computing information criteria.

#### Goodness of Fit<sup>a</sup>

		Value	df	Value/df				
	Deviance	95.694	588	.163				
q	Scaled Deviance	95.694	588					
DO	Pearson Chi-Square	112.410	588	.191				
1) n	Scaled Pearson Chi- Square	112.410	588					
13	Log Likelihood <sup>b</sup>	-1088.181						
om	Akaike's Information Criterion (AIC)	2186.362						
bin	Finite Sample Corrected AIC (AICC)	2186.464						
Ve	Bayesian Information Criterion (BIC)	2208.288						
ati	Consistent AIC (CAIC)	2213.288						
New (negative binomial) model	Dependent Variable: Num_of_Previous_Falls Model: (Intercept), HENDRICH Fall Risk Score, Age, Taking a drug on the Beers list, Avg temp between 6:54am and 6: 54pm https://www.timeanddate. com/weather/usa/indianapolis/historic?month=1&year=2015							
ંગ	a. Information criteria are in smaller-is-better form.							
$\mathbf{Z}$	b. The full log likelihood		played and	used in				

b. The full log likelihood function is displayed and used in computing information criteria.

# Outputs:

You'll have new Parameter Estimates ٠

The significance changes based on what kind of model you're running. The original (Poisson) model is better.

Again: Incidence Rate Ratio is the important statistic. Changes in IRR mean that for every one unit increase in predictor variable, you multiply the frequency of the event happening (count of your dependent variable) by that number. So a value of 1 means no changes in incidence rate (number of times the outcome happens). Value of less than 1 means incidence rate decreases. More than 1: incidence rate increases.



#### Original (Poisson) model:

				Parameter	Estimates					
			95% Wald Conf	idence Interval	Hypothesis Test				95% Wald Confi for E	
Parameter	В	Std. Error	Lower	Upper	Wald Chi- Square	df	Sig.	Exp(B)	Lower	Upper
(Intercept)	136	.2763	677	.406	.242	1	.623	.873	.508	1.500
HENDRICH Fall Risk Score	.097	.0099	.077	.116	94.798	1	.000	1.102	1.080	1.123
Age	.008	.0034	.002	.015	6.158	1	.013	1.008	1.002	1.015
Taking a drug on the Beers list	.115	.0601	003	.233	3.660	1	.056	1.122	.997	1.262
Avg temp between 6: 54am and 6:54pm https://www. timeanddate. com/weather/usa/india napolis/historic? month=1&year=2015	005	.0015	008	002	10.257	1	.001	.995	.992	.998
(Coole)	18									

enendent Variable: Num of Previous Fall

repeation of the second s

Fixed at the displayed value

#### New (negative binomial) model:

				Parameter	Estimates					
			95% Wald Conf	idence Interval	Hypothesis Test				95% Wald Confidence Interval for Exp(B)	
Parameter	в	Std. Error	Lower	Upper	Wald Chi- Square	df	Sig.	Exp(B)	Lower	Upper
(Intercept)	170	.4688	-1.089	.749	.132	1	.717	.844	.337	2.114
HENDRICH Fall Risk Score	.105	.0196	.066	.143	28.672	1	.000	1.110	1.069	1.154
Age	.008	.0058	003	.020	2.056	1	.152	1.008	.997	1.020
Taking a drug on the Beers list	.111	.1038	093	.314	1.137	1	.286	1.117	.911	1.369
Avg temp between 6: 54am and 6:54pm https://www. timeanddate. com/weather/usa/india napolis/historic? month=1&year=2015	004	.0026	009	.001	2.772	1	.096	.996	.991	1.001
(Scale)	1 <sup>a</sup>									
(Negative binomial)	1 <sup>a</sup>									
Demandent Variables Norm	of Description	- Celle								

Dependent Variable: Num\_ot\_Previous\_tains Model: (Intercept), HENDRICH Fall Risk Score, Age, Taking a drug on the Beers list, Avg temp between 6:54am and 6:54pm https://www.timeanddate com/weather/usa/indianapolis/historic?month=1&year=2015

Fixed at the displayed value

Decide which model type (Poisson or negative binomial) is most appropriate based on distribution of data (i.e., means vs. variances, ratio of deviance to degrees of freedom) and not by which model gives you more encouraging p-values.

