



What if Proc Power Can't Help? Calculating Power via Simulation, Example with Count Regression

Brandy is a statistician and database programmer at the University of Michigan Medical School. She has worked on a wide variety of research projects. She holds a bachelors degree in engineering, and masters degrees in mathematics and statistics. The American Public Health Association awarded Brandy a student research award for her statistics masters thesis on path analysis. She has also received “Advanced Analytics” and “Data for Good” awards from the Midwest SAS Users Group conference.

Outside of work, she enjoys playing the piano, composing music, singing in choir, yoga, tai chi, and martial arts aerobics.

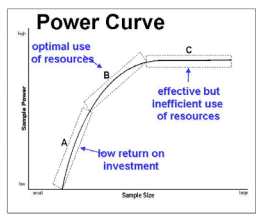


What if Proc Power Can't Help? Calculating Power via Simulation, Example with Count Regression

Brandy R. Sinco, MS, University of Michigan, Ann Arbor, MI

Joceline Vu, MD, University of Michigan, Ann Arbor, MI

Pasithorn Suwanabol, MD, University of Michigan, Ann Arbor, MI



Background for Power Request

- Two surgeons, Dr. Vu and Dr. Suwanabol, submitted a manuscript that compared post-operative opioid use between patients who had laparoscopic versus robotic colectomies.
- The outcome, number of pills consumed, was a count variable, and was analyzed with zero-inflated negative binomial model. Poisson and negative binomial regression models are two common models for count data. When the counts have more zeroes, than a Poisson or negative binomial distribution would have, zero-inflated models are a useful option.
- Reviewers for the article comparing opioid use after surgery questioned the amount of power for a Zero-Inflated Negative Binomial (ZINB) model.
- Brandy was asked to compute power via simulation to address the reviewers' concerns. Because neither SAS procedures nor Stata can compute power for ZINB models, Brandy turned to power via simulation.

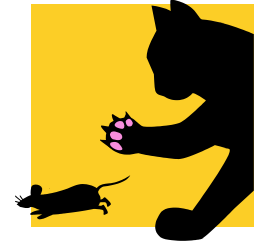


Road Map (Outline)

- 1)_ Definition of Statistical Power and Meaning of Power by Simulation. Useful tools in SAS.
- 2)_ Overview of Poisson, Negative Binomial, Zero-Inflated Models.
- 3)_ How to Use Proc GenMod to Estimate the Distribution Parameters for Count Regression.
- 4)_ Comparing Simulated Data to the Actual Data with Graphics and Descriptive Statistics.
- 5)_ Running the Outcomes Model on the Simulated Data and Computing the Power.
- 6)_ Bootstrapping to calculate power.
- 7)_ Conclusions.



Statistical Power (1 of 3)



- Power to detect a significant effect when it is present.
- Example 1: Mean temperature difference between two cities. Is it 0 (negligible) or is the mean > 5 ?
- Power depends on the minimum amount to detect, variation in the data, sample size.
- Detecting a temperature difference of 20 would require a smaller sample size than if we wanted to detect 5.
- Example 2: Detect presence of mice in a garage by chemical present on the bodies of mice.
- Cat responds to smaller amount of the chemical and has more power to detect the scent of a mouse than a human.



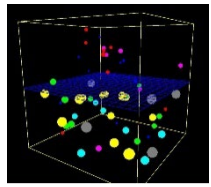
Statistical Power (2 of 3)

- μ = Parameter of interest, a mean, variance, rate, coefficient, or regression curve.
- Power = $\Pr(\text{Detecting difference in } \mu \mid \text{Difference in } \mu \text{ is present})$.
- Example: Post-operative opioid use between patients having laparoscopic versus robotic colectomies.
- H_0 : Null hypothesis, no difference, $\mu_1 = \mu_2$.
- No impact of surgical technique on post-operative opioid use.
- H_A : Alternative hypothesis, $\mu_1 \neq \mu_2$ or $\mu_1 > \mu_2$ or $\mu_1 < \mu_2$.
- Post-operative opioid use differs by surgical technique.



