

Power Approximations for Generalized Linear Models Using the Signal-to-Noise Transformation Method¹

Thomas H. Johnson, Laura J. Freeman, James D. Simpson, and Colin E. Anderson

Statistical power is a useful measure for assessing the adequacy of an operational test. It is the probability of correctly concluding that a factor in the experiment significantly impacts the response variable. For normally distributed response variables, power calculations are widely available in experimental design software. However, many defense testing applications use non-normal response variables. Generalized linear models provide many useful analysis methods for non-normal responses. While statistical software routinely includes generalized linear models in model-fitting packages, power calculations for generalized linear models are not widely available in experimental design modules. This paper proposes a signal-to-noise transformation method (SNRx) that enables generalized linear model power approximations using normal linear model power equations, making them generally available to all practitioners.



¹ Based on “Power Approximations for Generalized Linear Models Using the Signal-to-Noise Transformation Method,” *Quality Engineering*, October 2017, <https://www.tandfonline.com/doi/full/10.1080/08982112.2017.1361537>.

Introduction

Experimental designs are used to help with planning, executing, and analyzing an experiment. In the planning phase test objectives are determined. These objectives guide the development of the factors, levels, and response variables (Freeman et al. 2013). Recent Department of Defense policy has emphasized the importance of using principles of design of experiments in all operational testing (Johnson et al. 2012, 61).

Our goal is to provide a simple method to obtain power for a generalized linear model by transforming the effect size in the power calculation for a classical linear model.

Equally important in the planning phase is the assessment of the experimental design. An assortment of measures is available to assess the goodness of an experiment prior to data collection. Hahn, Meeker, and Feder (1976) call these *measures of precision*. These include standard error of predicted mean responses, standard error of coefficients, correlations metrics, and optimality criteria values. Measures of precision are affected by many aspects of the plan for the experiment, including the choice of factors and levels, the assumed model form, the combination of factor settings from run to run, and the total number of runs.

Power—the focus of this paper—is an important measure of precision. Power is the probability of correctly concluding that an effect has an impact on the response variable. In general, the power of an effect increases with sample size, making it a useful measure for determining the scope of an operational test. Here, we focus on a second-order model for designs with multilevel categorical factors. Effects considered include the main effects and two-factor interactions (Montgomery 2008, 4).

Experimental design software that calculates power for classical linear models is widely available. However, power calculations should reflect the knowledge that the result will not be normally distributed, when it is known before running the experiment. Techniques for calculating power for experimental designs with generalized linear models are not widely available in commercial software; such calculations usually require Monte Carlo simulation studies. Accounting for the knowledge of the planned analysis is important when planning the test because different distributions can require dramatically different sample sizes to achieve high-effect power.

Our goal is to provide a simple method to obtain power for a generalized linear model by *transforming* the effect size in the power calculation for a classical linear model. Existing software (e.g., JMP, Minitab, and Design Expert) that accommodates classical linear model power calculations allows the user to adjust the signal-to-noise ratio or alter the model coefficients under the alternative hypothesis. SNRx provides a means of setting the signal-to-noise ratio or the coefficients so that the calculation represents the generalized linear model power calculation. The target audience of SNRx is the analyst who has statistical design experience and is comfortable working with popular statistical

software, but who is not inclined to calculate power for generalized linear models using custom code and Monte Carlo simulation.

Model Formulation

A generalized linear model generalizes the classical linear model and is defined in terms of its three components (McCullagh and Nelder 1989, 27):

Random component. Response variables Y_1, \dots, Y_n share the same distribution from the exponential distribution family, where the v th response of the experiment has an expected value equal to the mean, μ_v .

Systematic component. The unknown coefficients systematically specify the linear predictor η_v such that $\eta_v = Z_v \psi + X_v \lambda$, where Z_v and X_v represent the v th row of the test and nuisance matrix.

Link between the random and systematic components. The link function $g(\cdot)$ relates the mean and linear predictor in the expression $(\mu_v) = \eta_v$.

Generalized linear models may also include as special cases linear regression, logistic regression, and log-linear models for count data.

Model Inference

We are interested in a hypothesis test for the significance of a multilevel categorical factor or interaction between multilevel categorical factors. Specifically, we want to be able to test whether the coefficients belonging to a main effect or two-factor interaction effect are equal to zero. Thus, the hypothesis test for an individual effect is

$$H_0: \psi = 0,$$

$$H_1: \psi \neq 0.$$

The classical and generalized linear models use similar techniques for evaluating these hypothesis tests. A classical linear model uses analysis of variance (ANOVA), which is based on an F statistic. The analogue of an ANOVA for generalized linear models is an analysis of deviance, which is based on a likelihood ratio statistic.

