**TECHNICAL REPORT # 59**
**MAY 2013**

# Interim sample size recalculation for linear and logistic regression models: a comprehensive Monte-Carlo study

Sergey Tarima, Peng He, Tao Wang, Aniko Szabo
Division of Biostatistics, Medical College of Wisconsin

**Abstract**

We propose a simple procedure for interim sample size recalculation when testing a hypothesis on a regression coefficient and explore its effects on type I and II errors. We consider hypothesis testing in linear and logistic regression models using the Wald test. We performed a comprehensive Monte-Carlo study comprised of 1092 experiments with $60,000$ repetitions each. In these experiments we varied the number of predictors (1, 2, or 10), the type of predictors (binary and continuous), the magnitude of the tested regression coefficient, the degree of association between predictors, and the lower and upper limits on the total sample size.

## 1 Introduction

The sample size (SS) calculation is complicated by the presence of nuisance parameters. The values of these parameters are often estimated from external or internal pilot data which could substantially decrease the influence of erroneous assumptions on the values of these nuisance parameters.

In this manuscript we explore a “naive” approach to interim sample size re-estimation in linear and logistic regression models. This approach recalculates nuisance parameters at the interim analysis and updates the sample size bounded by the size of the internal pilot, $n$, and an upper bound, $N_{max}$, often chosen from budgetary or recruitment considerations. The benefit of sample size recalculation comes with a price, it inflates the type I error and power, and the final sample size becomes a random quantity.

We consider regression models of the form

$$E\left(Y|X_1, ..., X_p\right) = g^{-1}\left(\beta_0 + \beta_1 X_1 + .... + \beta_p X_p\right), \tag{1}$$

where the mean of an outcome $Y$ conditional on $X_1, ..., X_p$ is parameterized by a continuous monotone link function $g(\cdot)$ and a linear combination of regression coefficients $\beta_0, ..., \beta_p$ and $X_1, ..., X_p$. The logistic regression is defined by a LOGIT link function, $g(EY) = \ln\left(\frac{EY}{1-EY}\right)$, for a binary outcome. The gaussian linear regression is defined by the identity link function, $g(EY) = EY$, and a normal outcome.

Internal pilot designs have been addressed by many authors, see [10] and [2]. However, most research literature on internal pilot designs targets two group comparisons with random subject allocation. This focus comes from clinical trials, whereas the coverage of interim sample size recalculation for observational data is very sparse. A rare exception is the sample size re-estimation for a general linear model explored by [1]. An extension to linear mixed models was considered by [4]. A sample size recalculation for the analysis of covariance model (ANCOVA) was investigated by [3]. The above papers assume that model covariates are fixed quantities. The paper of [5] attempted to analyze the effect of random group allocation. Internal pilot based estimates of regression parameters and the distribution of covariates were used in [6] for resampling-based power estimation and consequent sample size estimation.

Literature suggests several approaches to sample size calculation for regression models. The parameters of interest in these approaches may be different.

A popular approach to sample size calculation was suggested in [9] where the effect size and sample size formula were taken from a single predictor regression model and adjusted for multiple predictors using the coefficient of multiple correlation between the predictor of interest and the other model predictors.

Shieh [11] suggested a sample size calculation formula for likelihood ratio tests in Generalized Linear Models (GLM). His focus on GLM and a regression coefficient as a parameter of interest coincides with our objective, however, his sample size formula was tailored to the likelihood ratio test whereas our main interest is on Wald test.

Sample size calculations often target the quality of outcome prediction, which is traditionally expressed by the coefficient of determination, $R^2$. Modified $R^2$ measures such as the adjusted $R^2$, see [7], or the generalized $R^2$, see [8], are the popular extensions of $R^2$ to non-linear models.

Our main interest lies in the analysis of association between $X_1$ and $Y$ adjusted for $X_2, ..., X_p$. The conditional mean model (1) is fitted to assess this adjusted association, quantified by $\beta_1$. Sample sizes for testing the hypothesis about $\beta_1$ are often based on asymptotic normality of $\hat{\beta}_1$, the Maximum Likelihood Estimator (MLE) of $\beta_1$. Asymptotic normality of $\hat{\beta}_1$ justifies Wald, score, and asymptotic likelihood ratio tests. Consequently, the sample size formula is based on a $z$-approximation for testing $H_0 : \beta_1 = 0$ versus some alternative $H_1 : \beta_1 = \delta$.

Even though, error minimization for regression modeling is not in the focus of this manuscript, we tied $R^2$ with the values of $\beta_1$ in simple linear regression models, and the generalized $R^2$ with $\beta_1$ in simple logistic regressions.

Section 2 presents sample size formulas for testing $H_0$ on a linear or logistic regression model with the Wald test. Section 3 presents the results of comprehensive Monte-Carlo study performed for various artificial scenarios. The manuscript is concludes with a short discussion in Section 5.

## 2 Sample size calculation for testing $H_0 : \beta_1 = 0$

We consider independent and identically distributed sampling when random variables $(Y_i, X_{i1}, ..., X_{ip})$, $i = 1, ..., n, ..., N, ...,$ are observed from a joint distribution $F_{Y,X_1,...,X_p}(y, x_1, ..., x_p)$. In this scheme, we denote the internal pilot data sample size by $n$ and the total sample size by $N$.

### Wald test

Regression models traditionally assume that the outcome $Y$ is a random variable whereas the covariates $X_1, ..., X_p$ are quantities fixed by design. The randomness of covariates is often ignored and the modelling process is focused on a conditional mean defined by a conditional distribution $F_{Y|X_1,...,X_p}(y|x_1, ..., x_p)$.

If a parametric family is assumed, then under some regularity conditions the limiting distribution of the maximum likelihood estimator (MLE), $\hat{\mathbf{b}}$, is normal,

$$\sqrt{n}\left(\hat{\mathbf{b}} - \mathbf{b}\right) \xrightarrow{d} N\left(0, \mathcal{I}^{-1}(\mathbf{b})\right), \tag{2}$$

where $\mathcal{I}(\mathbf{b})$ is the (expected) Fisher information matrix, $\mathbf{b} = (\beta_0, \beta_1, ..., \beta_p)$. The Fisher information is defined as

$$\begin{aligned} \mathcal{I}(\mathbf{b}) &= -E_{Y,X_1,...,X_p}\left(\frac{\partial^2 \log f_{Y,X_1,...,X_p}}{\partial\mathbf{b}\partial\mathbf{b}^T}\right) \qquad (3) \\ &= -E_{Y,X_1,...,X_p}\left(\frac{\partial^2 \log f_{X_1,...,X_p}}{\partial\mathbf{b}\partial\mathbf{b}^T} + \frac{\partial^2 \log f_{Y|X_1,...,X_p}}{\partial\mathbf{b}\partial\mathbf{b}^T}\right) \\ &= -E_{Y,X_1,...,X_p}\left(\frac{\partial^2 \log f_{Y|X_1,...,X_p}}{\partial\mathbf{b}\partial\mathbf{b}^T}\right), \qquad (4) \end{aligned}$$

where $f$ is a distribution (p.m.f. or p.d.f.) of a random variable defined by its subscript. The transition from (3) to (4) assumes that the distribution of covariates does not depend on $\mathbf{b}$. Thus, the difference with the Fisher information defined for a conditional on $X_1, ..., X_p$ model is that the dependence on $X_1, ..., X_p$ have to be integrated out.

Let $\mathcal{I}^{-1}(\beta_1)$ be the second diagonal element of $\mathcal{I}^{-1}(\mathbf{b})$, the asymptotic variance of $\hat{\beta}_1$, then

$$\sqrt{n}\left(\hat{\beta}_1 - \beta_1\right) \xrightarrow{d} N\left(0, \mathcal{I}^{-1}(\beta_1)\right), \tag{5}$$

and the sample size formula for the Wald test is

$$N = \mathcal{I}^{-1}(\beta_1)\frac{\left(z_{1-\alpha/2} + z_{1-\beta}\right)^2}{\delta^2}, \tag{6}$$

where $\alpha$ is the type I error, $\beta$ is the type II error where the absence of subscripts distinguishes it from the regression coefficients, $z_\tau$ is a $\tau$-quantile of standard normal distribution, $\delta$ is a (non-standardized) effect size.

Despite the simplicity of formula (6), we have not seen its use for sample size calculations in generalized linear models except for a multiple linear regression, see Section 2.

The practical use of (6) is complicated by the asymptotic nature of the test and the unknown $\mathcal{I}^{-1}(\beta_1)$. We use internal pilot data to estimate $\mathcal{I}^{-1}(\beta_1)$ and recalculate the sample size. Regular regression model output contains a table of regression coefficients with estimates of their standard errors. From this table, the standard error of the internal pilot based estimate $\hat{\beta}_1$ is $SE(\hat{\beta}_1)$. Then, we use $n\left(SE(\hat{\beta}_1)\right)^2$ to approximate $\mathcal{I}^{-1}(\beta_1)$ and calculate the total sample size. Then, the final formula for the total sample size is

$$N = n\left(SE(\hat{\beta}_1)\right)^2 \frac{\left(z_{1-\alpha/2} + z_{1-\beta}\right)^2}{\delta^2}. \tag{7}$$

## Multiple linear regression

For a linear regression model

$$Y_i = \beta_0 + \beta_1 X_{i1} + ... + \beta_p X_{ip} + \epsilon$$

with non-random covariates $X_{i1}, ..., X_{ip}$ and $\epsilon \sim N(0, \sigma^2)$, the Fisher information matrix for a single $i^{th}$ observation $(Y_i, X_{i1}, ..., X_{ip})$ is a symmetric $(p+1) \times (p+1)$ matrix

$$\mathcal{I}(\mathbf{b}|X_{i1}, ..., X_{ip}) = \frac{1}{\sigma^2}\begin{pmatrix} X_{i0}X_{i0} & X_{i1}X_{i0} & \cdots & X_{ip}X_{i0} \\ X_{i0}X_{i1} & X_{i1}X_{i1} & \cdots & X_{ip}X_{i1} \\ \cdots & \cdots & \cdots & \cdots \\ X_{i0}X_{ip} & X_{i1}X_{ip} & \cdots & X_{ip}X_{ip} \end{pmatrix},$$

where $X_{i0} = 1$. Denote $\mathbf{X}_{i\cdot} = (X_{i0}, ..., X_{ip})^T$, then $\mathcal{I}(\mathbf{b}|X_{i1}, ..., X_{ip}) = \sigma^{-2}\mathbf{X}_{i\cdot}\mathbf{X}_{i\cdot}^T$.

If $X_{i1}, ..., X_{ip}$ are random variables, we need to integrate their distribution out,

$$\mathcal{I}(\mathbf{b}) = E_{X_{i1},...,X_{ip}}\mathcal{I}(\mathbf{b}|X_{i1}, ..., X_{ip}) = \sigma^{-2}E\left(\mathbf{X}_{i\cdot}\mathbf{X}_{i\cdot}^T\right). \tag{8}$$

The matrix of second moments $E\left(\mathbf{X}_{i\cdot}\mathbf{X}_{i\cdot}^T\right)$ becomes the variance-covariance matrix of covariates when all covariates are centered, $E(X_{ij}) = 0$ $(j = 1, ..., p)$, except for the constant term $X_{i0} = 1$. The distribution of covariates in observational studies is often unknown. A simple solution is to plug-in internal pilot data based estimates of $\sigma^2$ and $E\left(\mathbf{X}_{i\cdot}\mathbf{X}_{i\cdot}^T\right)$ in Equation (8), then

$$\hat{\mathcal{I}}(\mathbf{b}) = \hat{\sigma}^{-2}n^{-1}\mathbf{X}_n\mathbf{X}_n^T,$$

where

$$\mathbf{X}_n = \begin{pmatrix} 1 & X_{11} & \cdots & X_{1p} \\ \cdots & \cdots & \cdots & \cdots \\ 1 & X_{n1} & \cdots & X_{np} \end{pmatrix}.$$

For a sufficiently large $n$, $\mathcal{I}^{-1}(\mathbf{b}) \approx n\hat{\sigma}^2\left(\mathbf{X}_n^T\mathbf{X}_n\right)^{-1}$, and

$$\sqrt{n}\left(\hat{\beta}_1 - \beta_1\right) \overset{d}{\approx} N\left(0, n\hat{\sigma}^2\left(\mathbf{X}_n^T\mathbf{X}_n\right)^{-1}_{(11)}\right),$$

where $\left(\mathbf{X}_n^T\mathbf{X}_n\right)^{-1}_{(11)}$ denotes the second diagonal element of $\left(\mathbf{X}_n^T\mathbf{X}_n\right)^{-1}$. For centered covariates the use of some matrix algebra leads to

$$n\left(\mathbf{X}_n^T\mathbf{X}_n\right)^{-1}_{(11)} = \hat{\sigma}_{X_1}^{-2}\left(1-\hat{r}^2_{X_1|X_2,\ldots,X_p}\right)^{-1},$$

where $\hat{\sigma}^2_{X_1}$ is the sample variance of $X_1$ and $\hat{r}^2_{X_1|X_2,\ldots,X_p}$ is the sample proportion of variance in $X_1$ explained by other covariates (the coefficient of multiple determination). Then,

$$N = \frac{\hat{\sigma}^2\left(z_{1-\alpha/2}+z_{1-\beta}\right)^2}{\delta^2\hat{\sigma}^2_{X_1}\left(1-\hat{r}^2_{X_1|X_2,\ldots,X_p}\right)} \tag{9}$$

is the pilot-data-based estimate of the sample size. We emphasize that (9) is just another form of formula (7) rewritten for multiple linear regression.

We also considered a "gold standard" sample size

$$N_{gold} = \frac{\sigma^2\left(z_{1-\alpha/2}+z_{1-\beta}\right)^2}{\delta^2\sigma^2_{X_1}\left(1-r^2_{X_1|X_2,\ldots,X_p}\right)} \tag{10}$$

calculated at known values of $\sigma$, $\sigma_{X_1}$, and $r_{X_1|X_2,\ldots,X_p}$. Obviously, the use of our "gold standard" is limited to simulation studies only, see the column "Target SS" in Tables 2-53.

## Multiple Logistic Regression

Multiple logistic regression is defined in Equation (1) by, $g(p_i) = \log\left(\frac{p_i}{1-p_i}\right)$, where

$$p_i = E\left(Y_i|X_{i1},...,X_{ip}\right) = \frac{\exp\left(\beta_0+\beta_1X_{i1}+...+\beta_pX_{ip}\right)}{1+\exp\left(\beta_0+\beta_1X_{i1}+...+\beta_pX_{ip}\right)}.$$

The Fisher information matrix for a given design $\mathbf{X}_n$ is

$$\mathcal{I}\left(\mathbf{b}|\mathbf{X}_n\right) = \mathbf{X}_n\mathrm{diag}\left(p_i\left(1-p_i\right)\right)\mathbf{X}_n^T.$$

When $\mathbf{X}_n$ is a random matrix with i.i.d. rows, the sample mean

$$\hat{\mathcal{I}}\left(\mathbf{b}\right) = n^{-1}\sum_{i=1}^{n}\mathcal{I}\left(\hat{\mathbf{b}}|\mathbf{X}_n\right) = n^{-1}\sum_{i=1}^{n}\mathbf{X}_n\mathrm{diag}\left(\hat{p}_i\left(1-\hat{p}_i\right)\right)\mathbf{X}_n^T$$

is a consistent estimate of $\mathcal{I}\left(\mathbf{b}\right) = E\left(\mathcal{I}\left(\mathbf{b}|\mathbf{X}_n\right)\right)$ by the law of large numbers, where $\hat{p}_i = \left(1+\exp\left(-\mathbf{X}_{i\cdot}^T\hat{\mathbf{b}}\right)\right)^{-1}$. Then, $\hat{\mathcal{I}}^{-1}\left(\beta_1\right)$, the second diagonal element of $\hat{\mathcal{I}}^{-1}\left(\mathbf{b}\right)$, is used in the sample size formula (6).

By analogy with the "gold standard" sample size formula (10) for linear gaussian regression, under known regression parameters and the distribution of covariates, we can find $\mathcal{I}^{-1}\left(\beta_1\right)$ to calculate the the exact sample size, the column "Target SS" in Tables 54-79.