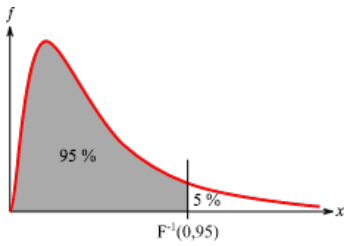
Statistical Power (3 of 3)

- Type 1 error, $\alpha = \Pr(H_A | H_0 \text{ is true})$. α also called “level of significance”.
- Type 1 error: Conclude that a difference is present, when difference is due to random variation.
- I.E., wrongfully conclude that opioid use differs by surgical technique, when there is really no difference.
- P-value = observed level of significance.
- Type 2 error, $\beta = \Pr(H_0 | H_A)$. Fail to detect a difference when a difference is present.
- **Power = $1 - \beta = \Pr(H_A | H_A)$. Probability of accurately detecting a difference.**



Calculating Power by Simulation

- Create a large number of datasets, 1000 or more with simulated distributions of the parameter of interest. Power = % of datasets with statistically significant results.
- Example: Suppose group 1 has a mean of 2, standard deviation of 2.5, and $N_1 = 10$; group 2 has a mean of 1, standard deviation of 2.25, and $N_2 = 20$.
- How much power do we have to detect a difference in means between groups 1 and 2?
- For this question, SAS Proc Power can help.
- Before diving into zero-inflated negative binomial regression, let's focus on this simpler example.



Simple Example: Formula, Proc Power, Simulation

- Formula for Power. $N = (\sigma_1^2 + \sigma_2^2)(Z_{\alpha/2} + Z_{\beta})^2 / d^2$; d = minimum detectable difference.
- $F = \sigma_1^2 / \sigma_2^2 = (2.5/2.25)^2 = 1.235$, $p = .332$; use pooled T-test.
- /* SAS code to find p-value. */
- Data _NULL_;
- p_value=1-probf(1.235, 9, 19);
- Put "F_05_9_19=" F_05_9_19 "P_Value=" P_Value;
- Run;

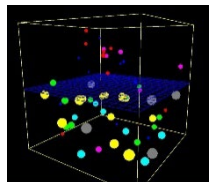
- **`/* Unequal Variances */`**
- `proc power;`
- `twosamplemeans test=diff_satt`
- `groupmeans = 2 | 1`
- `GROUPSTDDEVS = 2.5 | 2.25`
- `groupns = (10 20)`
- `power = .;`
- `run;`

- **`/* Equal Variances */`**
- `Test = diff` instead of `diff_satt`.
- `StdDev = 2.333` instead of `GroupStdDevs`.



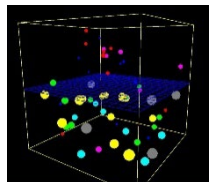
Proc Power Output for Unequal Variances

Distribution	Normal
Method	Exact
Group 1 Mean	2
Group 2 Mean	1
Group 1 Standard Deviation	2.5
Group 2 Standard Deviation	2.25
Group 1 Sample Size	10
Group 2 Sample Size	20
Number of Sides	2
Null Difference	0
Nominal Alpha	0.05
Computed Power	
Actual Alpha	Power
0.0505	0.172



Interpretation of Proc Power Output

- Assuming unequal variances, power = .172. If a simulated dataset is created for 1,000 repetitions, the means should be significantly different 17.2% of the time.
- Assuming equal variances, power = .188. So, simulated dataset should have differing means ~18.8% of the time.
- If the dataset has a mixture of equal and unequal variances, the power to detect a significant difference in means should be ~ 16% to 20%.



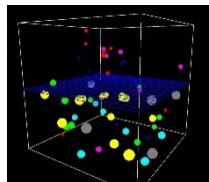
SAS Program for Simulation



- Data MeansTest1000;
- N1=10; N2=20;

- do Rep=1 to 1000; Call streaminit(Rep);
- do J1=1 to N1;
- X = Rand('Normal', 2, 2.5); Group=1; output;
- end; /* J1 loop */

- do J2=1 to N2;
- X = Rand('Normal', 1, 2.25); Group=2;output;
- end; /* J2 loop */
- end; /* Rep Loop */
- Run;



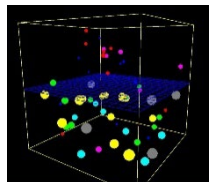
T-Tests on Simulated Data



- Proc TTest Data=MeansTest1000;
- class Group;
- var X;
- **by Rep;**
- ods output Equality=EqualVar1000 TTests=TTTests1000;
- run;

- Most SAS statistical procedures can run with a “By” statement.

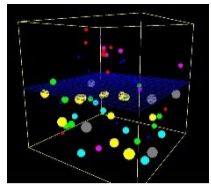
- Run the procedure “by Rep” to calculate power by simulation.



