**Power Analysis for Binary Logistic Regression**

G\*Power will estimate the sample size needed to have the desired amount of power for one predictor in a binary logistic regression analysis. The [G\*Power manual](http://www.psychologie.hhu.de/fileadmin/redaktion/Fakultaeten/Mathematisch-Naturwissenschaftliche_Fakultaet/Psychologie/AAP/gpower/GPowerManual.pdf) includes an example with one normally distributed predictor variable which has been standardized to mean zero, variance one. The results show that 337 cases are necessary to obtain 95% power to detect an effect of odds ratio = 1.5.

**z tests -** Logistic regression

**Options:** Large sample z-Test, Demidenko (2007) with var corr

**Analysis:** A priori: Compute required sample size

**Input:** Tail(s) = Two

Odds ratio = 1.5

Pr(Y=1|X=1) H0 = 0.5

α err prob = 0.05

Power (1-β err prob) = 0.95

R² other X = 0

X distribution = Normal

X parm μ = 0

X parm σ = 1

**Output:** Critical z = 1.9599640

Total sample size = 337

Actual power = 0.9500770

Suppose that there are additional predictors and the *R2* for predicting X from those other predictors is .1. Now 375 cases are needed to have 95% power for detecting the unique effect of X.

**z tests -** Logistic regression

**Options:** Large sample z-Test, Demidenko (2007) with var corr

**Analysis:** A priori: Compute required sample size

**Input:** Tail(s) = Two

Odds ratio = 1.5

Pr(Y=1|X=1) H0 = 0.5

α err prob = 0.05

Power (1-β err prob) = 0.95

R² other X = .1

X distribution = Normal

X parm μ = 0

X parm σ = 1

**Output:** Critical z = 1.9599640

Total sample size = 375

Actual power = 0.9503607

Suppose that X is more strongly related to the other predictors, such that *R2* = .2. Now 422 cases are needed. The greater the correlation between X and the other predictors, the more cases will be necessary to detect X’s unique effect.

**z tests -** Logistic regression

**Options:** Large sample z-Test, Demidenko (2007) with var corr

**Analysis:** A priori: Compute required sample size

**Input:** Tail(s) = Two

Odds ratio = 1.5

Pr(Y=1|X=1) H0 = 0.5

α err prob = 0.05

Power (1-β err prob) = 0.95

R² other X = .2

X distribution = Normal

X parm μ = 0

X parm σ = 1

**Output:** Critical z = 1.9599640

Total sample size = 422

Actual power = 0.9504172

The base rate of the target event will also affect power. The more the base rate deviates from 50%, the more cases will be needed. For example, suppose the base rate was only 20%. Now 628 cases are needed.

**z tests -** Logistic regression

**Options:** Large sample z-Test, Demidenko (2007) with var corr

**Analysis:** A priori: Compute required sample size

**Input:** Tail(s) = Two

Odds ratio = 1.5

Pr(Y=1|X=1) H0 = 0.2

α err prob = 0.05

Power (1-β err prob) = 0.95

R² other X = .2

X distribution = Normal

X parm μ = 0

X parm σ = 1

**Output:** Critical z = 1.9599640

Total sample size = 628

Actual power = 0.9500825

The G\*Power manual also gives an example with a binary predictor variable. We assume that when X has value 1 (instead of 0) the probability of the event is .1 under the alternative hypothesis, but under the null that probability is .05. This translates to an odds ratio of (.1/.9)/(.05/.95) = 2.1111. We assume that the proportion of cases for which X = 1 is .5 (π). We need 1,437 cases for 95% power.

**z tests -** Logistic regression

**Options:** Large sample z-Test, Demidenko (2007) with var corr

**Analysis:** A priori: Compute required sample size

**Input:** Tail(s) = Two

Odds ratio = 2.1111111

Pr(Y=1|X=1) H0 = 0.05

α err prob = 0.05

Power (1-β err prob) = 0.95

R² other X = 0

X distribution = Binomial

X parm π = 0.5

**Output:** Critical z = 1.9599640

Total sample size = 1437

Actual power = 0.9501044

Now, suppose there are other predictors which collectively correlated with X such that *R2* = .2. Now we need 1,796 cases.

**z tests -** Logistic regression

**Options:** Large sample z-Test, Demidenko (2007) with var corr

**Analysis:** A priori: Compute required sample size

**Input:** Tail(s) = Two

Odds ratio = 2.1111111

Pr(Y=1|X=1) H0 = 0.05

α err prob = 0.05

Power (1-β err prob) = 0.95

R² other X = .2

X distribution = Binomial

X parm π = 0.5

**Output:** Critical z = 1.9599640

Total sample size = 1796

Of course, if one desires to estimate the sample size for additional predictors in a multiple binary regression, e can conduct a power analysis for each of the predictors of interest.

OK, but how does one decide on a value for the odds ratio used in the power analysis? It may be possible to use past research results to determine what is a reasonable estimate for that parameter. Alternatively, one might be willing to assume that the odds ratio is small, medium, or large in size, but what are benchmark values for small, medium, and large? For dichotomous predictors, Ferguson( suggested that an odds ratio of 2 is small, 3 is medium, and 4 is large (but see Chen, Cohen, and Chen, 2010 and Oliver & Bell, 2013). For a continuous predictor the odds ratio reflects the changes in the odds of the target event when the predictor increases by one point. The value of that odds ratio depends greatly on the unit of measure (for example, grams or tons). One way to eliminate the effect of unit of measure is to standardize to mean 0, variance 1, the continuous predictor variable. The odds ratio will then indicate the change in the odds of the target event when the predictor increases by one standard deviation. I have not been able to find any discussion of benchmarks for an odds ratio where the predictor has been standardized.