The "gold standard" formulas are not perfect due to their asymptotic and not finite sample optimality, however, known Fisher information makes these formulas data independent. Thus, these "gold standard" formulas can be considered as the target quantities in sample size re-estimation.

# 3 Simulation studies

Each simulation experiment was based on 60000 repetitions, 50000 under $\beta_1 = 0$, and 10000 under the alternative $\beta_1 = \delta$. This secured the Monte-Carlo standard error of 0.001 for estimating the type I error and 0.004 for estimating statistical power.

## Covariates

Model covariates were generated from a multivariate normal, multinomial, or a mixture of normal distributions. To generate the distribution of $X_1, ..., X_p$ we first generated data from a multivariate normal distribution with zero mean, unit variances and an exchangeable correlation, $r$. Then, we categorized either $X_1$ into $B_1$ $(= sgn\,(X_1))$ and/or $X_j$ into $B_j$ $(= sgn\,(X_j);\ j = 2, ..., p)$ depending on a simulation experiment settings, where $sgn\,(\cdot)$ is equal to 1 for a positive argument and $-1$ for negative. For continuous predictors $r$ was equal to 0 (independence), 0.4 (weak correlation) or 0.8 (strong correlation). To secure the Person correlation of 0.4 (0.8) between two binary predictors, $B_j$ and $B_l$ the original correlation between $X_j$ and $X_l$, a.k.a. tetrachoric correlation, was set to 0.5878 (0.9511). To have the Pearson correlation of 0.4 (0.8) between a binary and a continuous predictors, say $X_j$ and $B_l$, the original correlation between $X_j$ and $X_l$ was set to 0.5037 (0.9991).

The extremely strong correlation between $X_j$ and $X_l$ created a substantial rate of perfect collinearity between $sgn(X_j)$ and $sgn(X_l)$ in our simulation studies with multiple predictors. To solve this we used backward elimination of linearly dependent columns from the design matrix (except for $X_1$) until a full rank design matrix was reached.

High collinearity creates a problem for estimating regression coefficients manifested in a higher variance.

In addition, the logistic regression periodically faced a complete separation problem, an common phenomenon in the presence of binary predictors and a rare, but possible, for continuous predictors. To avoid the separability issue we also performed backward elimination of linearly dependent columns until we reached full rank or until the design matrix had only the intercept column and $X_1$. If the rank of the two column matrix (intercept and $X_1$) was one, then we called the results of testing inconclusive and excluded the case from consideration. To limit the number of inconclusive results we decreased the correlation between $X_j$ and $X_l$ from 0.9991 to 0.9 in logistic regression analyses, which generated the Pearson correlation of 0.72 between $X_j$ and $B_l$.

## The Effect Sizes

To choose effect sizes, $\delta$, for our modelling, we targeted specific values of $R^2$. We considered $\beta_1 \in \{0.1429, 0.2294, 0.3333, 0.5, 1\}$ which defines $R^2 \in \{0.02, 0.05, 0.1, 0.2, 0.5\}$, respectively, in single predictor linear regression models. The definition of $R^2$ via the ratio of mean squared errors is not directly applicable to logistic regression models. However, a generalized $R^2$ can be used instead. We found that the generalized $R^2 \in \{0.02, 0.05, 0.1, 0.2\}$ for logistic regression model with a single continuous predictor is approximately reached at $\beta_1 \in \{0.291, 0.469, 0.702, 1.127\}$ at $n = 20$. The $\beta_1$ at the binary predictor leads to $\beta_1 \in \{0.286, 0.459, 0.667, 1.003\}$ at $n = 20$. As the sample size increases, the generalized $R^2$ converges to a constant, see Appendix. To keep track of the actual $R^2$, the columns $E\left(R^2\right)$ in Tables 54-67 reports the averaged generalized $R^2$.

## Sample size recalculation and exceptions

We use the word "exceptions" to identify situations where the total sample size cannot be calculated due to some reasons, or the calculated number is above high or below low sample size boundaries. Inability to make a decision with the final sample is also classified as an exceptional case.

When an exception happens, an alternative action should be applied to resolve the situation. The possibility of this alternative action must be taken into account when the type I error and power are evaluated.

We mainly observed exceptions within the logistic regression framework. A few examples of exceptions at the internal pilot stage, and their solutions as listed below:

- Perfect co-linearity. This mainly happens between strongly associated categorical predictors.
  - *Solution:* Eliminate co-linear predictors from the model.
  - *Remark:* We use backward elimination, always keeping $X_1$ in the model.
- High co-linearity.
  - *Solution:* When a standard error of $\hat{\beta}_1$ is above 100 we eliminate a predictor strongly associated with $X_1$.
  - *Remark:* This is not exactly an exception, however, high co-linearity often leads to high standard errors, and thus to a sample size hitting the upper bound.
- All values of $X_1$ or all values of $Y$ are the same.
  - *Solution:* We set $N = N_{max}$.
  - *Remark:* This is one of the least informative situation on the total sample size.
- The separation or quasi-separation problem between $Y$ and $X_1$ in logistic regression models
  - *Solution:* We set $N = N_{max}$.
  - *Remark:* This decision, however, is not obvious. If, for example, all $Y_i = X_{i1}$, we may see overwhelming evidence of association between $Y$ and $X_1$ given $P(Y)$ is bounded away from zero and one. The same situation ($Y_i = X_{i1}$) with a few zeros or ones is an indication of an insufficient sample size.

Table 1: Table numbers classified by model, the number and type of predictors, Pearson correlation ($r$), $B_i = I(X_i > 0)$

| Predictors | $r$ | Linear $N_{max} = 300$ | Linear $N_{max} = 600$ | Logistic $N_{max} = 600$ |
|---|---|---|---|---|
| $X_1$ | 0 | 2 | 28 | 54 |
| $B_1$ | 0 | 3 | 29 | 55 |
| $X_1, X_2$ | 0 | 4 | 30 | 56 |
| $X_1, X_2$ | 0.4 | 5 | 31 | 57 |
| $X_1, X_2$ | 0.8 | 6 | 32 | 58 |
| $B_1, B_2$ | 0 | 7 | 33 | 59 |
| $B_1, B_2$ | 0.4 | 8 | 34 | 60 |
| $B_1, B_2$ | 0.8 | 9 | 35 | 61 |
| $B_1, X_2$ | 0 | 10 | 36 | 62 |
| $B_1, X_2$ | 0.4 | 11 | 37 | 63 |
| $B_1, X_2$ | 0.8 | 12 | 38 | 64 |
| $X_1, B_2$ | 0 | 13 | 39 | 65 |
| $X_1, B_2$ | 0.4 | 14 | 40 | 66 |
| $X_1, B_2$ | 0.8 | 15 | 41 | 67 |
| $X_1, X_2, ..., X_{10}$ | 0 | 16 | 42 | 68 |
| $X_1, X_2, ..., X_{10}$ | 0.4 | 17 | 43 | 69 |
| $X_1, X_2, ..., X_{10}$ | 0.8 | 18 | 44 | 70 |
| $B_1, B_2, ..., B_{10}$ | 0 | 19 | 45 | 71 |
| $B_1, B_2, ..., B_{10}$ | 0.4 | 20 | 46 | 72 |
| $B_1, B_2, ..., B_{10}$ | 0.8 | 21 | 47 | 73 |
| $B_1, X_2, ..., X_{10}$ | 0 | 22 | 48 | 74 |
| $B_1, X_2, ..., X_{10}$ | 0.4 | 23 | 49 | 75 |
| $B_1, X_2, ..., X_{10}$ | – | 24 ($r = 0.8$) | 50 ($r = 0.8$) | 76 ($r = 0.72$) |
| $X_1, B_2, ..., B_{10}$ | 0 | 25 | 51 | 77 |
| $X_1, B_2, ..., B_{10}$ | 0.4 | 26 | 52 | 78 |
| $X_1, B_2, ..., B_{10}$ | – | 27 ($r = 0.8$) | 53 ($r = 0.8$) | 79 ($r = 0.72$) |

- The sample size formula is applicable but the calculated sample size does not belong to the sample size range.
  - *Solution:* If $N < n_1$, then $N = n_1$. If $N > N_{max}$, then $N = N_{max}$.

## Other settings

We explored three possibilities for the internal pilot sample size, $n \in \{20, 50, 100\}$. All results of our simulation experiments are reported in Tables 2-27 (Linear regression with $N_{max} = 300$), Tables 28-53 (Linear regression with $N_{max} = 600$), and Tables 54-79 (Logistic regression with $N_{max} = 600$) with 12 to 15 experiments per table. The information on the type I error and power for multiple regression models are also summarized in Figures 1-10.

For convenience the summary of all Tables is presented in Table 1.

# 4 Empirical Conclusions

The Monte-Carlo Error for type I error assessments is 0.001, the Monte-Carlos error for power estimation was 0.004.

## Linear regression

In linear models, the type I error often stayed close to the desired level, which is consistent with previously published simulation results that used single predictor models. However, there were a few situations where the type I error was elevated.

**Single predictor:** Occasionally, the type I error was increased to 5.5% − 5.7% when the internal pilot was relatively small, $n = 20$, and the "gold standard" sample size was not high as well.

Table 2 shows that the "Target SS" is 70.68 for detecting $\beta_1 = 0.3333$, whereas actual total sample size had a mean of 83.46 with the SD of 41.33. The type I error for this model was 0.0547. Periodically, the total sample size falls below $n$, which as we see in all our simulations one of main contributor to the type I inflation. This problem increased at $\beta_1 = 0.5$, the type I error was estimated as 5.66%. A surprising effect shows up when the "Target SS" is equal to 7.86, which comes from a large effect size, $\beta_1 = 1$. This case had $EN = 20.18$ and $SD(EN) = 1.39$, which means that with a rare exception our total sample size was equal to 20. This makes this scenario almost indistinguishable from the fixed sample size, and we are obviously overpowered in this case. The type I error was well controlled, the estimated type I error was 0.0494.

Similar inference is applicable for $n = 50$, but inflation of the type I error is present at a lower level, 0.0528 (at $\beta_1 = 0.05$) and 0.0523 ($\beta_1 = 0.0523$).

A linear regression model with a binary predictor, see Table 3, is slightly more sensitive to the type I error inflation, the type I error at $n = 20$ is 0.0563 ($\beta_1 = 0.2294$) and 0.0586 ($\beta_1 = 0.3333$).

The upper bound for the total sample size for Tables 2 and 3 was equal to 300. This created situations with substantial underpowering. At $n = 20$ and $\beta_1 = 0.1429$, the power for a continuous predictor was only 66% and 67.6% for a binary one. The increase in the internal pilot to $n = 100$ provided a closer to the gold standard estimate of the total sample size, but did not resolve underpowering, we reached only 68.9% and 69.1% power for continuous and binary predictors, respectively.

Tables 28 and 28 report single predictors models for continuous and binary predictors, respectively, when the upper bound for the total sample size was 600. The conclusions about the type I error are essentially the same. The difference only in power properties. The upper bound of 600 observations was less likely to be reached and power less then 80% was seen only two times. For the continuous predictor the lowest power was observed at $n = 20$ and $\beta_1 = 0.1429$, the same scenario for a binary predictor had power of 78.3%.

Overall, we observed underpowering when there was a substantial chance that the final sample size reached the upper bound. Similarly, overpowering could happen when the distribution of the final sample size "spilled over" the lower bound.

**Multiple predictors:** The inflation of the type I error at $n = 20$ increases with increased number of predictors, see Tables 4-21. The worst type I error we have seen is 0.0662, Table 24, at highly correlated predictors, $B_1, X_2, ..., X_{10}$. The binary predictor of interest, $B_1$, was also associated with an increased type I error when compared to a continuous predictor $X_1$. As the size of internal pilot increases the inflation of type I error becomes less and less visible.

For illustrative convenience we also reported the type I error and power in eight figures, Figures 1 - 4 (linear regressions) and 7 - 10 (logistic regressions).

The type I errors for linear regression models with multiple predictors are reported in Figures 1 ($X_1$ is binary) and 2 ($X_1$ is continuous).

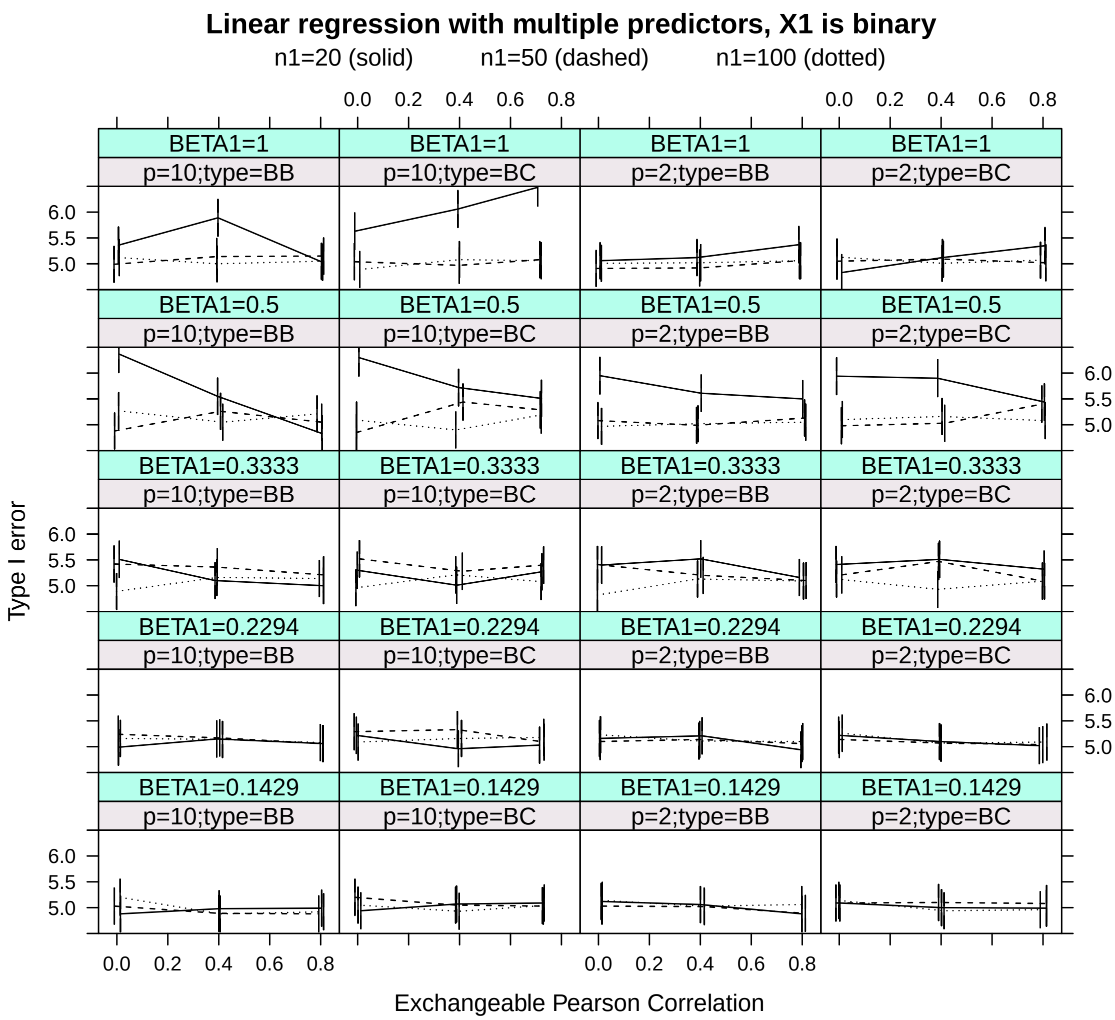


Figure 1: Type I error for testing a binary predictor of interest in multiple linear regression; $p$ is the number of predictors; BB, all binary predictors; BC, $X_1$ is binary and the rest are continuous predictors

Power for linear regression models with multiple predictors are reported in Figures 3 ($X_1$ is binary) and 4 ($X_1$ is continuous).

For large effect sizes, when the target sample size is close to $n$ we often see power above the desired level. On the other hand, the sample sizes hitting the upper bound, $N_{max}$, lead to underpowered studies.