View Proc T-test Output



Obs	Rep	Variable	Method	Variances	tValue	DF	Probt
1	1	X	Pooled	Equal	2.80	28	0.0091
2	1	X	Satterthwaite	Unequal	3.08	23.212	0.0052
3	2	X	Pooled	Equal	0.79	28	0.4353
4	2	X	Satterthwaite	Unequal	0.87	23.307	0.3927
5	3	X	Pooled	Equal	0.26	28	0.7944
6	3	X	Satterthwaite	Unequal	0.27	18.671	0.7930
7	4	X	Pooled	Equal	2.07	28	0.0483
8	4	X	Satterthwaite	Unequal	1.86	14.091	0.0832
9	5	X	Pooled	Equal	0.82	28	0.4171
10	5	X	Satterthwaite	Unequal	0.88	21.3	0.3912



Calculate Power of T-Tests



- Data TTests1000Sig;
- set TTests1000;
- Sig=(ProbT<.05); run;
-
- Title 'Power if variance pooled'; /* Power = 20.2%, compared to 18.8% Proc Power */
- proc freq data=TTests1000Sig;
- tables Sig; where Method='Pooled';run;
-
- Title 'Power if Satterthwaite';/* Power = 17.5%, compared to 17.2% Proc Power */
- proc freq data=TTests1000Sig;
- tables Sig;where Method='Satterthwaite'; run;



Poisson and Negative Binomial Distributions for Counts

- **Poisson.** Let Y = number of occurrences of an event, such as the number of a driver's traffic accidents. In a Poisson distribution, the number of occurrences increases with the area of measure. For example, $\text{Pr}(\text{accident})$ increases with distance of commute to work.
- $f(Y=y; \mu) = e^{-\mu}\mu^y/y!$, μ = mean = variance, $y = 0, 1, 2, \dots, \infty$.
- Poisson Regression $\ln(y) = X\beta$.

- **Negative Binomial.** Negative binomial regression is an alternative to Poisson regression when variance > mean, also called over-dispersion.
- Negative Binomial Regression uses the log link, same as Poisson regression.
- Scale parameter = Φ , $\Phi > 0$.
- Mean = μ .
- Variance = $\mu + \Phi\mu^2$.



Zero Inflated Distributions

- Models for distributions with excess zeroes, where y could be modeled as Poisson or Negative Binomial, if the distribution didn't contain excessive 0's.
- (Lambert, 1992 and Greene, 1994).
- **Zero-Inflated Poisson.** Mixture of a Bernoulli distribution and a Poisson distribution. Model for distribution with excess zeroes.
- When $y=0$, $p \cdot \text{Logistic} + (1-p) \cdot \text{Poisson}$.
- When $y>0$, $(1-p) \cdot \text{Poisson}$.
- **Zero-Inflated Negative Binomial.** Mixture of a Bernoulli distribution and a Negative Binomial distribution.