Some classical linear model software allows the user to specify the details of a planned experiment, and the software outputs the power associated with this hypothesis test. The user can input the design matrix, choose the model form, set the anticipated coefficients (i.e., set ψ under H_1), and obtain power.

The SNRx method is useful in situations where the practitioner only has access to classical linear model software, but is interested in calculating power for a specific generalized linear model. In this situation, the SNRx method sets ψ under H_1 so that the ANOVA hypothesis test well represents an analysis of deviance for the specific generalized linear model.

SNRx Method

The approach assumes that for each run in the experiment ($v = 1, 2, \dots, N$) the linear predictors η_v in a generalized linear model can be modeled as the response variable Y_v in a classical linear model. That is, $Y_v = \eta_v = Z_v \psi + X_v \lambda + \epsilon_v$, where $\epsilon_v \sim N(0, \sigma^2)$, and the error term ϵ_v is independent and identically distributed. The variance σ^2 is the transformed noise, meaning it represents the variance of the linear predictor for the generalized linear model.

Another assumption in this approach is that σ^2 is constant and is evaluated at the overall mean across the design space $\bar{\mu}$. For example, an analyst may anticipate a 70 percent average probability of success across the design space that can be fit with a logistic regression model. The overall mean $\bar{\mu}$ impacts σ^2 and, in turn, affects power.

A tenet of generalized linear models is that the variance of Y depends on the mean μ and the dispersion parameter ϕ . Since we are assuming a nonzero effect size for ψ under the alternative hypothesis, an implication is that μ is not constant; thus, neither is σ^2 . For this reason, only small effect sizes should be considered.

Another assumption is that the hypothesis test is constructed without considering nuisance effects. That is, for the hypothesis test $\psi = 0$, the nuisance coefficients take the form $\lambda = (\lambda_{int}|0)T$. Without this assumption, significant values of λ could further invalidate the assumption that σ^2 is constant because λ impacts μ , which, in turn, affects the variance of Y .

We define the signal-to-noise ratio as $\kappa = \delta/\sigma$. For SNRx, we must transform δ and σ to the linear predictor space. Since Y is a random variable with $E(Y) = \mu$, we can use $g(Y)$ as an estimator of $g(\mu)$. Using the delta method from Casella and Berger (2002), we can approximate that

$$E(g(Y)) \approx g(\mu),$$
$$\text{Var}(g(Y)) \approx [g'(\mu)]^2 \text{Var}(Y).$$

We also know that $\text{Var}(Y) = a(\phi)\text{Var}(\mu)$ for generalized linear models. Substituting this into the above equation, taking the square root, and evaluating $g'(\mu)$ and $\text{Var}(\mu)$ at $\bar{\mu}$, we obtain the following estimate of the noise:

$$\sigma = \sqrt{\text{Var}(g(Y))} = g'(\bar{\mu})\sqrt{a(\phi)\text{Var}(\bar{\mu})}.$$

Now that the noise is transformed, we turn our attention to the signal. If the upper and lower bounds of the signal of interest are $\bar{\mu} + \delta/2$ and $\bar{\mu} - \delta/2$, we can convert this quantity to a value in the linear predictor space as $g(\bar{\mu} + \delta/2)$ and $g(\bar{\mu} - \delta/2)$, respectively, where $g(\cdot)$ is the link function for the generalized linear model of interest.

The signal-to-noise ratio is described as the ratio of the signal and noise within the linear predictor space, as shown in the equation below.

$$\kappa = \frac{g(\bar{\mu} + \delta/2) - g(\bar{\mu} - \delta/2)}{g'(\bar{\mu})\sqrt{a(\phi)V(\bar{\mu})}}$$

Mission Success Example

In logistic regression, the response variable is binary (1 or 0). For this example, let 1 and 0 represent a mission success and failure, respectively. In an experiment with N groups or strata, Y_v represents the number of successes in the v th group out of m_v attempts, where $v = 1, 2, \dots, n$. Then, a logistic regression model assumes that $Y_v \sim \text{binom}(m_v, \pi_v)$.