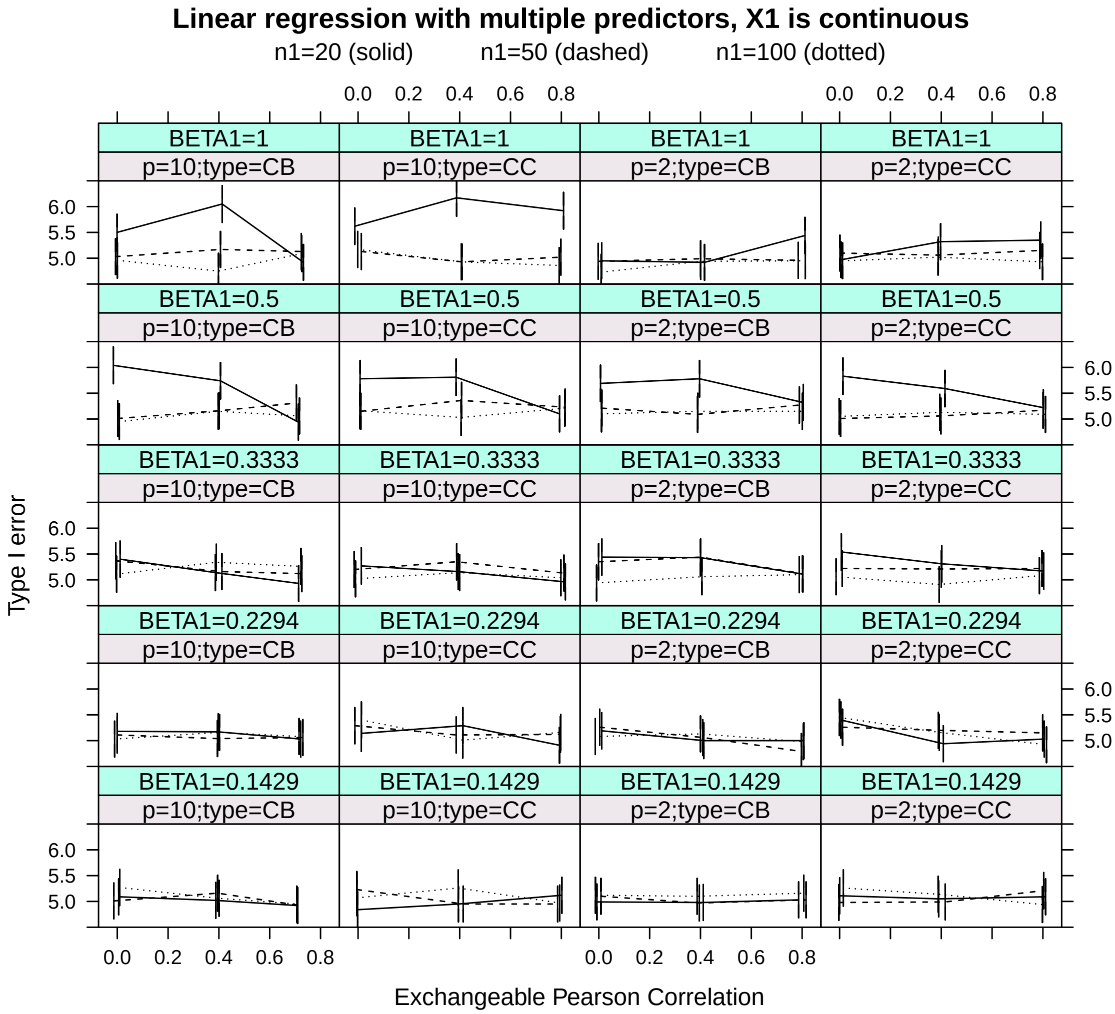


Figure 2: Type I error for testing a continuous predictor of interest in multiple linear regression; $p$ is the number of predictors; CB, $X_1$ is continuous and the rest are binary predictors; CC, all continuous predictors

The desired power is approximately achieved when the chances for $N$ to be below $n$ and above $N_{max}$ are low. This is seen clearly for $p = 2$ at $\beta_1 = 0.2294$.

## Logistic regression models

**Single predictor:** Single predictor models are reported in Tables 54 and 55 for continuous and binary predictors. Surprisingly the type I error was mainly below the designed 5%.

The lowest type I error for a single continuous predictor was 2.92% for $n = 20$ with $\beta_1 = 1.127$. In this scenario $E(N) = 39.21$ and $SD(N) = 26.9$, which means we observed a few $N$ below $n = 20$. This effect was also associate with increased power, 84.4%. A good control of type I error and power was reached at $n = 100$ with $\beta_1 =\in \{0.291, 0.469\}$. These cases secured the type I error of 4.67% and 4.48%, and the power of 81.3% and 80.2%.

The smallest observed type I error for single binary predictor models was 3.7%, at $n = 20$

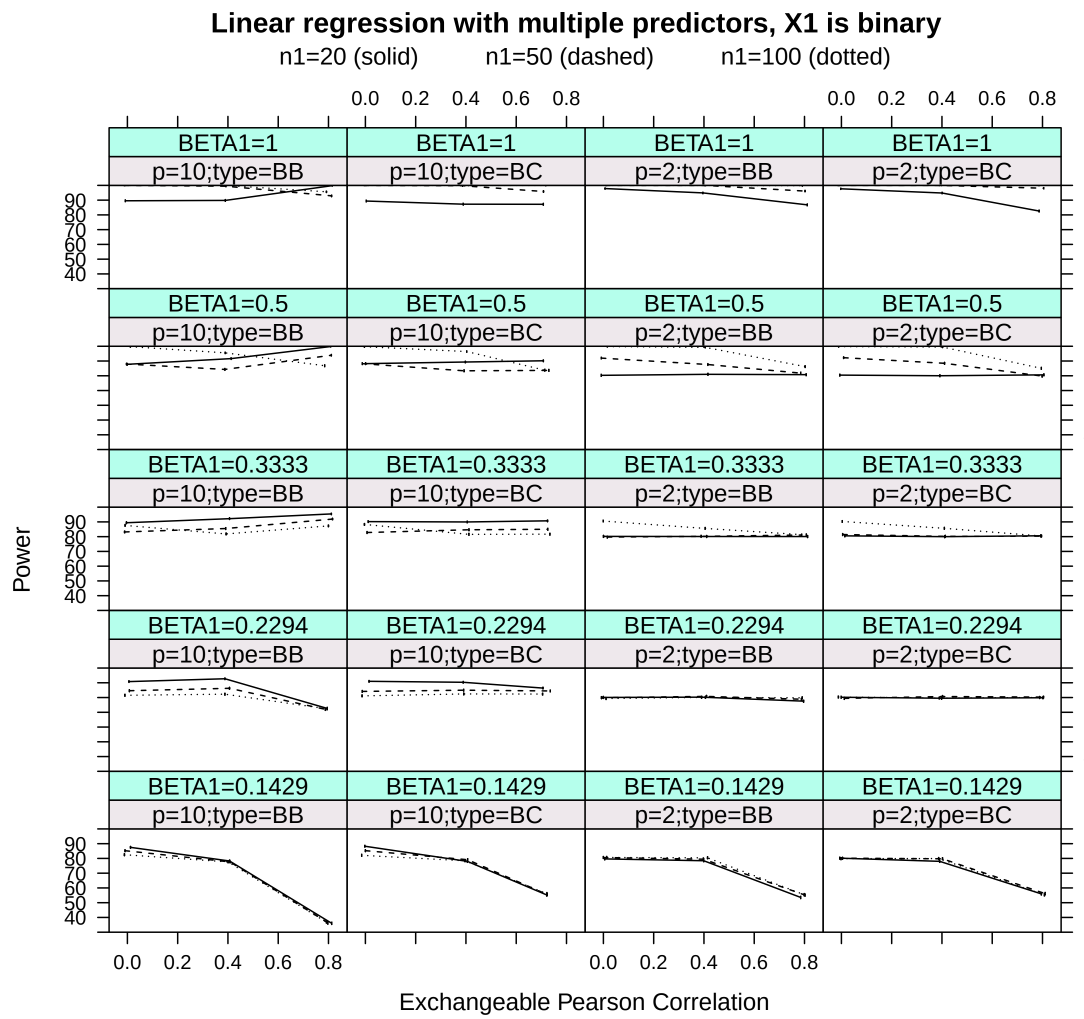


Figure 3: Power for testing a binary predictor of interest in multiple linear regression; $p$ is the number of predictors; BB, all binary predictors; BC, $X_1$ is binary and the rest are continuous predictors

and $\beta_1 = 1$. The power for this case was 83.7%. Good type I error and power were reached at $n = 100$ with $\beta_1 = \in \{0.286, 0.459\}$.

**Toy Example:** In most cases the type I and II errors are being inflated when the distribution of the total sample size, $N$, hits upper andor lower lower bounds. Figues 5 and 6 show, respectively, a good and a bad scenarios for interim sample size recalculations for logistic regression model.

The absence of exceptions along with the distribution of the total sample size fully contained between the $n$ and $N_{max}$ represent the most favorable scenario for interim sample size recalculation.

The bad scenario had two highly collinear binary predictors and a strong effect size. In our simulations, strong collinearity (when detected) was resolved by eliminating collinear predictors. The separability cases were addressed by administratively setting $N = N_{max}$. On the other hand, the high value of $\beta_1$ often led to $N < n_1$, which was resolved by $N = n_1$.

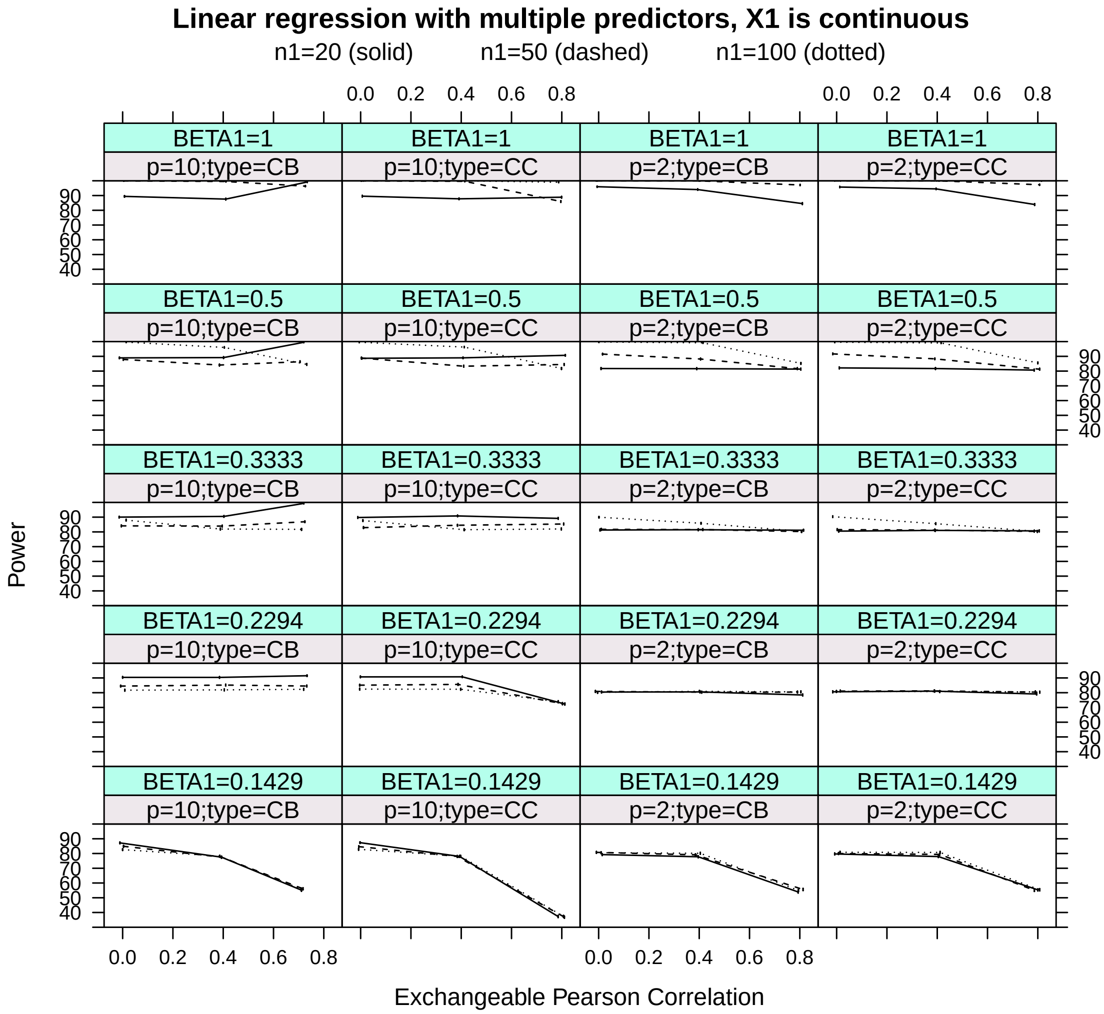


Figure 4: Power for testing a continuous predictor of interest in multiple linear regression; $p$ is the number of predictors; CB, $X_1$ is continuous and the rest are binary predictors; CC, all continuous predictors

Figure 6 reports one of worst situations we observed in our simulation studies.

**Multiple predictors:** The type I error for logistic regression models with multiple predictors are summarized in Figures 7 ($X_1$ is binary) and 8 ($X_1$ is continuous).

Figure 7 shows a surprising effect on type I error. It becomes much lower than 5% for highly correlated binary predictors at high values of $\beta_1$. The toy example (Figures 5 and 6) clearly shows that this comes from the high rate of exceptions. Simiularly we observed excessive rates of exceptions for the cases of continuous $X_1$ with high $\beta_1$ and small $n_1$.

Power for logistic regression models with multiple predictors are reported in Figures 9 ($X_1$ is binary) and 10 ($X_1$ is continuous).

Power properties were reduced for cases with highly collinear binary predictors. This collinearity was often detected, $B_2$ was removed from the model, which consequently led to a biased (downward) estimate of the variance of $\hat{\beta}_1$ and to a lower final sample size.

We also observed lower power for the cases with a low effect size ($\beta_1 = 0.286$). These

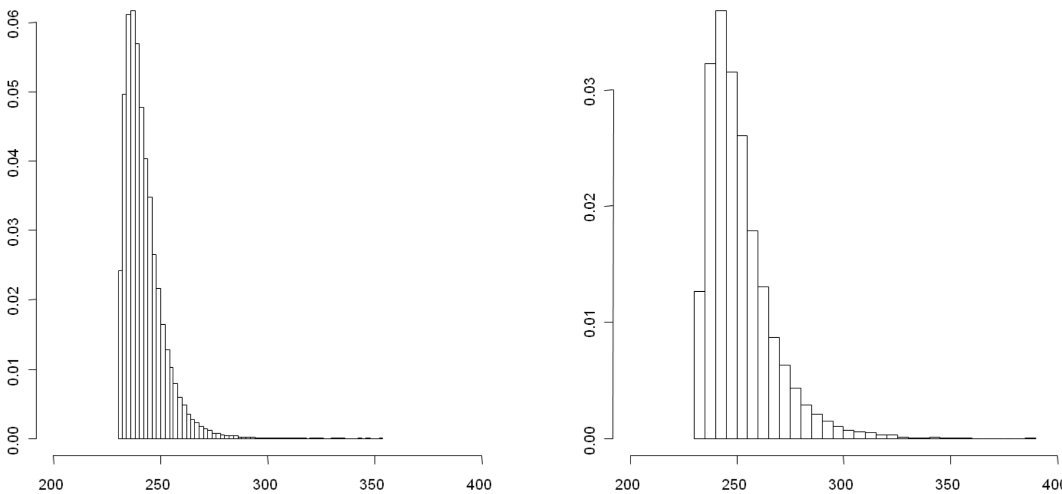


Figure 5: Logistic regression, two binary predictors, $n_1 = 100$, $\delta = 0.37$, $r = 0$, type I error = 4.9%, power=81.8%. Left panel presented the final sample size distribution under the null, the right panel is under the alternative hypothesis.