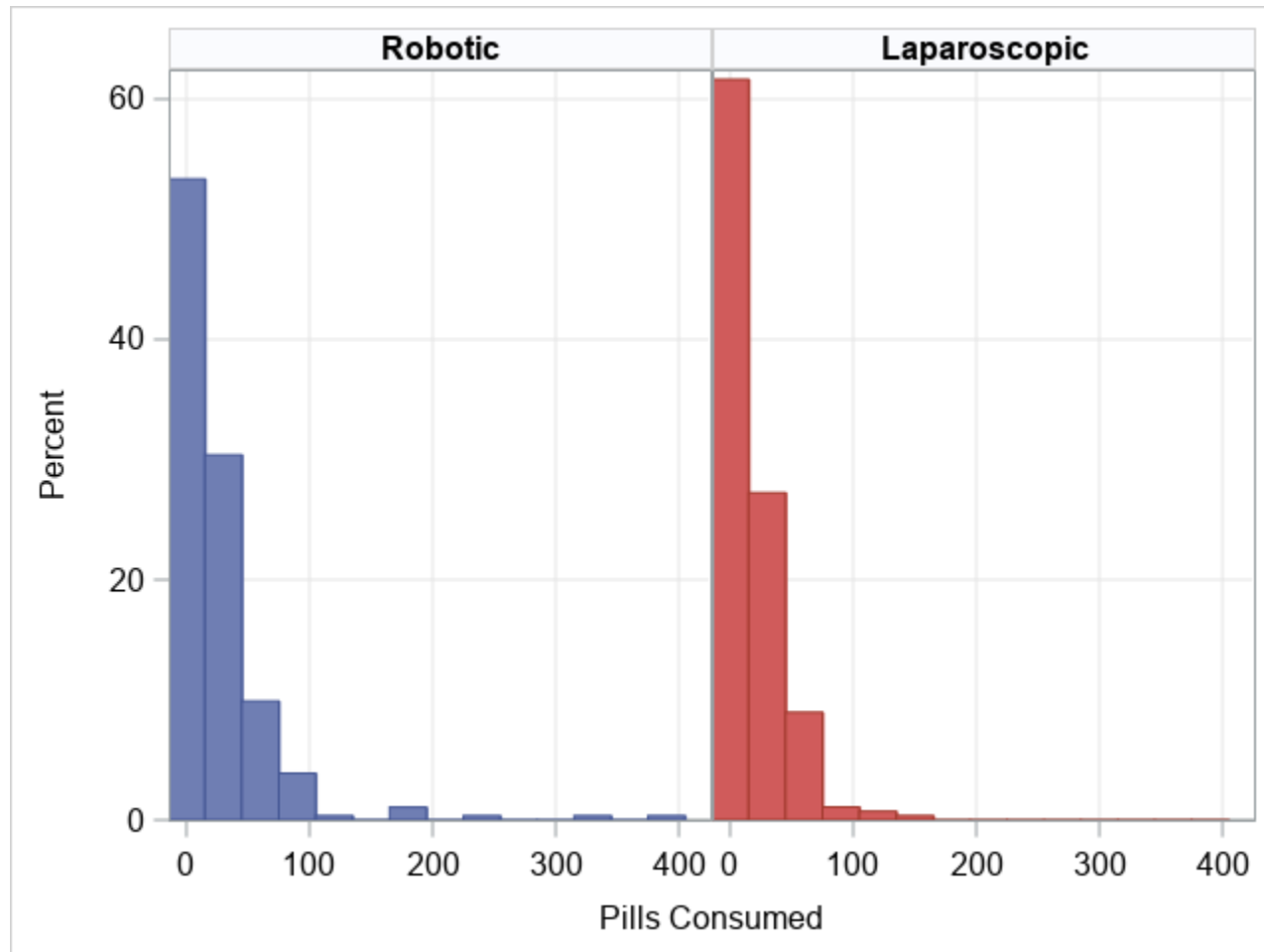


Colectomy Data – ZINB Example, 562 Patients

Pills Consumed	Robotic (N = 283)	Laparoscopic (N = 279)
0	67 (23.7%)	94 (33.7%)
>0 to <5	41 (14.5%)	38 (13.6%)
≥ 5	175 (61.8%)	147 (52.7%)

- Pill consumption trend significantly different between laparoscopic and robotic surgery patients, $p = .010$, Cochran-Armitage trend test. Use “trend” option in Proc Freq Tables statement to get the Cochran-Armitage trend test.
- Difference in pill consumption is significant before adjusting for covariates (Proc EyeBall).
- 28.6% zeroes in data.

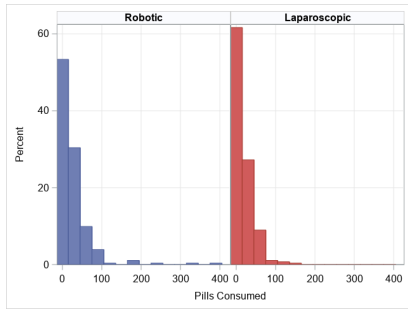
Histogram of Colectomy Data By Surgery Type



Robotic. N=283, mean=23.52, std dev =42.04

Laparoscopic. N=279, mean=15.02, std dev = 21.03

Wilcoxon Test p = 0.007



SAS Code for Histogram

- Proc SGPANEL data=Col_Opioid NOAUTOLEGEND;
- Panelby lap / rows=1 columns=2 HEADERATTRS=(Color=Black Size=12 Weight=Bold) NoVarName;
- Histogram pills_consumed / group=lap;
- rowaxis grid LABELATTRS=(Size=12) VALUEATTRS=(Size=12);
- colaxis grid LABELATTRS=(Size=12) VALUEATTRS=(Size=12) values=(0 to 400 by 50);
- run;



Estimation of ZINB Parameters in SAS (1 of 4)



- `proc genmod data=Col_opioid; *** lap=0, (robotic colectomy) ***;`
- `model pills_consumed= / dist=zinb;`
- `zeromodel;`
- `ods output ParameterEstimates=Parm_GenMod_lap0;`
- `where lap=0; run;`

- `proc genmod data=Col_opioid; *** lap=1, (laparoscopic colectomy) ***;`
- `model pills_consumed= / dist=zinb;`
- `zeromodel;`
- `ods output ParameterEstimates=Parm_GenMod_lap1;`
- `where lap=1; run;`



Estimation of ZINB Parameters in SAS (2 of 4)



- Use estimation method in SAS Global Forum paper, (Erdman, Jackson, Sinko, 2008).
- Proc GenMod output for negative binomial part of ZINB model, lap=0.