It is, in my humble opinion, unreasonably difficult to conduct a power analysis to test the null hypothesis that a predictor has no unique effect in a multiple binary logistic regression. The difficulty arises from needing to estimate not only the individual effects of the predictors but also the correlations between predictors and the shape of the predictors’ distributions. Sometimes it is assumed that the predictors are independent of each other, a condition likely only true when they are experimentally manipulated.

I think it important to estimate the effect of each predictor ignoring all other predictors as well as the unique effects of each predictor. While it is true that the power for detection of the unique effect of a predictor will be reduced when it is correlated with other predictors in the model, I would feel comfortable with a sample size that would give me adequate power for each of the important predictors when evaluated ignoring overlap with the other predictors.

Consider this example. Suzie Cue wants to predict binary outcome Y from four predictors, two of which are binary and the other two continuous and normally distributed. She thinks all four predictors equally important. She is willing to assume that the continuous predictors are normally distributed. She can determine the power/sample size for the continuous predictors by doing a power analysis for a *t* test (or for a nonparametric test if distributions are not expected to be normal). Suppose that she wants enough data to have 80% power for a medium-sized effect and she estimates that 25% of the total sample will display the target even and 75% not. She will need 170 cases.

**t tests -** Means: Difference between two independent means (two groups)

**Analysis:** A priori: Compute required sample size

**Input:** Tail(s) = Two

Effect size d = 0.5

α err prob = 0.05

Power (1-β err prob) = 0.80

Allocation ratio N2/N1 = 3

**Output:** Noncentrality parameter δ = 2.8338811

Critical t = 1.9741852

Df = 168

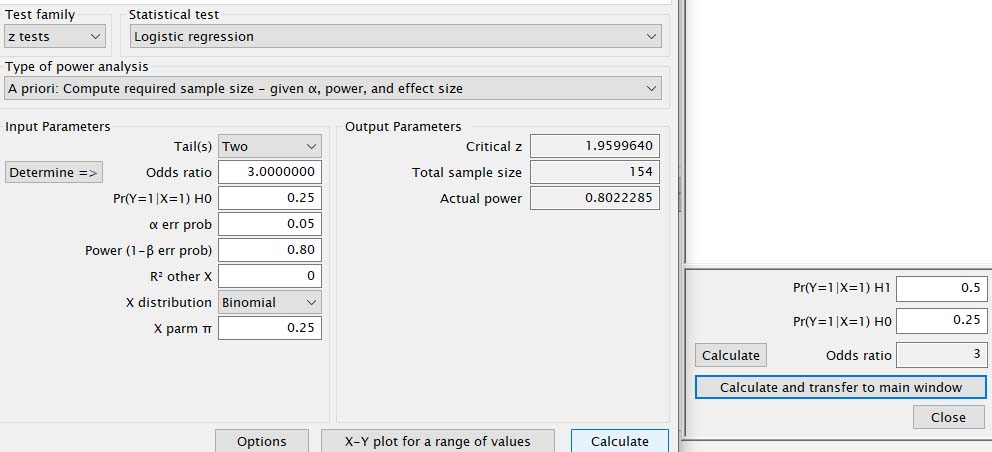
Sample size group 1 = 43

Sample size group 2 = 127

Total sample size = 170

Actual power = 0.8044668

She also wants 80% power to detect a medium-sized effect of the dichotomous predictors. She uses Ferguson’s recommended value of 3 for a medium-sized odds ratio. She expects the percentage of cases having the target outcome to be 25% for the total sample. She will need 154 cases.



Alternatively, Suzie decides to use phi as the effect size estimate, and a phi of .3 is a medium-sized phi. She needs only 88 cases.

**χ² tests -** Goodness-of-fit tests: Contingency tables

**Analysis:** A priori: Compute required sample size

**Input:** Effect size w = 0.3

α err prob = 0.05

Power (1-β err prob) = 0.80

Df = 1

**Output:** Noncentrality parameter λ = 7.9200000

Critical χ² = 3.8414588

Total sample size = 88

Actual power = 0.8035275

References

Ferguson, C. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice*, *40*: 532–538.

Chen, H., Cohen, P., & Chen, S. (2010). [How big is a big odds ratio? Interpreting the magnitudes of odds ratios in epidemiological studies](https://www.tandfonline.com/doi/full/10.1080/03610911003650383). *Communications in Statistics - Simulation and Computation*, *39*, 860-864.

Oliver, J., & Bell, M. L. (2013). [Effect sizes for 2×2 contingency tables](https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0058777). *PLoS ONE*, *8*(3): e58777. <https://doi.org/10.1371/journal.pone.0058777>

Wuensch, K. L. [The correspondence between phi and odds ratios](http://core.ecu.edu/psyc/wuenschk/StatHelp/Phi-OddsRatio.docx) – it depends the distribution of the marginals.

[Karl L. Wuensch](http://core.ecu.edu/psyc/wuenschk/klw.htm), July, 2019.