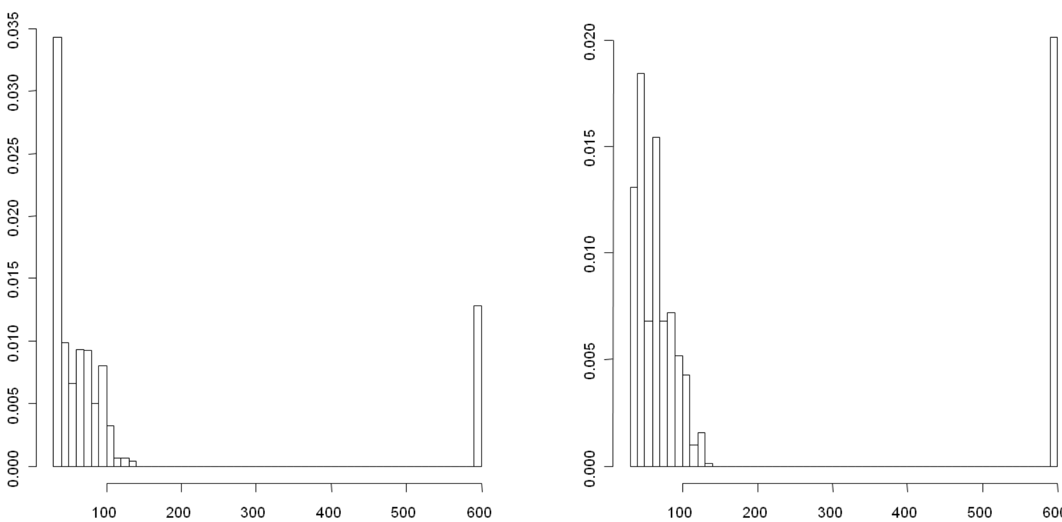


Figure 6: Logistic regression, two binary predictors, $n_1 = 20$, $\delta = 1$, $r = 0.8$, type I error = 1.3%, power=54.0%. Left panel presented the final sample size distribution under the null, the right panel is under the alternative hypothesis.

cases were mainly driven by the distribution of $N$ and how often it hit $N_{max}$.

# 5 Discussion

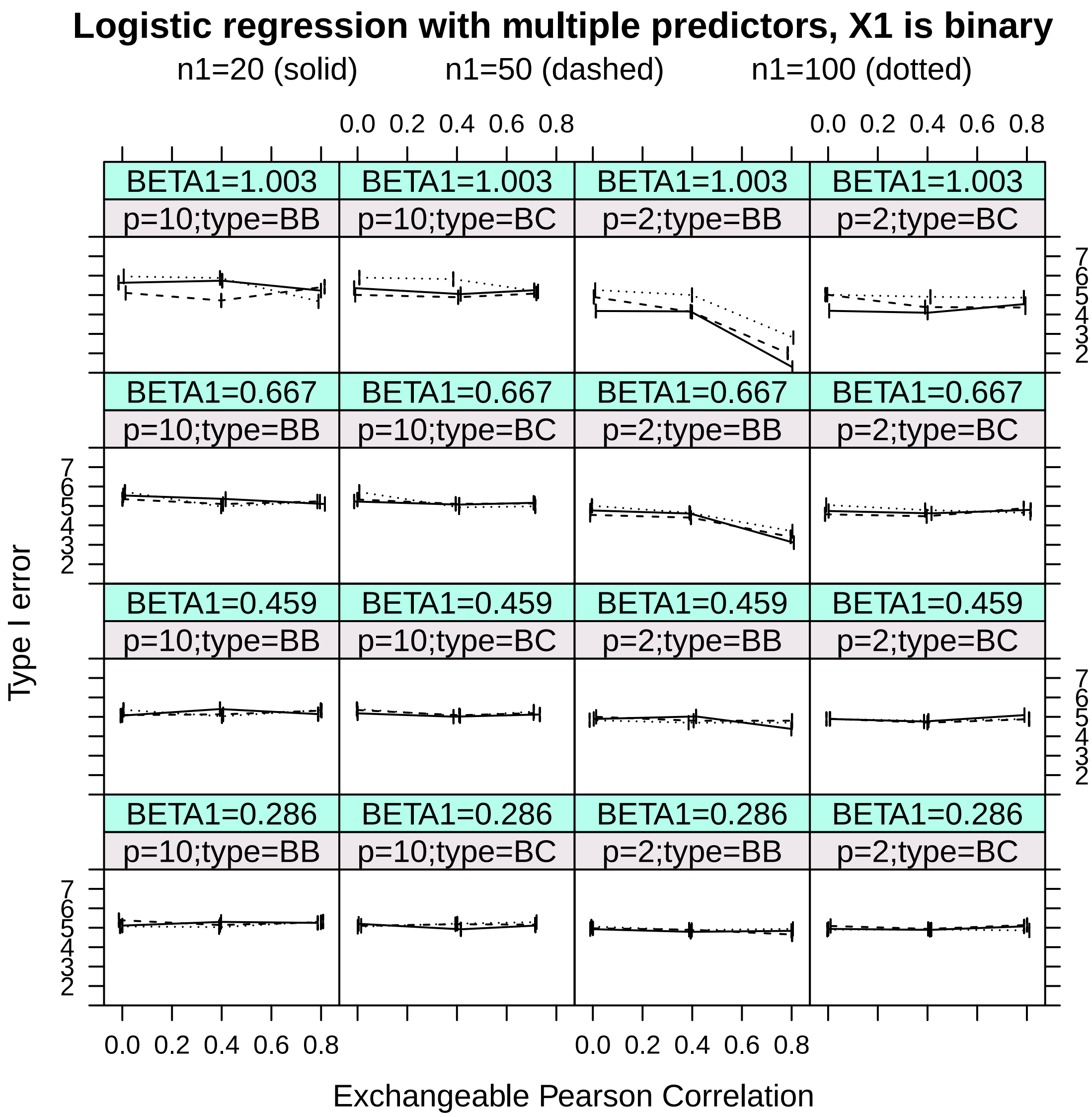


Figure 7: Type I error for testing a binary predictor of interest in multiple logistic regression; $p$ is the number of predictors; BB, all binary predictors; BC, $X_1$ is binary and the rest are continuous predictors

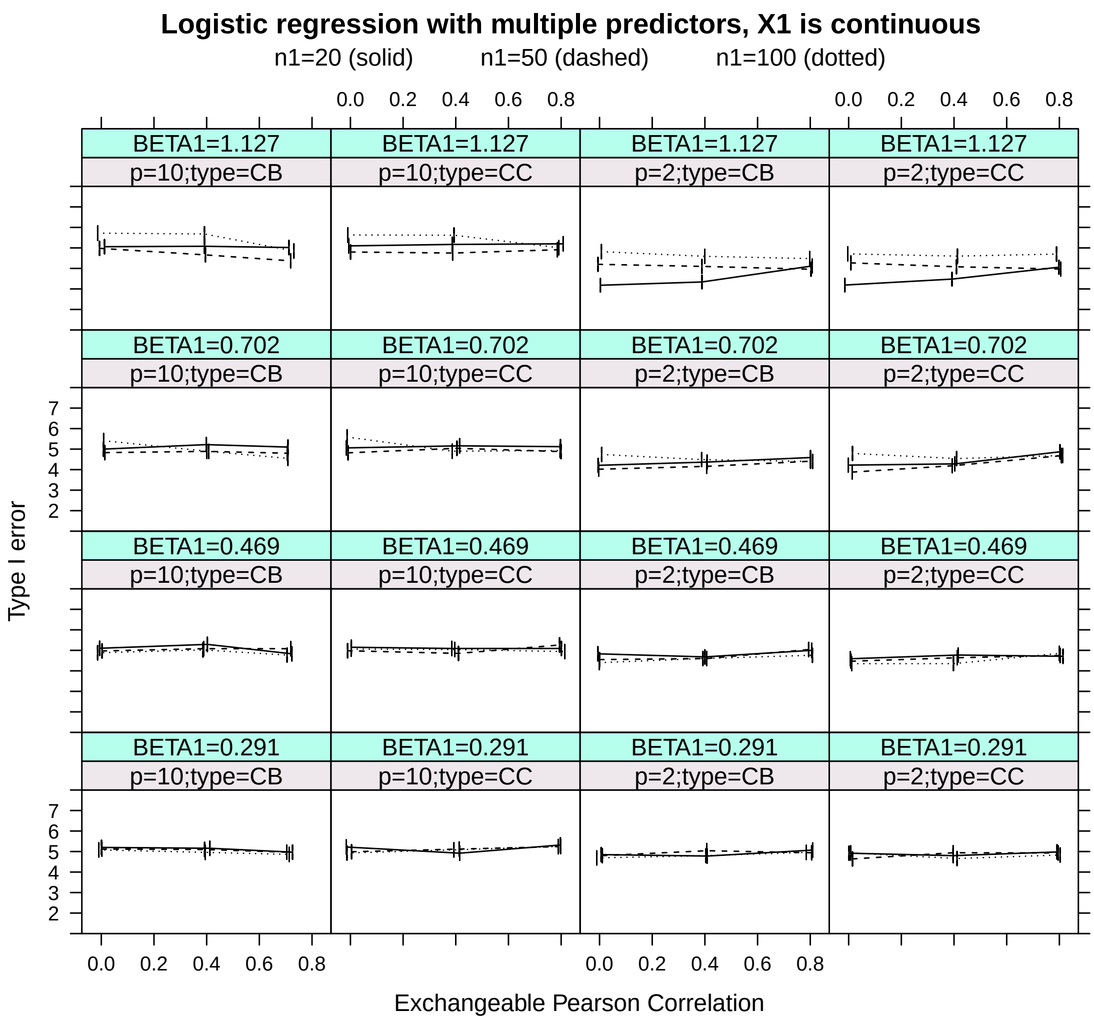


Figure 8: Type I error for testing a continuous predictor of interest in multiple logistic regression; $p$ is the number of predictors; CB, $X_1$ is continuous and the rest are binary predictors; CC, all continuous predictors

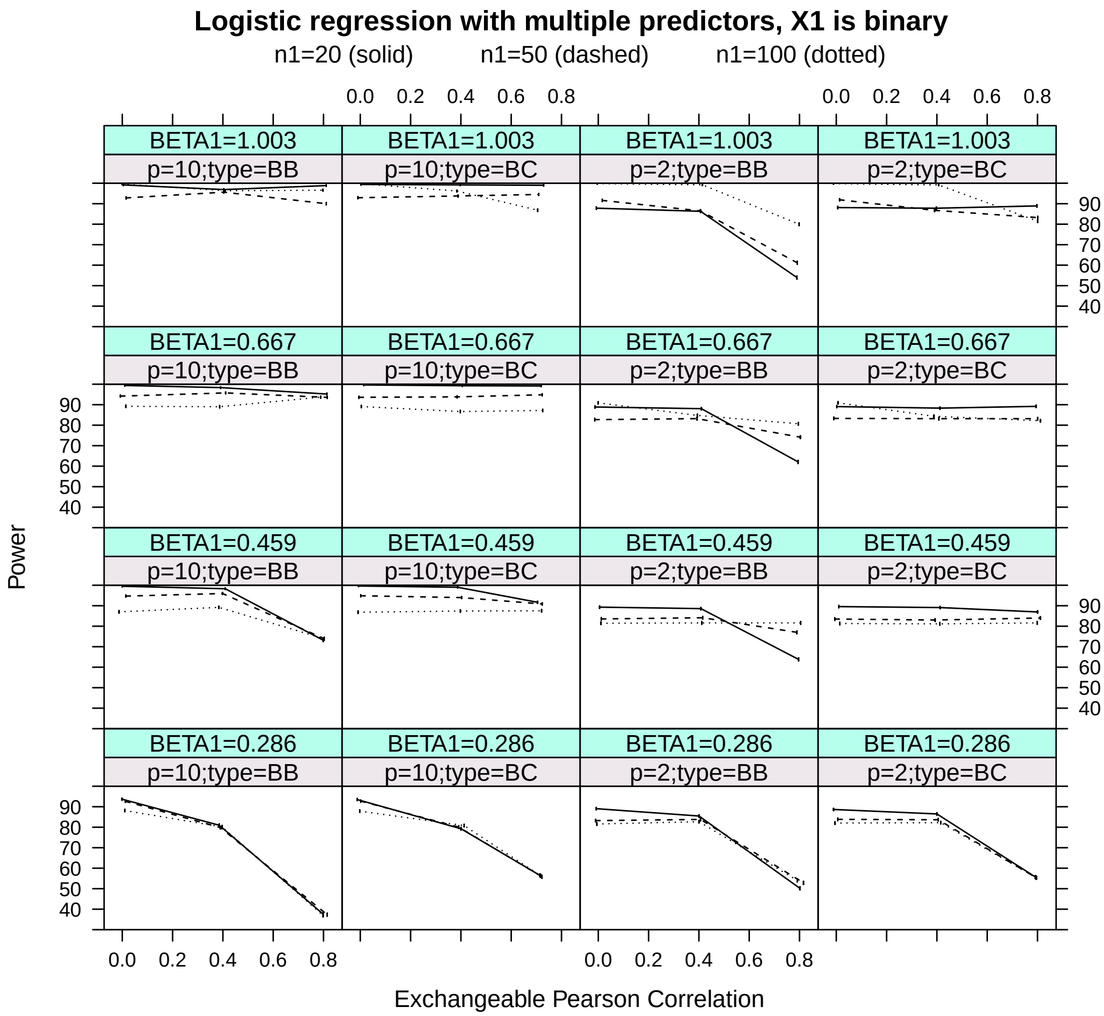


Figure 9: Power for testing a binary predictor of interest in multiple logistic regression; $p$ is the number of predictors; BB, all binary predictors; BC, $X_1$ is binary and the rest are continuous predictors

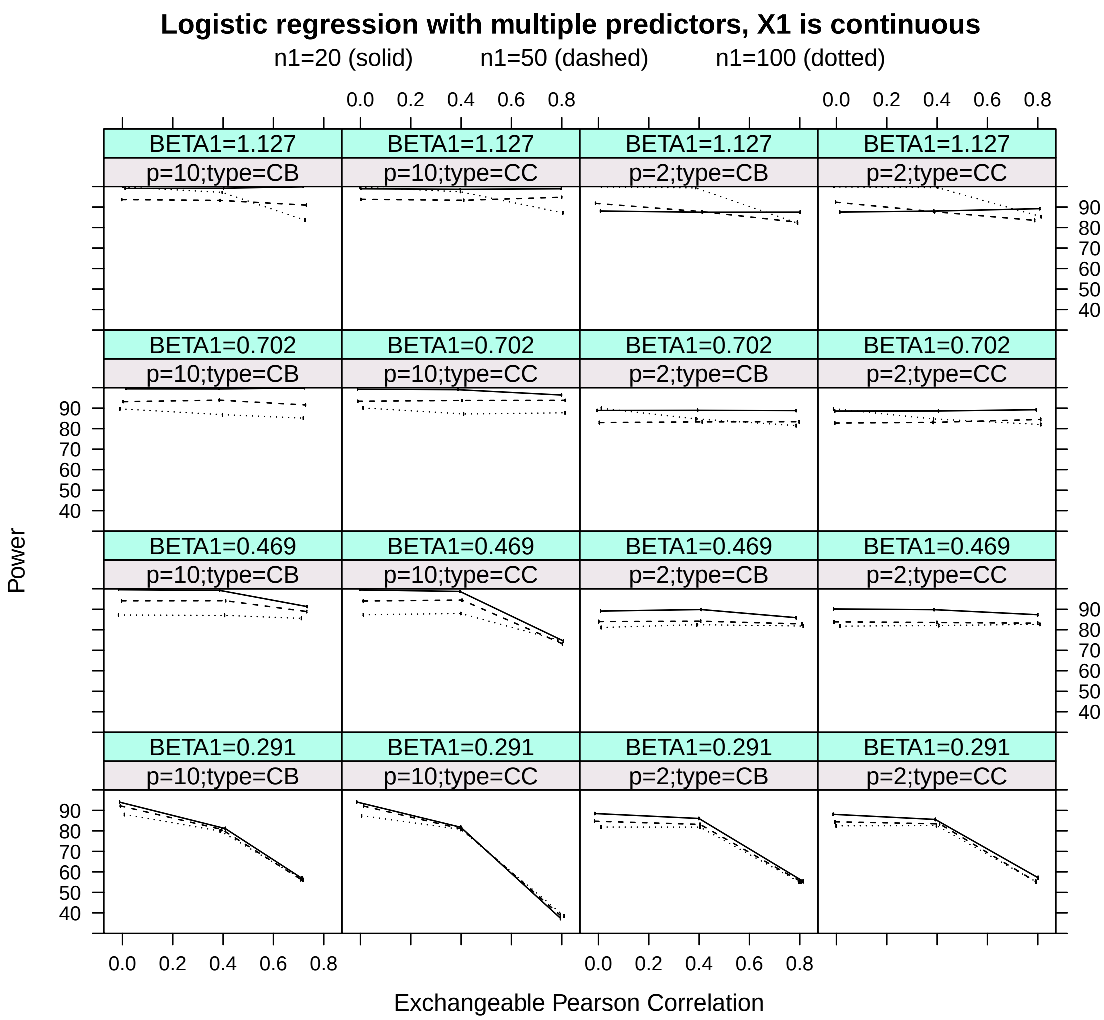


Figure 10: Power for testing a continuous predictor of interest in multiple logistic regression; $p$ is the number of predictors; CB, $X_1$ is continuous and the rest are binary predictors; CC, all continuous predictors

Naive internal pilot designs are commonly used in practice. These designs update the total sample size using new values of nuisance parameters at the interim analysis. These methods, however, use unadjusted sample size formulas and test statistics. This naive approach to sample size recalculation is known to inflate the type I and II errors. Meanwhile, a few statistical methods suggest various remedies for a better control of the size of the test and its power properties. We call these adjusted internal pilot designs as non-naive internal pilots designs.