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	3.3307	0.0902	3.1539	3.5075	1363.65	<.0001
Dispersion	1	1.6097	0.2309	1.2151	2.1324		

- Proc GenMod output for logistic part of ZINB model for Y=0 vs Y>0.

Parameter	DF	Estimate	Standard Error	Wald 95% Confidence Limits		Wald Chi-Square	Pr > ChiSq
Intercept	1	-1.6673	0.2738	-2.2040	-1.1307	37.08	<.0001



Estimation of ZINB Parameters in SAS (3 of 4)



- data colect_param0; set Parm_GenMod_lap0;
- /* Initialize parameters for lap=0 */
- Dispersion=1.6097;
- B0_ZINB=3.3307; /* intercept ZINB model */
- B0_ZeroLogistic =-1.6673; /* B0 logistic model for count=0 */

- /* Computed parameters */
- $p_ZeroLogistic = \exp(B0_ZeroLogistic) / (1 + \exp(B0_ZeroLogistic))$;
- $NB_k = 1 / \text{Dispersion}$; /* Number of successes, negative binomial */
- $NB_p = 1 / (1 + (\text{Dispersion} * \exp(B0_ZINB)))$; /* p=Prob success, negative binomial */
- run;

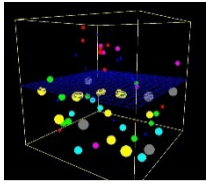


Estimation of ZINB Parameters in SAS (4 of 4)



- **** First, simulate 1 dataset of N=283, lap=0 ***;**
- `data SimLap0_1rep; set colect_param0;`
- `Call streaminit(11262); /* Initialize stream of random number with a seed */`

- `do i=1 to 283;`
- `p_Uniform = Rand('Uniform',0, 1);`
- `if (p_Uniform > p_Zerologistic) then pills_consumed = rand('NEGB', NB_p, NB_k);`
- `else pills_consumed=0;`
- `Simulate=1; lap=0;`
- `val_los=rand('Normal', 7.76, 5.19); /* unequal lap 1 & 0 */`
- `Age4564=rand('BERNOULLI', .431); /* equal lap 1 & 0 */`
- `Output; end;`



Simulate Covariates



- `val_los=rand('Normal', 7.76, 5.19);`
- `Age4564=rand('BERNOULLI', .431);`
- `AgeGE65=rand('BERNOULLI', .491);`
- `smoker= rand('BERNOULLI', .235);`
- `reg_obese= rand('BERNOULLI', .432);`
- `stoma = rand('BERNOULLI', .336);`
- `ASACat2_3 = rand('BERNOULLI', .898);`
- `ASACat4_5 = rand('BERNOULLI', .085);`
- `CancerAdema = rand('BERNOULLI', .350);`
- `Diverticular= rand('BERNOULLI', .336);`
- `Urgent = rand('BERNOULLI', .477);`
- `flg_cmp_any = rand('BERNOULLI', .173);`



Compare Pills Consumed, Actual and Simulated, 1 Rep

- **/* Example with Robotic Colectomy (lap=0) */**
- Data Both_lap0;
- set Col_Opioid SimLap0_1rep;
- keep lap simulate pills_consumed ConsumeCat3;
- where lap=0; run;

- **/* Wilcoxon p-value = 0.734 for pills_consumed between simiuated & actual data */**
- Proc NPar1Way Wilcoxon Data = Both_lap0;
- Class simulate;
- Var pills_consumed;
- run;

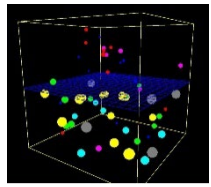
Simulate	N	Mean	Std Dev
0	283	23.5	42.0
1	283	23.7	32.8



Compare All Covariates, Actual and Simulated, 1 Rep

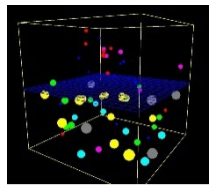
- /* Compare Binary Variables with Chi-Square Test */
- Proc Freq Data= Combined_1rep;
- Tables (Age4564 AgeGE65 smoker reg_obese stoma ASACat2_3 ASACat4_5 CancerAdema Diverticular Urgent flg_cmp_any) * simulate / ChiSq Out=FreqOut ;
- Run;