A few pieces of information are needed to set up the power calculation. The first is the assumed mean response across the design space $\bar{\mu}$. For logistic regression, the mean response is bounded between zero and one and represents the average probability of success across the design space. For this example, we assume a nominal 70 percent probability of success, or $\bar{\mu} = 0.7$.

The second element is the effect size δ . Recall that δ is the change in the mean response that is symmetric about $\bar{\mu}$. In this example, we assume $\delta = 0.3$ so that the change of interest ranges from 55 to 85 percent probability of success.

The next step is to calculate the signal-to-noise ratio κ . The signal-to-noise ratio can be directly inputted into some software, such as Design Expert, and the corresponding effect power is outputted. In other software, such as JMP, the coefficients anticipated under the alternative hypothesis must be manually inputted using the approach outlined below. Using the assumed values for this example, we get

$$\kappa = \frac{\log\left(\frac{.7 + .3/2}{1 - (.7 + .3/2)}\right) - \log\left(\frac{.7 - .3/2}{1 - (.7 - .3/2)}\right)}{\frac{1}{.7(1 - .7)}\sqrt{(1).7(1 - .7)}} = .70.$$

To obtain the approximate coefficients, we first construct the marginal mean effect so that its range is equal to κ and then convert it to coefficients. The coefficients for a three-level main effect are

$$\psi = [.70/2 - .70/2]T.$$

In this example, assume the experiment includes three factors and the sample size is 96. That is, the operational test includes 96 missions. The experimental design is a full factorial that is replicated four times so the model matrix M is size 96×18 . The first column of M corresponds to the intercept, columns 2

through 7 correspond to the main effects, and columns 8 through 18 correspond to the two factor interactions. The coefficient vector β is size 18×1 . The power calculation requires that we split the model matrix into the test matrix Z and the nuisance matrix X .

For the test on the main effect, the test matrix Z is size 96×2 , and the previously calculated test coefficient vector ψ is size 2×1 . The nuisance matrix is 96×16 . We calculate the hat matrix W , and use W , Z , and ψ in the equation for the noncentrality parameter, which is given as

$$\gamma_F = (Z\psi)^T (I - W)(Z\psi),$$

and we find that $\gamma_F = 7.91$. By setting the significance $\alpha = 0.05$, we then calculate the critical F value that is equal to $f_{crit} = 3.11$. Finally, we calculate power, which is equal to 0.69. Clearly, 96 missions does not provide enough power to determine if the main effect significantly affects mission success. Additional missions are required to provide a robust evaluation. Further details about this calculation can be found in the full-length version of this paper.

Conclusion

This work provides a practical approach for sizing operational tests. Compared to current approaches, our hope is that this methodology will be more accessible to the test and evaluation community. Properly scoped tests should lead to more rigorous evaluations, which, in turn, should lead to well-informed acquisition decisions.

References

- Casella, G., and R. L. Berger. 2002. *Statistical Inference*. 2nd Edition. Pacific Grove, CA: Duxbury.
- Freeman, L. J., A. G. Ryan, J. L. Kensler, R. M. Dickinson, and G. G. Vining. 2013. "A Tutorial on the Planning of Experiments." *Quality Engineering* 25: 315-332.
- Hahn, G. J., W. Q. Meeker Jr., and P. I. Feder. 1976. "The Evaluation and Comparison of Experimental Designs for Fitting Regression Relationships." *Journal of Quality Technology* 8: 140-157. <https://doi.org/10.1080/00224065.1976.11980735>.
- Johnson, R. T., G. T. Hutto, J. R. Simpson, and D. C. Montgomery. 2012. "Designed Experiments for the Defense Community." *Quality Engineering* 24: 60-79. <https://doi.org/10.1080/08982112.2012.627288>.
- McCullagh, P., and J. A. Nelder. 1972. *Generalized Linear Models*. 2nd edition. Chapman and Hall, Inc.
- Montgomery, D. C. 2008. *Design and Analysis of Experiments*. 7th edition. John Wiley & Sons.



Colin Anderson (left), a Research Staff Member in the Operational Evaluation Division (OED) of IDA's Systems and Analyses Center, holds a doctorate in physics from Yale University. *Laura Freeman*, former Assistant Director in OED, holds a doctorate in statistics from Virginia Tech. *Tom Johnson* (right), a Research Staff Member in OED, holds a doctorate in aerospace engineering from Old Dominion University. *Jim Simpson* (no photo available), a consultant in OED, holds a doctorate in industrial engineering from Arizona State University.