Despite the existence of substantial amount of statistical literature on non-naive internal pilot designs, we did not see enough evidence to justify the use of non-naive internal pilot designs.

In this work, we did not make any modifications to the naive internal pilot designs. We investigated how much the type I and II errors are inflated under various scenarios with the use of the ordinary linear and logistic regression models as the primary analytic tools.

The detailed analysis is presented in Section 4. Overall our recommendations are the following.

- **Avoid small sizes for internal pilot designs.** We have seen substantial inflations and deflations for the size of the test. The type I error occasionally increased up to 6.62% in linear regression and decreased down to 1.3% in some instances of logistic regression model. When the internal pilot sample size increases, the type I error variations (inflation or deflation) becomes much smaller, which decreases the need for non-naive internal pilot designs.

- **Avoid collinearity.** Categorical predictors may generate perfectly collinear design matrices. The chance of observing collinearity increases with small sample sizes and for the cases or highly associated predictors. In the case of perfect collinearity we have to resolve the issue administratively, which may create a biased estimate of the total sample size. In a similar manner, high collinearity between predictors increases standard errors and leads to a skyrocketing sample size.

- **Avoid inadequate restrictions on the total sample size.** The distribution of the total sample size is a random quantity, and if it hits lower or upper bounds the type I and II errors are affected.

- **Use large internal pilots for logistic regression modelling of rare events.** Rare events lead to high variability in regression parameter estimation, elevated chances for quasi-separability and separability, and there is a substantial probability that no events are observed in the internal pilot.

We found that the recipe for a well control of type I and II errors in naive internal pilot designs include (1) a sufficiently large internal pilot sample size, (2) the distribution support of the total sample size ($N$) fully located between the lower ($n_1$) and upper ($N_{max}$) bounds, and (3) very low probability exceptions.

The chance that the distribution of $N$ hits the lower bound $n_1$ is of a lesser concerns if the internal pilot is sufficiently large. The type I error inflation decreases with the increase of $n_1$.

Overall, naive internal pilot designs are useful and legitimate way for sample size recalculation provided that the aforementioned pitfalls are avoided.

# Appendix

The generalized $R^2_G$ (see [8]) is defined by

$$R^2_G = 1 - \left(\frac{L(0)}{L(\widehat{\beta})}\right)^{2/n}, \tag{11}$$

where $L(0)$ is the likelihood of the intercept-only model, $L(\hat{\beta})$ is the likelihood based on the estimated model parameters. We consider a simple single predictor case when $\beta = (\beta_0, \beta_1)$ and substitute $L(\hat{\beta})$ with $L(\beta)$ in Equation 11. Then, we consider

$$R^2_{\mathbf{X}} = 1 - \left(\frac{L(0)}{L(\beta)}\right)^{2/n}, \tag{12}$$

where the subscript $\mathbf{X}$ highlights the dependence on the design matrix. A bit of algebra leads to

$$\ln(1 - R^2_{\mathbf{X}}) = \frac{2}{n}\left\{\ln L(0) - \ln L(\widehat{\beta})\right\} = \frac{1}{n}\left(-2\ln\frac{L(\hat{\beta})}{L(0)}\right).$$

For an i.i.d. sample $(Y_1, X_1), ..., (Y_n, X_n)$ the log-likelihood conditional on $X_i$ is

$$\ln L(\beta) = \sum_{i=1}^{n} \ln f(Y_i|\beta, X_i)$$

and

$$\begin{aligned} \ln(1 - R^2_{\mathbf{X}}) &= \frac{1}{n}\sum_{i=1}^{n}\left\{-2\ln\frac{f(Y_i|\hat{\beta}, X_i)}{f(Y_i|0)}\right\} \\ &= -2\int \ln\frac{f(y|\hat{\beta}, x)}{f(y|0, x)} dP_n(x, y), \end{aligned} \tag{13}$$

where $P_n(x, y)$ is the empirical measure with $n^{-1}$ weights on $(Y_i, X_i)$. Under some regularity and properties of empirical processes, as $n \to \infty$,

$$\ln(1 - R^2_{\mathbf{X}}) \to -2\int \ln\frac{f(y|\beta, x)}{f(y|0, x)} f(y|\beta, x)dy.$$

Thus, the generalized $R^2$ asymptotically is independent of the sample size.

To establish connection between (12) and $\beta_1$ we used numeric approximation and found that the generalized $R^2 \in \{0.02, 0.05, 0.1, 0.2\}$ is approximately reached at $\beta_1 \in \{0.291, 0.469, 0.702, 1.127\}$, respectively at $n = 20$. For larger sample sizes and the same $\beta_1$, the generalized $R^2$ become

Table 2: Linear regression with a single continuous predictor, $N_{max} = 300$.

| n | $\beta_1$ | $R^2$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *Target SS* |
|---|---|---|---|---|---|---|---|
| 20 | 0.1429 | 0.02 | 0.0505 | 0.6600 | 281.50 | 39.78 | 385.11 |
| | 0.2294 | 0.05 | 0.0504 | 0.7931 | 168.37 | 69.37 | 149.21 |
| | 0.3333 | 0.1 | 0.0547 | 0.7988 | 83.46 | 41.33 | 70.68 |
| | 0.5 | 0.2 | 0.0566 | 0.8170 | 38.21 | 18.49 | 31.36 |
| | 1 | 0.5 | 0.0494 | 0.9674 | 20.18 | 1.39 | 7.86 |
| 50 | 0.1429 | 0.02 | 0.0503 | 0.6831 | 292.65 | 20.67 | 385.11 |
| | 0.2294 | 0.05 | 0.0528 | 0.7935 | 159.25 | 45.81 | 149.21 |
| | 0.3333 | 0.1 | 0.0523 | 0.8106 | 76.08 | 21.24 | 70.68 |
| | 0.5 | 0.2 | 0.0501 | 0.9194 | 50.48 | 2.59 | 31.36 |
| | 1 | 0.5 | 0.0508 | 1 | 50 | 0 | 7.86 |
| 100 | 0.1429 | 0.02 | 0.0501 | 0.6892 | 297.41 | 10.35 | 385.11 |
| | 0.2294 | 0.05 | 0.0503 | 0.8074 | 154.56 | 31.08 | 149.21 |
| | 0.3333 | 0.1 | 0.0484 | 0.8988 | 100.46 | 2.80 | 70.68 |
| | 0.5 | 0.2 | 0.0488 | 0.9980 | 100 | 0 | 31.36 |
| | 1 | 0.5 | 0.0504 | 1 | 100 | 0 | 7.86 |

a little bit higher. This increase in $R^2$ is not substantial, which means that the new values of $\beta_1$ for other sample sizes are not needed.

Table 3: Linear regression with a single binary predictor, $N_{max} = 300$.

| n | $\beta_1$ | $R^2$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *Target SS* |
|---|---|---|---|---|---|---|---|
| 20 | 0.1429 | 0.02 | 0.0502 | 0.6757 | 286.84 | 31.70 | 384.36 |
| | 0.2294 | 0.05 | 0.0508 | 0.7833 | 157.06 | 52.59 | 149.15 |
| | 0.3333 | 0.1 | 0.0563 | 0.7902 | 75.26 | 26.14 | 70.65 |
| | 0.5 | 0.2 | 0.0586 | 0.7956 | 34.10 | 11.25 | 31.40 |
| | 1 | 0.5 | 0.0496 | 0.9825 | 20.01 | 1.20 | 7.85 |
| 50 | 0.1429 | 0.02 | 0.0515 | 0.6892 | 296.46 | 12.94 | 384.36 |
| | 0.2294 | 0.05 | 0.0519 | 0.7980 | 153.03 | 31.55 | 149.15 |
| | 0.3333 | 0.1 | 0.0544 | 0.7986 | 72.85 | 14.53 | 70.65 |
| | 0.5 | 0.2 | 0.0506 | 0.9306 | 50.03 | 0.44 | 31.40 |
| | 1 | 0.5 | 0.0515 | 1 | 50 | 0 | 7.85 |
| 100 | 0.1429 | 0.02 | 0.0502 | 0.6916 | 299.19 | 5.02 | 384.36 |
| | 0.2294 | 0.05 | 0.0518 | 0.7970 | 151.20 | 21.64 | 149.15 |
| | 0.3333 | 0.1 | 0.0497 | 0.9111 | 100.03 | 0.51 | 70.65 |
| | 0.5 | 0.2 | 0.0486 | 0.9986 | 100 | 0 | 31.40 |
| | 1 | 0.5 | 0.0509 | 1 | 100 | 0 | 7.85 |

Table 4: Linear regression with two independent continuous predictors, $N_{max} = 300$.

| n | $\beta_1$ | $R^2$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *Target SS* |
|---|---|---|---|---|---|---|---|
| 20 | 0.1429 | 0.02 | 0.0521 | 0.6674 | 283.91 | 37.52 | 384.58 |
| | 0.2294 | 0.05 | 0.0512 | 0.8022 | 176.83 | 71.56 | 149.03 |
| | 0.3333 | 0.1 | 0.0535 | 0.8094 | 88.57 | 45.11 | 70.69 |
| | 0.5 | 0.2 | 0.0579 | 0.8283 | 40.20 | 20.00 | 31.34 |
| | 1 | 0.5 | 0.0496 | 0.9596 | 20.26 | 1.70 | 7.86 |
| 50 | 0.1429 | 0.02 | 0.0497 | 0.6799 | 293.41 | 19.68 | 384.58 |
| | 0.2294 | 0.05 | 0.0516 | 0.8175 | 162.30 | 47.08 | 149.03 |
| | 0.3333 | 0.1 | 0.0528 | 0.8116 | 77.75 | 22.19 | 70.69 |
| | 0.5 | 0.2 | 0.0506 | 0.9157 | 50.56 | 2.84 | 31.34 |
| | 1 | 0.5 | 0.0506 | 0.9999 | 50 | 0 | 7.86 |
| 100 | 0.1429 | 0.02 | 0.0494 | 0.6810 | 297.65 | 9.77 | 384.58 |
| | 0.2294 | 0.05 | 0.0498 | 0.8072 | 155.91 | 31.67 | 149.03 |
| | 0.3333 | 0.1 | 0.0487 | 0.9019 | 100.53 | 3.07 | 70.69 |
| | 0.5 | 0.2 | 0.0491 | 0.9966 | 100 | 0 | 31.34 |
| | 1 | 0.5 | 0.0508 | 1 | 100 | 0 | 7.86 |

Table 5: Linear regression with two continuous predictors, Pearson correlation between the predictors = 0.4, $N_{max} = 300$.

| n | $\beta_1$ | $R^2$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *Target SS* |
|---|---|---|---|---|---|---|---|
| 20 | 0.1429 | 0.02 | 0.0511 | 0.5979 | 291.13 | 27.94 | 457.70 |
| | 0.2294 | 0.05 | 0.0515 | 0.7972 | 201.82 | 72.60 | 177.49 |
| | 0.3333 | 0.1 | 0.0557 | 0.8118 | 104.77 | 51.88 | 84.19 |
| | 0.5 | 0.2 | 0.0588 | 0.8135 | 47.40 | 24.18 | 37.45 |
| | 1 | 0.5 | 0.0506 | 0.9377 | 20.55 | 2.59 | 9.34 |
| 50 | 0.1429 | 0.02 | 0.0499 | 0.6165 | 298.03 | 10.48 | 457.70 |
| | 0.2294 | 0.05 | 0.0511 | 0.8022 | 191.49 | 52.31 | 177.49 |
| | 0.3333 | 0.1 | 0.0544 | 0.8092 | 92.27 | 27.06 | 84.19 |
| | 0.5 | 0.2 | 0.0498 | 0.8775 | 51.91 | 5.59 | 37.45 |
| | 1 | 0.5 | 0.0498 | 0.9999 | 50 | 0 | 9.34 |
| 100 | 0.1429 | 0.02 | 0.0494 | 0.6125 | 299.71 | 3.13 | 457.70 |
| | 0.2294 | 0.05 | 0.0517 | 0.8012 | 185.29 | 37.45 | 177.49 |
| | 0.3333 | 0.1 | 0.0515 | 0.8517 | 103.08 | 8.06 | 84.19 |
| | 0.5 | 0.2 | 0.0500 | 0.9919 | 100 | 0 | 37.45 |
| | 1 | 0.5 | 0.0492 | 1 | 100 | 0 | 9.34 |

Table 6: Linear regression with two continuous predictors, Pearson correlation between the predictors = 0.8, $N_{max} = 300$.

| n | $\beta_1$ | $R^2$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *Target SS* |
|---|---|---|---|---|---|---|---|
| 20 | 0.1429 | 0.02 | 0.0498 | 0.3156 | 299.84 | 0.64 | 1067.68 |
| | 0.2294 | 0.05 | 0.0489 | 0.6400 | 287.22 | 33.80 | 414.85 |
| | 0.3333 | 0.1 | 0.0505 | 0.7972 | 216.26 | 70.93 | 196.10 |
| | 0.5 | 0.2 | 0.0548 | 0.8087 | 100.51 | 53.75 | 87.23 |
| | 1 | 0.5 | 0.0555 | 0.8293 | 29.36 | 12.79 | 21.79 |
| 50 | 0.1429 | 0.02 | 0.0498 | 0.3111 | 300.00 | 0.02 | 1067.68 |
| | 0.2294 | 0.05 | 0.0502 | 0.6506 | 295.99 | 14.94 | 414.85 |
| | 0.3333 | 0.1 | 0.0494 | 0.8031 | 209.49 | 53.58 | 196.10 |
| | 0.5 | 0.2 | 0.0519 | 0.8086 | 95.38 | 28.18 | 87.23 |
| | 1 | 0.5 | 0.0495 | 0.9736 | 50.02 | 0.41 | 21.79 |
| 100 | 0.1429 | 0.02 | 0.0503 | 0.3105 | 300 | 0 | 1067.68 |
| | 0.2294 | 0.05 | 0.0514 | 0.6584 | 298.91 | 6.58 | 414.85 |
| | 0.3333 | 0.1 | 0.0516 | 0.8067 | 204.36 | 40.22 | 196.10 |
| | 0.5 | 0.2 | 0.0507 | 0.8437 | 104.12 | 9.27 | 87.23 |
| | 1 | 0.5 | 0.0483 | 0.9998 | 100 | 0 | 21.79 |

Table 7: Linear regression with two independent binary predictors, $N_{max} = 300$.