- /* Compare Normal Variables with Proc Test */
- Proc Ttest Data= Combined_1rep;
- Class Simulate;
- Var LOS;
- Run;



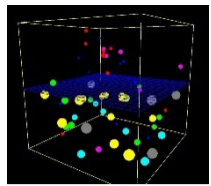
Simulate 1000 Reps for Robotic and Laparoscopic

- Begin by estimating model parameters from Proc GenMod for lap=0 (robotic colectomy).
- ods html path="c:\temp";
- proc genmod data=Col_opioid;
- model pills_consumed= / dist=zinb;
- zeromodel;
- ods output ParameterEstimates=Parm_GenMod_lap0;
- where lap=0;
- run;
- ods html close;



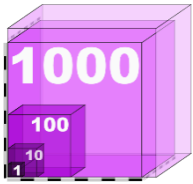
Simulate 1000 Reps for Robotic and Laparoscopic

- data colect_param0;
- /* Initialize parameters for lap=0 */
- Dispersion=1.6097;
- B0_ZINB=3.3307; /* intercept ZINB model */
- B0_ZeroLogistic =-1.6673; /* B0 logistic model for count=0 */
-
- /* Computed parameters */
- p_Zerologistic=exp(B0_ZeroLogistic)/(1+exp(B0_ZeroLogistic));
- NB_k = 1/Dispersion; /* Number of successes, negative binomial */
- NB_p = 1/(1 + (Dispersion*exp(B0_ZINB))); /* p=Prob success, negative binomial */
- run;



Simulate 1000 Reps for Robotic and Laparoscopic

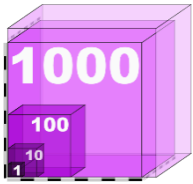
- data SimLap0_1000rep;
- set colect_param0;
- Call streaminit(11263); /* Initialize stream of random number with a seed */
-
- do rep=1 to 1000;
- do i=1 to 283;
- p_Uniform = Rand('Uniform',0, 1);
- if (p_Uniform > p_Zerologistic) then
- pills_consumed = rand('NEGB', NB_p, NB_k);
- else pills_consumed=0;
- pills_prescribed=rand('Normal', 43.46, 45.86); /* unequal lap 1 & 0 */
- Simulate=1; lap=0; output;
- End; end; run;



Repeat for Laparoscopic Surgery

```
/* Step 1: Estimate ZINB Model Parameters */
```

```
ods html path="c:\temp";  
proc genmod data=Col_opioid;  
model pills_consumed= / dist=zinb;  
zeromodel;  
ods output ParameterEstimates=Parm_GenMod_lap1;  
where lap=1;  
run;  
ods html close;
```

Repeat for Laparoscopic Surgery

```
/* Initialize parameters for lap=1 */
```

```
data colect_param1;
```

```
Dispersion=1.0668;
```

```
B0_ZINB=3.0682; /* intercept ZINB model */
```

```
B0_ZeroLogistic =-0.8411; /* B0 logistic model for count=0 */
```

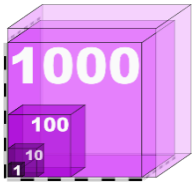
```
/* Computed parameters */
```

```
p_ZeroLogistic=exp(B0_ZeroLogistic)/(1+exp(B0_ZeroLogistic));
```

```
NB_k = 1/Dispersion; /* Number of successes, negative binomial */
```

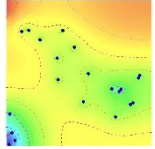
```
NB_p = 1/(1 + (Dispersion*exp(B0_ZINB))); /* p=Prob success, negative  
binomial */
```

```
run;
```



Repeat for Laparoscopic Surgery

```
data SimLap1_1000rep;
set colect_param1;
Call streaminit(20194);
do rep=1 to 1000;
do i=1 to 279;
  p_Uniform = Rand('Uniform',0, 1);
  if (p_Uniform > p_Zerologistic) then
    pills_consumed = rand('NEGB', NB_p, NB_k);
  else pills_consumed=0;
Simulate=1;
lap=1;
pills_prescribed=rand('Normal', 36.4, 23.3); /* unequal lap 1 & 0 */
output;
end; /* end do 1 to 279 */
end; /* end do rep=1 to 1000 */
run;
```

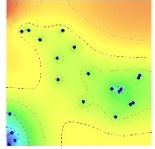


Combine Simulations and Run Model

```
/* Combine simulated datasets and sort. */
```

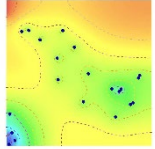
- Data SimLap_1000rep;
- Set SimLap0_1000rep SimLap1_1000rep;
- run;

- proc sort data=SimLap_1000rep;
- by Rep Lap;
- run;



Combine Simulations and Run Model

- `proc datasets library=work;`
 - `modify SimLap_1000rep;`
 - **`index create Rep;`**
 - `run;`
 - `quit;`
- Adding an index to Rep will help the power calculation to run faster. When Proc GenMod (or other SAS procedures) are run with a “by” statement and the variable in the “by” statement is indexed, the procedure will run faster. (Raithel, Creating and Exploiting SAS® Indexes, 2000).

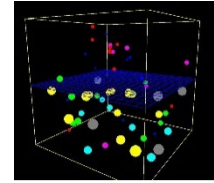


Combine Simulations and Run Model

- Proc GenMod data=SimLap_1000Rep;
- Model pills_consumed=lap pills_prescribed/dist=zinb;
-
- Zeromodel lap pills_prescribed;
- by Rep;
- Ods Output ParameterEstimates=Parm_NB_LapOnly
ZeroParameterEstimates=Parm_ZeroModel_LapOnly;
- Run;
- run;



Calculate Power from Simulation

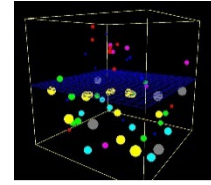


- **/* Power for Negative Binomial Part of Model */**
- data Parm_NB;
- set Parm_NB(keep=Rep Parameter Estimate ProbChiSq);
- Where parameter='lap';
- if ProbChiSq<.05 then lapSignificant=1;
- if ProbChiSq>=.05 then lapSignificant=0;

- **/* Power for 0 vs >0 Model */**
- Data Parm_ZeroModel;
- setParm_ZeroModel(keep=Rep Parameter Estimate ProbChiSq);
- Where parameter='lap';
- if ProbChiSq<.05 then lapSignificant=1; if ProbChiSq>=.05 then lapSignificant=0;



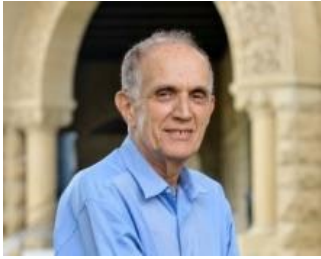
Calculate Power from Simulation



- **/* Power for Negative Binomial Part of Model */**
- Proc Freq data = Parm_NB;
- Tables lapSignificant;
- Run;

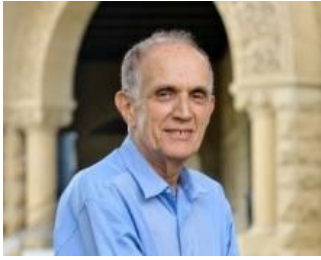
- Negative binomial power was of interest to researchers, not the zero model.
- Power for the negative binomial model was 78% with no predictors other than lap and pills_prescribed.

- To investigate the effect on power by increasing the sample size to 500 or larger, change the N's of 279 and 283 to 500 or larger, and re-run Procs GenMod and Freq to get the resulting power.



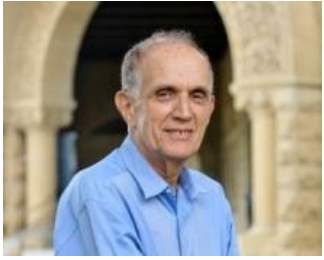
Bootstrapping, 1 of 4

- In previous power calculation, pills_consumed and pill_prescribed were generated from independent simulations. Correlations between variables not taken into account.
- Similar to the way that Proc Power estimates power for independent predictors in a logistic regression.
- **proc power;** logistic vardist("Heat") = ordinal((5 10 15 20) : (0.2 0.3 0.3 0.2))
vardist("Soak") = ordinal((2 4 6) : (0.4 0.4 0.2)) vardist("Mass1")=normal(4, 1)
- Bootstrapping for power calculation is another option.(Roberts and Fan, 2002), (Efron and Tibshirani, 1993).