| n | $\beta_1$ | $R^2$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *Target SS* |
|---|---|---|---|---|---|---|---|
| 20 | 0.1429 | 0.02 | 0.0515 | 0.6706 | 288.52 | 30.12 | 384.36 |
| | 0.2294 | 0.05 | 0.0514 | 0.8081 | 166.85 | 57.89 | 149.15 |
| | 0.3333 | 0.1 | 0.0551 | 0.8098 | 80.03 | 29.73 | 70.65 |
| | 0.5 | 0.2 | 0.0585 | 0.8008 | 36.17 | 12.89 | 31.40 |
| | 1 | 0.5 | 0.0486 | 0.9762 | 20.03 | 0.50 | 7.85 |
| 50 | 0.1429 | 0.02 | 0.0493 | 0.6841 | 296.88 | 12.24 | 384.36 |
| | 0.2294 | 0.05 | 0.0505 | 0.8048 | 156.10 | 32.79 | 149.15 |
| | 0.3333 | 0.1 | 0.0536 | 0.8012 | 74.53 | 15.25 | 70.65 |
| | 0.5 | 0.2 | 0.0500 | 0.9291 | 50.05 | 0.58 | 31.40 |
| | 1 | 0.5 | 0.0492 | 1 | 50 | 0 | 7.85 |
| 100 | 0.1429 | 0.02 | 0.0520 | 0.6936 | 299.24 | 4.87 | 384.36 |
| | 0.2294 | 0.05 | 0.0518 | 0.7919 | 152.57 | 22.08 | 149.15 |
| | 0.3333 | 0.1 | 0.0493 | 0.9116 | 100.04 | 0.54 | 70.65 |
| | 0.5 | 0.2 | 0.0506 | 0.9976 | 100 | 0 | 31.40 |
| | 1 | 0.5 | 0.0498 | 1 | 100 | 0 | 7.85 |

Table 8: Linear regression with two binary predictors, Pearson correlation between the predictors = 0.4 (tetrachoric correlation = 0.5878), $N_{max} = 300$.

| n | $\beta_1$ | $R^2$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *Target SS* |
|---|---|---|---|---|---|---|---|
| 20 | 0.1429 | 0.02 | 0.0502 | 0.6074 | 293.42 | 23.06 | 457.61 |
| | 0.2294 | 0.05 | 0.0507 | 0.8057 | 196.93 | 67.10 | 177.69 |
| | 0.3333 | 0.1 | 0.0535 | 0.8014 | 100.26 | 46.99 | 84.18 |
| | 0.5 | 0.2 | 0.0578 | 0.8111 | 45.57 | 24.15 | 37.40 |
| | 1 | 0.5 | 0.0514 | 0.9519 | 20.68 | 8.41 | 9.35 |
| 50 | 0.1429 | 0.02 | 0.0488 | 0.6238 | 299.05 | 6.66 | 457.61 |
| | 0.2294 | 0.05 | 0.0513 | 0.8020 | 188.28 | 45.50 | 177.69 |
| | 0.3333 | 0.1 | 0.0540 | 0.8075 | 90.01 | 23.10 | 84.18 |
| | 0.5 | 0.2 | 0.0508 | 0.8821 | 51.19 | 4.24 | 37.40 |
| | 1 | 0.5 | 0.0524 | 0.9998 | 50 | 0 | 9.35 |
| 100 | 0.1429 | 0.02 | 0.0494 | 0.6209 | 299.92 | 1.55 | 457.61 |
| | 0.2294 | 0.05 | 0.0506 | 0.8065 | 183.26 | 31.62 | 177.69 |
| | 0.3333 | 0.1 | 0.0517 | 0.8513 | 101.87 | 5.74 | 84.18 |
| | 0.5 | 0.2 | 0.0505 | 0.9935 | 100 | 0 | 37.40 |
| | 1 | 0.5 | 0.0496 | 1 | 100 | 0 | 9.35 |

Table 9: Linear regression with two binary predictors, Pearson correlation between the predictors = 0.8 (tetrachoric correlation = 0.9511), $N_{max} = 300$.

| n | $\beta_1$ | $R^2$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *Target SS* |
|---|---|---|---|---|---|---|---|
| 20 | 0.1429 | 0.02 | 0.0496 | 0.3154 | 299.53 | 5.94 | 1066.30 |
| | 0.2294 | 0.05 | 0.0501 | 0.6061 | 279.60 | 43.74 | 415.22 |
| | 0.3333 | 0.1 | 0.0529 | 0.7875 | 214.63 | 81.50 | 196.77 |
| | 0.5 | 0.2 | 0.0544 | 0.8103 | 129.42 | 86.09 | 87.35 |
| | 1 | 0.5 | 0.0557 | 0.8576 | 62.84 | 89.82 | 21.83 |
| 50 | 0.1429 | 0.02 | 0.0488 | 0.3184 | 300.00 | 0.64 | 1066.30 |
| | 0.2294 | 0.05 | 0.0498 | 0.6372 | 291.21 | 24.26 | 415.22 |
| | 0.3333 | 0.1 | 0.0512 | 0.8033 | 211.15 | 66.65 | 196.77 |
| | 0.5 | 0.2 | 0.0512 | 0.8219 | 108.10 | 57.39 | 87.35 |
| | 1 | 0.5 | 0.0490 | 0.9632 | 53.46 | 21.37 | 21.83 |
| 100 | 0.1429 | 0.02 | 0.0480 | 0.3132 | 300 | 0 | 1066.30 |
| | 0.2294 | 0.05 | 0.0502 | 0.6514 | 296.33 | 13.46 | 415.22 |
| | 0.3333 | 0.1 | 0.0526 | 0.7958 | 208.47 | 55.17 | 196.77 |
| | 0.5 | 0.2 | 0.0512 | 0.8672 | 111.88 | 26.59 | 87.35 |
| | 1 | 0.5 | 0.0490 | 0.9971 | 100.03 | 1.73 | 21.83 |

Table 10: Linear regression with two independent predictors (binary and continuous), the predictor of interest is binary, $N_{max} = 300$.

| n | $\beta_1$ | $R^2$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *Target SS* |
|---|---|---|---|---|---|---|---|
| 20 | 0.1429 | 0.02 | 0.0514 | 0.6729 | 288.23 | 38.38 | 384.36 |
| | 0.2294 | 0.05 | 0.0503 | 0.8008 | 167.12 | 57.49 | 149.15 |
| | 0.3333 | 0.1 | 0.0555 | 0.8059 | 79.92 | 29.32 | 70.65 |
| | 0.5 | 0.2 | 0.0594 | 0.7992 | 36.18 | 12.80 | 31.40 |
| | 1 | 0.5 | 0.0510 | 0.9768 | 20.03 | 1.31 | 7.85 |
| 50 | 0.1429 | 0.02 | 0.0501 | 0.6946 | 296.78 | 12.57 | 384.36 |
| | 0.2294 | 0.05 | 0.0514 | 0.7992 | 155.85 | 32.89 | 149.15 |
| | 0.3333 | 0.1 | 0.0536 | 0.8037 | 74.40 | 15.23 | 70.65 |
| | 0.5 | 0.2 | 0.0502 | 0.9275 | 50.05 | 0.59 | 31.40 |
| | 1 | 0.5 | 0.0491 | 1 | 50 | 0 | 7.85 |
| 100 | 0.1429 | 0.02 | 0.0519 | 0.6966 | 299.28 | 4.76 | 384.36 |
| | 0.2294 | 0.05 | 0.0514 | 0.8045 | 152.76 | 21.98 | 149.15 |
| | 0.3333 | 0.1 | 0.0509 | 0.9016 | 100.04 | 0.60 | 70.65 |
| | 0.5 | 0.2 | 0.0492 | 0.9987 | 100 | 0 | 31.40 |
| | 1 | 0.5 | 0.0484 | 1 | 100 | 0 | 7.85 |

Table 11: Linear regression with two predictors (binary and continuous), the predictor of interest is binary, Pearson correlation between the predictors = 0.4 (original correlation = 0.5037), $N_{max} = 300$.

| n | $\beta_1$ | $R^2$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *Target SS* |
|---|---|---|---|---|---|---|---|
| 20 | 0.1429 | 0.02 | 0.0505 | 0.6073 | 293.40 | 23.14 | 458.25 |
| | 0.2294 | 0.05 | 0.0510 | 0.7999 | 195.05 | 65.22 | 177.87 |
| | 0.3333 | 0.1 | 0.0552 | 0.7967 | 96.67 | 40.36 | 84.21 |
| | 0.5 | 0.2 | 0.0565 | 0.8080 | 43.54 | 17.94 | 37.43 |
| | 1 | 0.5 | 0.0495 | 0.9482 | 20.18 | 1.29 | 9.37 |
| 50 | 0.1429 | 0.02 | 0.0500 | 0.6151 | 299.15 | 6.29 | 458.25 |
| | 0.2294 | 0.05 | 0.0517 | 0.8020 | 186.46 | 43.36 | 177.87 |
| | 0.3333 | 0.1 | 0.0524 | 0.8065 | 88.91 | 21.08 | 84.21 |
| | 0.5 | 0.2 | 0.0507 | 0.8816 | 50.89 | 3.21 | 37.43 |
| | 1 | 0.5 | 0.0511 | 0.9999 | 50 | 0 | 9.37 |
| 100 | 0.1429 | 0.02 | 0.0506 | 0.6182 | 299.92 | 1.49 | 458.25 |
| | 0.2294 | 0.05 | 0.0520 | 0.7896 | 182.36 | 30.05 | 177.87 |
| | 0.3333 | 0.1 | 0.0506 | 0.8527 | 101.62 | 5.01 | 84.21 |
| | 0.5 | 0.2 | 0.0472 | 0.9949 | 100 | 0 | 37.43 |
| | 1 | 0.5 | 0.0511 | 1 | 100 | 0 | 9.37 |

Table 12: Linear regression with two predictors (binary and continuous), the predictor of interest is binary, Pearson correlation between the predictors = 0.8 (original correlation = 0.9991), $N_{max} = 300$.

| n | $\beta_1$ | $R^2$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *Target SS* |
|---|---|---|---|---|---|---|---|
| 20 | 0.1429 | 0.02 | 0.0489 | 0.3179 | 299.94 | 1.98 | 1052.06 |
| | 0.2294 | 0.05 | 0.0501 | 0.6451 | 289.09 | 30.03 | 408.97 |
| | 0.3333 | 0.1 | 0.0515 | 0.7946 | 212.20 | 66.78 | 193.72 |
| | 0.5 | 0.2 | 0.0568 | 0.8061 | 101.79 | 43.46 | 86.16 |
| | 1 | 0.5 | 0.0554 | 0.8340 | 27.27 | 9.72 | 21.54 |
| 50 | 0.1429 | 0.02 | 0.0499 | 0.3136 | 300 | 0 | 1052.06 |
| | 0.2294 | 0.05 | 0.0493 | 0.6532 | 297.11 | 12.26 | 408.97 |
| | 0.3333 | 0.1 | 0.0511 | 0.7986 | 204.51 | 47.42 | 193.72 |
| | 0.5 | 0.2 | 0.0526 | 0.8100 | 92.05 | 23.10 | 86.16 |
| | 1 | 0.5 | 0.0519 | 0.9779 | 50.00 | 0.07 | 21.54 |
| 100 | 0.1429 | 0.02 | 0.0488 | 0.3163 | 300 | 0 | 1052.06 |
| | 0.2294 | 0.05 | 0.0498 | 0.6650 | 299.39 | 4.56 | 408.97 |
| | 0.3333 | 0.1 | 0.0514 | 0.8037 | 199.78 | 34.08 | 193.72 |
| | 0.5 | 0.2 | 0.0509 | 0.8451 | 102.43 | 6.38 | 86.16 |
| | 1 | 0.5 | 0.0504 | 0.9999 | 100 | 0 | 21.54 |

Table 13: Linear regression with independent predictors (binary and continuous), the predictor of interest is continuous, $N_{max} = 300$.

| n | $\beta_1$ | $R^2$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *Target SS* |
|---|---|---|---|---|---|---|---|
| 20 | 0.1429 | 0.02 | 0.0510 | 0.6602 | 283.63 | 38.05 | 384.02 |
| | 0.2294 | 0.05 | 0.0505 | 0.8109 | 177.43 | 71.87 | 149.09 |
| | 0.3333 | 0.1 | 0.0547 | 0.8133 | 88.66 | 45.29 | 70.56 |
| | 0.5 | 0.2 | 0.0563 | 0.8219 | 40.22 | 20.21 | 31.34 |
| | 1 | 0.5 | 0.0511 | 0.9590 | 20.25 | 1.69 | 7.85 |
| 50 | 0.1429 | 0.02 | 0.0518 | 0.6701 | 293.37 | 19.84 | 384.02 |
| | 0.2294 | 0.05 | 0.0520 | 0.8045 | 161.84 | 46.86 | 149.09 |
| | 0.3333 | 0.1 | 0.0529 | 0.8140 | 77.91 | 22.33 | 70.56 |
| | 0.5 | 0.2 | 0.0496 | 0.9174 | 50.58 | 2.89 | 31.34 |
| | 1 | 0.5 | 0.0502 | 1 | 50 | 0 | 7.85 |
| 100 | 0.1429 | 0.02 | 0.0497 | 0.6849 | 297.56 | 10.15 | 384.02 |
| | 0.2294 | 0.05 | 0.0516 | 0.8111 | 155.86 | 31.48 | 149.09 |
| | 0.3333 | 0.1 | 0.0496 | 0.9008 | 100.53 | 3.02 | 70.56 |
| | 0.5 | 0.2 | 0.0502 | 0.9979 | 100 | 0 | 31.34 |
| | 1 | 0.5 | 0.0510 | 1 | 100 | 0 | 7.85 |

Table 14: Linear regression with two predictors (continuous and binary), the predictor of interest is continuous, Pearson correlation between the predictors = 0.4 (original correlation = 0.5037), $N_{max} = 300$.

| n | $\beta_1$ | $R^2$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *Target SS* |
|---|---|---|---|---|---|---|---|
| 20 | 0.1429 | 0.02 | 0.0513 | 0.6030 | 291.16 | 27.89 | 458.75 |
| | 0.2294 | 0.05 | 0.0518 | 0.7942 | 201.45 | 72.39 | 178.24 |
| | 0.3333 | 0.1 | 0.0537 | 0.8163 | 105.01 | 52.59 | 84.22 |
| | 0.5 | 0.2 | 0.0576 | 0.8183 | 47.60 | 24.30 | 37.45 |
| | 1 | 0.5 | 0.0506 | 0.9386 | 20.57 | 2.65 | 9.37 |
| 50 | 0.1429 | 0.02 | 0.0512 | 0.6204 | 298.10 | 10.35 | 458.75 |
| | 0.2294 | 0.05 | 0.0512 | 0.8011 | 191.60 | 52.09 | 178.24 |
| | 0.3333 | 0.1 | 0.0541 | 0.8115 | 92.03 | 27.08 | 84.22 |
| | 0.5 | 0.2 | 0.0506 | 0.8862 | 52.02 | 5.82 | 37.45 |
| | 1 | 0.5 | 0.0497 | 0.9997 | 50 | 0 | 9.37 |
| 100 | 0.1429 | 0.02 | 0.0481 | 0.6128 | 299.72 | 3.18 | 458.75 |
| | 0.2294 | 0.05 | 0.0514 | 0.8100 | 185.74 | 37.36 | 178.24 |
| | 0.3333 | 0.1 | 0.0516 | 0.8646 | 103.12 | 8.09 | 84.22 |
| | 0.5 | 0.2 | 0.0502 | 0.9919 | 100 | 0 | 37.45 |
| | 1 | 0.5 | 0.0507 | 1 | 100 | 0 | 9.37 |

Table 15: Linear regression with two predictors (continuous and binary), the predictor of interest is continuous, Pearson correlation between the predictors = 0.8 (original correlation = 0.9991), $N_{max} = 300$.

| n | $\beta_1$ | $R^2$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *Target SS* |
|---|---|---|---|---|---|---|---|
| 20 | 0.1429 | 0.02 | 0.0510 | 0.3162 | 299.77 | 4.37 | 1052.62 |
| | 0.2294 | 0.05 | 0.0507 | 0.6410 | 285.53 | 36.52 | 408.51 |
| | 0.3333 | 0.1 | 0.0508 | 0.8045 | 215.12 | 73.92 | 193.89 |
| | 0.5 | 0.2 | 0.0539 | 0.8127 | 111.57 | 59.11 | 86.19 |
| | 1 | 0.5 | 0.0571 | 0.8481 | 30.27 | 14.78 | 21.53 |
| 50 | 0.1429 | 0.02 | 0.0491 | 0.3146 | 300.00 | 0.02 | 1052.62 |
| | 0.2294 | 0.05 | 0.0488 | 0.6516 | 294.59 | 18.21 | 408.51 |
| | 0.3333 | 0.1 | 0.0523 | 0.8109 | 208.59 | 56.84 | 193.89 |
| | 0.5 | 0.2 | 0.0534 | 0.8180 | 95.52 | 31.07 | 86.19 |
| | 1 | 0.5 | 0.0505 | 0.9732 | 50.04 | 0.71 | 21.53 |
| 100 | 0.1429 | 0.02 | 0.0495 | 0.3169 | 300 | 0 | 1052.62 |
| | 0.2294 | 0.05 | 0.0503 | 0.6610 | 298.17 | 8.90 | 408.51 |
| | 0.3333 | 0.1 | 0.0527 | 0.7971 | 203.27 | 43.44 | 193.89 |
| | 0.5 | 0.2 | 0.0505 | 0.8525 | 104.68 | 10.62 | 86.19 |
| | 1 | 0.5 | 0.0498 | 0.9999 | 100 | 0 | 21.53 |