Bootstrapping 2 of 4

- Generate B bootstrap samples of all variables to be included in the model.
- Use actual sample size in the data, but sample with replacement. Bootstrapped samples capture correlation between variables.
- Easy to implement in SAS. Use method = urs with Proc SurveySelect
- /* Create bootstrap samples */
- PROC SURVEYSELECT data=Col_OpioidLap0 Seed=11202019 method=urs n=283
reps=1000 out=BootLap0 outhits; run;
- PROC SURVEYSELECT data=Col_OpioidLap1 Seed=11212019 method=urs n=279
reps=1000 out=BootLap1 outhits; run;

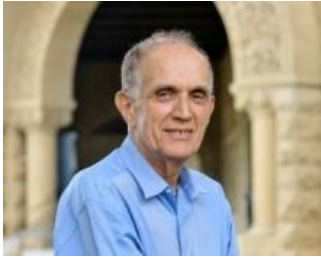


Bootstrapping 3 of 4

- `/* Combine datasets and sort by Rep */`
- `Data BootLap;`
- `Set BootLap0(Rename=(Replicate=Rep)) BootLap1(Rename=(Replicate=Rep));`

- `proc sort data=BootLap;`
- `by Rep; run;`

- `/* Put index on Rep for faster execution */`
- `proc datasets library=work;`
- `modify BootLap;`
- `index create Rep;`
- `run; quit;`



Bootstrapping 4 of 4

- `/* Model with Pills Prescribed */`
- `Proc GenMod data=BootLap;`
- `Model pills_consumed=lap pills_prescribed / dist=zinb;`
- `Zeromodel lap pills_prescribed;`
- `by Rep;`
- `Ods Output ParameterEstimates=Parm_NB_presc`
`ZeroParameterEstimates=Parm_ZeroModel_presc; Run;`
- Process output with data step and Proc Freq, same as before.
- With bootstrapping, power = 72%.



Power, Final Results

- Power, based on simulation, power was 78% with the data step simulation method and 72% with bootstrapping.
- Researchers proceeded with original data and included Brandy's power analysis in the response to the reviewers, rather than waiting to collect more data for 80% power.
- Vu J, Cleary RK, Englesbe MJ, Brummett CM, Waljee JF, Suwanabol PA. Post-discharge Opioid Consumption After Minimally Invasive and Open Colectomy: Does Operative Approach Matter? *Ann Surg*. 2020 Jul 9. doi: 10.1097/SLA.0000000000004240. Epub ahead of print. PMID: 32657943.



Conclusions

- Begin simulation with a single rep. Compare original and simulated distributions of the outcome variable and covariates before proceeding.
- Most SAS statistical procedures can be run with a “by Rep” option to calculate power by simulation. Adding an index to the Rep variable helps the simulations to run faster.
- To compute power, output the results from the statistical procedure, calculate the proportion of significant findings, and display the percentage of significant findings with Proc Freq.
- When the distribution of an outcome is not available in Proc Power or the investigator questions the results of Proc Power, simulation is a useful tool.



References

- Efron B, Tibshirani R. An introduction to the bootstrap. New York: Chapman & Hall; 1993.
- Erdman D., Jackson L., Sinko A. Zero-Inflated Poisson and Zero-Inflated Negative Binomial Models Using the COUNTREG Procedure. SAS Global Forum (2008).
<https://support.sas.com/resources/papers/sgf2008/countreg.pdf>.
- Fan X., Roberts J.K. Bootstrapping within the Multilevel/Hierarchical Linear Modeling Framework: A Primer for Use with SAS and SPLUS. Multiple Linear Regression Viewpoints, 2004, Vol. 30(1), 23-34.
- Greene W. H. (1994), Accounting for Excess Zeros and Sample Selection in Poisson and Negative Binomial Regression Models, Technical report.
- Lambert D. (1992), “Zero-Inflated Poisson Regression Models with an Application to Defects in Manufacturing,” Technometrics, 34, 1–14.
- Raithel M.A. (2000). Creating and Exploiting SAS® Indexes. SUGI Paper 123-29.
- Vu J, Cleary RK, Englesbe MJ, Brummett CM, Waljee JF, Suwanabol PA. Post-discharge Opioid Consumption After Minimally Invasive and Open Colectomy: Does Operative Approach Matter? Ann Surg. 2020 Jul 9. doi: 10.1097/SLA.0000000000004240.



Contact Information

- Brandy R. Sinco, Statistician and Programmer/Analyst
- Michigan Medicine
- CHOP (Center for Healthcare Outcomes and Policy)
- Ann Arbor, MI 48109-2800
- E-Mail: brsinco@med.umich.edu