Table 16: Linear regression with 10 independent continuous predictors, $N_{max} = 300$.

| n | $\beta_1$ | $R^2$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *Target SS* |
|---|---|---|---|---|---|---|---|
| 20 | 0.1429 | 0.02 | 0.0503 | 0.6607 | 293.49 | 26.76 | 384.49 |
| | 0.2294 | 0.05 | 0.0499 | 0.8949 | 242.98 | 73.32 | 149.09 |
| | 0.3333 | 0.1 | 0.0554 | 0.8952 | 157.92 | 83.33 | 70.71 |
| | 0.5 | 0.2 | 0.0611 | 0.8899 | 78.05 | 55.09 | 31.39 |
| | 1 | 0.5 | 0.0538 | 0.8977 | 25.23 | 12.48 | 7.86 |
| 50 | 0.1429 | 0.02 | 0.0489 | 0.6809 | 297.25 | 12.98 | 384.49 |
| | 0.2294 | 0.05 | 0.0522 | 0.8417 | 193.11 | 56.15 | 149.09 |
| | 0.3333 | 0.1 | 0.0550 | 0.8369 | 93.67 | 30.46 | 70.71 |
| | 0.5 | 0.2 | 0.0498 | 0.8809 | 52.59 | 6.94 | 31.39 |
| | 1 | 0.5 | 0.0502 | 0.9997 | 50 | 0 | 7.86 |
| 100 | 0.1429 | 0.02 | 0.0509 | 0.6746 | 298.82 | 6.92 | 384.49 |
| | 0.2294 | 0.05 | 0.0518 | 0.8155 | 169.91 | 35.98 | 149.09 |
| | 0.3333 | 0.1 | 0.0515 | 0.8800 | 101.55 | 5.65 | 70.71 |
| | 0.5 | 0.2 | 0.0502 | 0.9954 | 100 | 0 | 31.39 |
| | 1 | 0.5 | 0.0511 | 1 | 100 | 0 | 7.86 |

Table 17: Linear regression with 10 continuous predictors, Pearson correlation among the predictors = 0.4, $N_{max} = 300$.

| n | $\beta_1$ | $R^2$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *Target SS* |
|---|---|---|---|---|---|---|---|
| 20 | 0.1429 | 0.02 | 0.0483 | 0.5065 | 298.08 | 14.39 | 584.82 |
| | 0.2294 | 0.05 | 0.0498 | 0.8375 | 274.37 | 52.56 | 227.17 |
| | 0.3333 | 0.1 | 0.0526 | 0.9023 | 207.36 | 83.29 | 107.53 |
| | 0.5 | 0.2 | 0.0558 | 0.8914 | 114.66 | 72.23 | 47.72 |
| | 1 | 0.5 | 0.0612 | 0.8796 | 33.05 | 21.64 | 11.94 |
| 50 | 0.1429 | 0.02 | 0.0493 | 0.4973 | 299.91 | 2.06 | 584.82 |
| | 0.2294 | 0.05 | 0.0498 | 0.8303 | 261.32 | 47.15 | 227.17 |
| | 0.3333 | 0.1 | 0.0522 | 0.8484 | 141.39 | 45.84 | 107.53 |
| | 0.5 | 0.2 | 0.0531 | 0.8370 | 65.80 | 18.23 | 47.72 |
| | 1 | 0.5 | 0.0493 | 0.9964 | 50.00 | 0.09 | 11.94 |
| 100 | 0.1429 | 0.02 | 0.0504 | 0.4964 | 300.00 | 0.19 | 584.82 |
| | 0.2294 | 0.05 | 0.0515 | 0.8173 | 249.97 | 41.13 | 227.17 |
| | 0.3333 | 0.1 | 0.0521 | 0.8175 | 124.72 | 23.50 | 107.53 |
| | 0.5 | 0.2 | 0.0503 | 0.9646 | 100.01 | 0.37 | 47.72 |
| | 1 | 0.5 | 0.0509 | 1 | 100 | 0 | 11.94 |

Table 18: Linear regression with 10 continuous predictors, Pearson correlation among the predictors = 0.8, $N_{max} = 300$.

| n | $\beta_1$ | $R^2$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *Target SS* |
|---|---|---|---|---|---|---|---|
| 20 | 0.1429 | 0.02 | 0.0483 | 0.2104 | 299.96 | 1.89 | 1732.42 |
| | 0.2294 | 0.05 | 0.0502 | 0.4481 | 298.80 | 11.50 | 672.42 |
| | 0.3333 | 0.1 | 0.0494 | 0.7460 | 289.01 | 34.98 | 318.95 |
| | 0.5 | 0.2 | 0.0495 | 0.8891 | 237.71 | 75.33 | 141.67 |
| | 1 | 0.5 | 0.0603 | 0.8916 | 87.31 | 59.72 | 35.28 |
| 50 | 0.1429 | 0.02 | 0.0495 | 0.2112 | 300 | 0 | 1732.42 |
| | 0.2294 | 0.05 | 0.0506 | 0.4513 | 299.98 | 0.83 | 672.42 |
| | 0.3333 | 0.1 | 0.0511 | 0.7463 | 291.38 | 23.37 | 318.95 |
| | 0.5 | 0.2 | 0.0529 | 0.8508 | 184.03 | 55.30 | 141.67 |
| | 1 | 0.5 | 0.0528 | 0.8631 | 54.81 | 9.80 | 35.28 |
| 100 | 0.1429 | 0.02 | 0.0507 | 0.2063 | 300 | 0 | 1732.42 |
| | 0.2294 | 0.05 | 0.0489 | 0.4513 | 300 | 0 | 672.42 |
| | 0.3333 | 0.1 | 0.0519 | 0.7500 | 292.81 | 18.06 | 318.95 |
| | 0.5 | 0.2 | 0.0519 | 0.8185 | 161.66 | 34.23 | 141.67 |
| | 1 | 0.5 | 0.0509 | 0.9895 | 100.00 | 0.04 | 35.28 |

Table 19: Linear regression with 10 independent binary predictors, $N_{max} = 300$.

| n | $\beta_1$ | $R^2$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *Target SS* |
|---|---|---|---|---|---|---|---|
| 20 | 0.1429 | 0.02 | 0.0484 | 0.6678 | 294.66 | 23.91 | 384.37 |
| | 0.2294 | 0.05 | 0.0512 | 0.8905 | 242.58 | 70.45 | 149.15 |
| | 0.3333 | 0.1 | 0.0525 | 0.8953 | 149.99 | 76.24 | 70.65 |
| | 0.5 | 0.2 | 0.0620 | 0.8770 | 71.10 | 45.28 | 31.40 |
| | 1 | 0.5 | 0.0540 | 0.8944 | 23.59 | 9.60 | 7.85 |
| 50 | 0.1429 | 0.02 | 0.0486 | 0.6831 | 298.85 | 7.72 | 384.37 |
| | 0.2294 | 0.05 | 0.0561 | 0.8438 | 188.33 | 45.53 | 149.15 |
| | 0.3333 | 0.1 | 0.0529 | 0.8254 | 89.91 | 22.49 | 70.65 |
| | 0.5 | 0.2 | 0.0492 | 0.8842 | 51.10 | 3.66 | 31.40 |
| | 1 | 0.5 | 0.0501 | 1 | 50 | 0 | 7.85 |
| 100 | 0.1429 | 0.02 | 0.0485 | 0.6713 | 299.68 | 3.19 | 384.37 |
| | 0.2294 | 0.05 | 0.0521 | 0.8106 | 166.30 | 26.30 | 149.15 |
| | 0.3333 | 0.1 | 0.0499 | 0.8838 | 100.37 | 2.09 | 70.65 |
| | 0.5 | 0.2 | 0.0497 | 0.9972 | 100 | 0 | 31.40 |
| | 1 | 0.5 | 0.0493 | 1 | 100 | 0 | 7.85 |

Table 20: Linear regression with 10 binary predictors, Pairwise Pearson correlations = 0.4 (pairwise tetrachoric correlations = 0.5878), $N_{max} = 300$.

| n | $\beta_1$ | $R^2$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *Target SS* |
|---|---|---|---|---|---|---|---|
| 20 | 0.1429 | 0.02 | 0.0511 | 0.5035 | 298.73 | 11.56 | 584.24 |
| | 0.2294 | 0.05 | 0.0488 | 0.8453 | 281.09 | 45.89 | 227.21 |
| | 0.3333 | 0.1 | 0.0498 | 0.9230 | 226.07 | 81.83 | 107.63 |
| | 0.5 | 0.2 | 0.0556 | 0.9160 | 144.14 | 90.51 | 47.75 |
| | 1 | 0.5 | 0.0618 | 0.8952 | 54.28 | 65.52 | 11.93 |
| 50 | 0.1429 | 0.02 | 0.0492 | 0.4963 | 299.96 | 1.25 | 584.24 |
| | 0.2294 | 0.05 | 0.0497 | 0.8205 | 266.70 | 44.03 | 227.21 |
| | 0.3333 | 0.1 | 0.0524 | 0.8543 | 149.41 | 50.48 | 107.63 |
| | 0.5 | 0.2 | 0.0509 | 0.8381 | 68.91 | 22.27 | 47.75 |
| | 1 | 0.5 | 0.0492 | 0.9939 | 50.01 | 0.40 | 11.93 |
| 100 | 0.1429 | 0.02 | 0.0507 | 0.5002 | 300.00 | 0.06 | 584.24 |
| | 0.2294 | 0.05 | 0.0506 | 0.8181 | 253.68 | 39.73 | 227.21 |
| | 0.3333 | 0.1 | 0.0503 | 0.8196 | 126.95 | 24.89 | 107.63 |
| | 0.5 | 0.2 | 0.0494 | 0.9596 | 100.03 | 0.68 | 47.75 |
| | 1 | 0.5 | 0.0493 | 1 | 100 | 0 | 11.93 |

Table 21: Linear regression with 10 binary predictors, Pairwise Pearson correlations = 0.8 (pairwise tetrachoric correlations = 0.9511), $N_{max} = 300$.

| n | $\beta_1$ | $R^2$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *Target SS* |
|---|---|---|---|---|---|---|---|
| 20 | 0.1429 | 0.02 | 0.0504 | 0.1982 | 300.00 | 0.00 | 1736.86 |
| | 0.2294 | 0.05 | 0.0492 | 0.4207 | 299.94 | 2.59 | 672.07 |
| | 0.3333 | 0.1 | 0.0495 | 0.7181 | 299.67 | 6.53 | 318.79 |
| | 0.5 | 0.2 | 0.0494 | 0.9552 | 298.53 | 15.65 | 141.61 |
| | 1 | 0.5 | 0.0506 | 0.9978 | 295.87 | 30.41 | 35.47 |
| 50 | 0.1429 | 0.02 | 0.0515 | 0.1914 | 300 | 0 | 1736.86 |
| | 0.2294 | 0.05 | 0.0505 | 0.4267 | 299.95 | 1.63 | 672.07 |
| | 0.3333 | 0.1 | 0.0499 | 0.7171 | 295.90 | 18.47 | 318.79 |
| | 0.5 | 0.2 | 0.0515 | 0.9143 | 257.76 | 64.17 | 141.61 |
| | 1 | 0.5 | 0.0519 | 0.9296 | 149.81 | 104.10 | 35.47 |
| 100 | 0.1429 | 0.02 | 0.0482 | 0.2022 | 300 | 0 | 1736.86 |
| | 0.2294 | 0.05 | 0.0491 | 0.4212 | 299.99 | 0.66 | 672.07 |
| | 0.3333 | 0.1 | 0.0504 | 0.7177 | 294.09 | 19.19 | 318.79 |
| | 0.5 | 0.2 | 0.0515 | 0.8599 | 210.81 | 64.76 | 141.61 |
| | 1 | 0.5 | 0.0520 | 0.9529 | 104.06 | 20.08 | 35.47 |

Table 22: Linear regression with 10 independent predictors, the predictor of interest is binary, other 9 predictors are continuous, $N_{max} = 300$.

| n | $\beta_1$ | $R^2$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *Target SS* |
|---|---|---|---|---|---|---|---|
| 20 | 0.1429 | 0.02 | 0.0513 | 0.6604 | 294.81 | 23.66 | 384.37 |
| | 0.2294 | 0.05 | 0.0524 | 0.8961 | 243.06 | 70.45 | 149.15 |
| | 0.3333 | 0.1 | 0.0552 | 0.9009 | 149.92 | 75.88 | 70.65 |
| | 0.5 | 0.2 | 0.0625 | 0.8828 | 71.20 | 44.52 | 31.40 |
| | 1 | 0.5 | 0.0540 | 0.8936 | 23.47 | 8.81 | 7.85 |
| 50 | 0.1429 | 0.02 | 0.0501 | 0.6741 | 298.80 | 7.89 | 384.37 |
| | 0.2294 | 0.05 | 0.0503 | 0.8442 | 188.54 | 45.46 | 149.15 |
| | 0.3333 | 0.1 | 0.0541 | 0.8310 | 89.87 | 22.29 | 70.65 |
| | 0.5 | 0.2 | 0.0506 | 0.8753 | 51.08 | 3.60 | 31.40 |
| | 1 | 0.5 | 0.0503 | 1 | 50 | 0 | 7.85 |
| 100 | 0.1429 | 0.02 | 0.0515 | 0.6737 | 299.70 | 3.07 | 384.37 |
| | 0.2294 | 0.05 | 0.0520 | 0.8152 | 166.57 | 26.31 | 149.15 |
| | 0.3333 | 0.1 | 0.0501 | 0.8781 | 100.38 | 2.14 | 70.65 |
| | 0.5 | 0.2 | 0.0499 | 0.9970 | 100 | 0 | 31.40 |
| | 1 | 0.5 | 0.0491 | 1 | 100 | 0 | 7.85 |

Table 23: Linear regression with 10 predictors, the predictor of interest is binary, other 9 predictors are continuous, pairwise Pearson correlation = 0.4 (original pairwise correlation = 0.5037), $N_{max} = 300$.

| n | $\beta_1$ | $R^2$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *Target SS* |
|---|---|---|---|---|---|---|---|
| 20 | 0.1429 | 0.02 | 0.0501 | 0.4951 | 298.40 | 13.12 | 587.90 |
| | 0.2294 | 0.05 | 0.0536 | 0.8356 | 274.34 | 52.01 | 228.31 |
| | 0.3333 | 0.1 | 0.0517 | 0.8981 | 204.34 | 81.87 | 108.14 |
| | 0.5 | 0.2 | 0.0609 | 0.8876 | 108.48 | 66.28 | 48.04 |
| | 1 | 0.5 | 0.0608 | 0.8701 | 31.09 | 18.58 | 12.00 |
| 50 | 0.1429 | 0.02 | 0.0502 | 0.4999 | 299.96 | 1.26 | 587.90 |
| | 0.2294 | 0.05 | 0.0509 | 0.8266 | 262.08 | 44.58 | 228.31 |
| | 0.3333 | 0.1 | 0.0528 | 0.8416 | 138.44 | 40.61 | 108.14 |
| | 0.5 | 0.2 | 0.0510 | 0.8301 | 63.94 | 15.83 | 48.04 |
| | 1 | 0.5 | 0.0495 | 0.9974 | 50.00 | 0.04 | 12.00 |
| 100 | 0.1429 | 0.02 | 0.0500 | 0.4975 | 300.00 | 0.10 | 587.90 |
| | 0.2294 | 0.05 | 0.0523 | 0.8168 | 250.47 | 38.24 | 228.31 |
| | 0.3333 | 0.1 | 0.0521 | 0.8141 | 123.10 | 20.92 | 108.14 |
| | 0.5 | 0.2 | 0.0512 | 0.9613 | 100.00 | 0.20 | 48.04 |
| | 1 | 0.5 | 0.0492 | 1 | 100 | 0 | 12.00 |

Table 24: Linear regression with 10 predictors, the predictor of interest is binary, other 9 predictors are continuous, pairwise Pearson correlation = 0.72 (original pairwise correlation = 0.9), $N_{max} = 300$.

| n | $\beta_1$ | $R^2$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *Target SS* |
|---|---|---|---|---|---|---|---|
| 20 | 0.1429 | 0.02 | 0.0514 | 0.3108 | 299.83 | 3.93 | 1030.71 |
| | 0.2294 | 0.05 | 0.0483 | 0.6532 | 294.64 | 24.12 | 400.25 |
| | 0.3333 | 0.1 | 0.0514 | 0.8685 | 263.97 | 59.99 | 189.36 |
| | 0.5 | 0.2 | 0.0536 | 0.9017 | 174.57 | 81.95 | 84.07 |
| | 1 | 0.5 | 0.0662 | 0.8785 | 50.45 | 34.42 | 21.07 |
| 50 | 0.1429 | 0.02 | 0.0501 | 0.3128 | 300 | 0 | 1030.71 |
| | 0.2294 | 0.05 | 0.0504 | 0.6577 | 298.65 | 8.72 | 400.25 |
| | 0.3333 | 0.1 | 0.0515 | 0.8430 | 233.78 | 51.53 | 189.36 |
| | 0.5 | 0.2 | 0.0540 | 0.8413 | 108.76 | 31.23 | 84.07 |
| | 1 | 0.5 | 0.0498 | 0.9564 | 50.06 | 0.79 | 21.07 |
| 100 | 0.1429 | 0.02 | 0.0497 | 0.3193 | 300 | 0 | 1030.71 |
| | 0.2294 | 0.05 | 0.0492 | 0.6688 | 299.62 | 3.65 | 400.25 |
| | 0.3333 | 0.1 | 0.0514 | 0.8224 | 213.11 | 38.16 | 189.36 |
| | 0.5 | 0.2 | 0.0522 | 0.8425 | 104.82 | 9.50 | 84.07 |
| | 1 | 0.5 | 0.0510 | 0.9999 | 100 | 0 | 21.07 |

Table 25: Linear regression with 10 independent predictors, the predictor of interest is continuous, other 9 predictors are binary, $N_{max} = 300$.

| n | $\beta_1$ | $R^2$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *Target SS* |
|---|---|---|---|---|---|---|---|
| 20 | 0.1429 | 0.02 | 0.0508 | 0.6699 | 293.43 | 27.05 | 384.34 |
| | 0.2294 | 0.05 | 0.0527 | 0.8906 | 242.68 | 73.41 | 149.16 |
| | 0.3333 | 0.1 | 0.0542 | 0.8988 | 158.12 | 83.02 | 70.62 |
| | 0.5 | 0.2 | 0.0614 | 0.8877 | 78.07 | 54.59 | 31.41 |
| | 1 | 0.5 | 0.0553 | 0.9024 | 25.24 | 12.97 | 7.86 |
| 50 | 0.1429 | 0.02 | 0.0495 | 0.6771 | 297.21 | 13.04 | 384.34 |
| | 0.2294 | 0.05 | 0.0501 | 0.8405 | 193.23 | 56.22 | 149.16 |
| | 0.3333 | 0.1 | 0.0529 | 0.8463 | 93.37 | 30.29 | 70.62 |
| | 0.5 | 0.2 | 0.0500 | 0.8823 | 52.57 | 6.90 | 31.41 |
| | 1 | 0.5 | 0.0508 | 0.9999 | 50 | 0 | 7.86 |
| 100 | 0.1429 | 0.02 | 0.0492 | 0.6793 | 298.86 | 6.76 | 384.34 |
| | 0.2294 | 0.05 | 0.0509 | 0.8174 | 169.88 | 36.04 | 149.16 |
| | 0.3333 | 0.1 | 0.0515 | 0.8780 | 101.56 | 5.64 | 70.62 |
| | 0.5 | 0.2 | 0.0496 | 0.9948 | 100 | 0 | 31.41 |
| | 1 | 0.5 | 0.0491 | 1 | 100 | 0 | 7.86 |

Table 26: Linear regression with 10 predictors, the predictor of interest is continuous, other 9 predictors are binary, pairwise Pearson correlation = 0.4 (original pairwise correlation = 0.5037), $N_{max} = 300$.

| n | $\beta_1$ | $R^2$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *Target SS* |
|---|---|---|---|---|---|---|---|
| 20 | 0.1429 | 0.02 | 0.0515 | 0.4985 | 298.19 | 14.05 | 587.73 |
| | 0.2294 | 0.05 | 0.0506 | 0.8325 | 274.60 | 52.62 | 227.98 |
| | 0.3333 | 0.1 | 0.0522 | 0.9054 | 209.27 | 83.73 | 108.06 |
| | 0.5 | 0.2 | 0.0551 | 0.8929 | 119.29 | 77.78 | 48.09 |
| | 1 | 0.5 | 0.0601 | 0.8769 | 41.13 | 51.65 | 11.99 |
| 50 | 0.1429 | 0.02 | 0.0506 | 0.5020 | 299.93 | 1.80 | 587.73 |
| | 0.2294 | 0.05 | 0.0508 | 0.8236 | 260.84 | 47.26 | 227.98 |
| | 0.3333 | 0.1 | 0.0515 | 0.8345 | 141.34 | 45.65 | 108.06 |
| | 0.5 | 0.2 | 0.0529 | 0.8371 | 65.78 | 18.34 | 48.09 |
| | 1 | 0.5 | 0.0505 | 0.9955 | 50.00 | 0.08 | 11.99 |
| 100 | 0.1429 | 0.02 | 0.0493 | 0.5051 | 300.00 | 0.23 | 587.73 |
| | 0.2294 | 0.05 | 0.0512 | 0.8167 | 250.24 | 41.16 | 227.98 |
| | 0.3333 | 0.1 | 0.0523 | 0.8197 | 124.87 | 23.68 | 108.06 |
| | 0.5 | 0.2 | 0.0512 | 0.9592 | 100.01 | 0.40 | 48.09 |
| | 1 | 0.5 | 0.0493 | 1 | 100 | 0 | 11.99 |

Table 27: Linear regression with 10 predictors, the predictor of interest is continuous, other 9 predictors are binary, pairwise Pearson correlation = 0.72 (original pairwise correlation = 0.9), $N_{max} = 300$.

| n | $\beta_1$ | $R^2$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *Target SS* |
|---|---|---|---|---|---|---|---|
| 20 | 0.1429 | 0.02 | 0.0499 | 0.3237 | 299.98 | 1.30 | 1030.10 |
| | 0.2294 | 0.05 | 0.0512 | 0.6682 | 299.67 | 6.22 | 399.13 |
| | 0.3333 | 0.1 | 0.0507 | 0.9337 | 298.02 | 16.66 | 189.62 |
| | 0.5 | 0.2 | 0.0498 | 0.9935 | 293.74 | 33.59 | 84.09 |
| | 1 | 0.5 | 0.0506 | 0.9941 | 207.98 | 53.95 | 21.05 |
| 50 | 0.1429 | 0.02 | 0.0499 | 0.3218 | 300.00 | 0.02 | 1030.10 |
| | 0.2294 | 0.05 | 0.0499 | 0.6619 | 297.03 | 13.78 | 399.13 |
| | 0.3333 | 0.1 | 0.0492 | 0.8490 | 239.05 | 59.37 | 189.62 |
| | 0.5 | 0.2 | 0.0549 | 0.8621 | 139.79 | 79.60 | 84.09 |
| | 1 | 0.5 | 0.0505 | 0.9649 | 91.39 | 92.72 | 21.05 |
| 100 | 0.1429 | 0.02 | 0.0506 | 0.3247 | 300 | 0 | 1030.10 |
| | 0.2294 | 0.05 | 0.0509 | 0.6604 | 298.65 | 7.66 | 399.13 |
| | 0.3333 | 0.1 | 0.0487 | 0.8115 | 209.54 | 44.87 | 189.62 |
| | 0.5 | 0.2 | 0.0511 | 0.8545 | 106.26 | 13.99 | 84.09 |
| | 1 | 0.5 | 0.0491 | 0.9997 | 100.22 | 6.63 | 21.05 |

Table 28: Linear regression with a single continuous predictor, $N_{max} = 600$.

| n | $\beta_1$ | $R^2$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *Target SS* |
|---|---|---|---|---|---|---|---|
| 20 | 0.1429 | 0.02 | 0.0510 | 0.7859 | 410.55 | 142.71 | 385.11 |
| | 0.2294 | 0.05 | 0.0522 | 0.7909 | 176.14 | 87.23 | 149.21 |
| | 0.3333 | 0.1 | 0.0543 | 0.8063 | 83.61 | 42.45 | 70.68 |
| | 0.5 | 0.2 | 0.0551 | 0.8195 | 38.04 | 18.30 | 31.36 |
| | 1 | 0.5 | 0.0506 | 0.9654 | 20.16 | 1.26 | 7.86 |
| 50 | 0.1429 | 0.02 | 0.0502 | 0.8019 | 404.31 | 105.69 | 385.11 |
| | 0.2294 | 0.05 | 0.0522 | 0.8013 | 159.40 | 47.11 | 149.21 |
| | 0.3333 | 0.1 | 0.0519 | 0.8068 | 76.23 | 21.48 | 70.68 |
| | 0.5 | 0.2 | 0.0516 | 0.9220 | 50.47 | 2.57 | 31.36 |
| | 1 | 0.5 | 0.0500 | 1 | 50 | 0 | 7.86 |
| 100 | 0.1429 | 0.02 | 0.0510 | 0.8077 | 396.08 | 79.36 | 385.11 |
| | 0.2294 | 0.05 | 0.0511 | 0.8096 | 154.26 | 30.90 | 149.21 |
| | 0.3333 | 0.1 | 0.0512 | 0.9029 | 100.46 | 2.80 | 70.68 |
| | 0.5 | 0.2 | 0.0477 | 0.9970 | 100 | 0 | 31.36 |
| | 1 | 0.5 | 0.0502 | 1 | 100 | 0 | 7.86 |

Table 29: Linear regression with a single binary predictor, $N_{max} = 600$.

| n | $\beta_1$ | $R^2$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *Target SS* |
|---|---|---|---|---|---|---|---|
| 20 | 0.1429 | 0.02 | 0.0493 | 0.7832 | 399.02 | 120.29 | 385.11 |
| | 0.2294 | 0.05 | 0.0493 | 0.7874 | 158.76 | 54.84 | 149.21 |
| | 0.3333 | 0.1 | 0.0541 | 0.7946 | 75.45 | 26.12 | 70.68 |
| | 0.5 | 0.2 | 0.0577 | 0.7999 | 34.07 | 11.12 | 31.36 |
| | 1 | 0.5 | 0.0507 | 0.9841 | 20.01 | 0.30 | 7.86 |
| 50 | 0.1429 | 0.02 | 0.0481 | 0.7943 | 393.13 | 79.49 | 385.11 |
| | 0.2294 | 0.05 | 0.0498 | 0.8003 | 152.86 | 31.56 | 149.21 |
| | 0.3333 | 0.1 | 0.0537 | 0.7979 | 72.96 | 14.41 | 70.68 |
| | 0.5 | 0.2 | 0.0489 | 0.9284 | 50.03 | 0.45 | 31.36 |
| | 1 | 0.5 | 0.0501 | 1 | 50 | 0 | 7.86 |
| 100 | 0.1429 | 0.02 | 0.0497 | 0.8018 | 388.69 | 55.79 | 385.11 |
| | 0.2294 | 0.05 | 0.0494 | 0.7970 | 151.20 | 21.55 | 149.21 |
| | 0.3333 | 0.1 | 0.0489 | 0.9097 | 100.03 | 0.53 | 70.68 |
| | 0.5 | 0.2 | 0.0490 | 0.9988 | 100 | 0 | 31.36 |
| | 1 | 0.5 | 0.0507 | 1 | 100 | 0 | 7.86 |

Table 30: Linear regression with two independent continuous predictors, $N_{max} = 600$.

| n | $\beta_1$ | $R^2$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *Target SS* |
|---|---|---|---|---|---|---|---|
| 20 | 0.1429 | 0.02 | 0.0511 | 0.7971 | 425.13 | 143.08 | 384.58 |
| | 0.2294 | 0.05 | 0.0540 | 0.8062 | 186.48 | 94.79 | 149.03 |
| | 0.3333 | 0.1 | 0.0554 | 0.8051 | 88.43 | 46.16 | 70.69 |
| | 0.5 | 0.2 | 0.0583 | 0.8216 | 40.08 | 19.87 | 31.34 |
| | 1 | 0.5 | 0.0497 | 0.9576 | 20.25 | 1.70 | 7.86 |
| 50 | 0.1429 | 0.02 | 0.0498 | 0.7996 | 411.43 | 107.09 | 384.58 |
| | 0.2294 | 0.05 | 0.0526 | 0.8109 | 162.80 | 48.49 | 149.03 |
| | 0.3333 | 0.1 | 0.0522 | 0.8152 | 77.96 | 22.23 | 70.69 |
| | 0.5 | 0.2 | 0.0501 | 0.9167 | 50.60 | 2.97 | 31.34 |
| | 1 | 0.5 | 0.0510 | 1 | 50 | 0 | 7.86 |
| 100 | 0.1429 | 0.02 | 0.0526 | 0.8085 | 399.72 | 79.29 | 384.58 |
| | 0.2294 | 0.05 | 0.0545 | 0.8079 | 155.90 | 31.57 | 149.03 |
| | 0.3333 | 0.1 | 0.0506 | 0.9033 | 100.56 | 3.15 | 70.69 |
| | 0.5 | 0.2 | 0.0505 | 0.9960 | 100 | 0 | 31.34 |
| | 1 | 0.5 | 0.0495 | 1 | 100 | 0 | 7.86 |

Table 31: Linear regression with two continuous predictors, Pearson correlation between the predictors = 0.4, $N_{max} = 600$.

| n | $\beta_1$ | $R^2$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *Target SS* |
|---|---|---|---|---|---|---|---|
| 20 | 0.1429 | 0.02 | 0.0505 | 0.7793 | 471.79 | 134.14 | 457.70 |
| | 0.2294 | 0.05 | 0.0494 | 0.8104 | 221.17 | 109.63 | 177.49 |
| | 0.3333 | 0.1 | 0.0531 | 0.8102 | 105.96 | 55.61 | 84.19 |
| | 0.5 | 0.2 | 0.0559 | 0.8173 | 47.41 | 24.39 | 37.45 |
| | 1 | 0.5 | 0.0532 | 0.9451 | 20.57 | 2.78 | 9.34 |
| 50 | 0.1429 | 0.02 | 0.0499 | 0.7942 | 472.35 | 104.28 | 457.70 |
| | 0.2294 | 0.05 | 0.0520 | 0.8122 | 193.13 | 57.91 | 177.49 |
| | 0.3333 | 0.1 | 0.0521 | 0.8121 | 91.91 | 27.17 | 84.19 |
| | 0.5 | 0.2 | 0.0506 | 0.8830 | 51.95 | 5.60 | 37.45 |
| | 1 | 0.5 | 0.0506 | 0.9999 | 50 | 0 | 9.34 |
| 100 | 0.1429 | 0.02 | 0.0514 | 0.8074 | 470.18 | 83.33 | 457.70 |
| | 0.2294 | 0.05 | 0.0516 | 0.8076 | 185.41 | 38.42 | 177.49 |
| | 0.3333 | 0.1 | 0.0491 | 0.8557 | 102.99 | 7.86 | 84.19 |
| | 0.5 | 0.2 | 0.0513 | 0.9907 | 100 | 0 | 37.45 |
| | 1 | 0.5 | 0.0502 | 1 | 100 | 0 | 9.34 |

Table 32: Linear regression with two continuous predictors, Pearson correlation between the predictors = 0.8, $N_{max} = 600$.

| n | $\beta_1$ | $R^2$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *Target SS* |
|---|---|---|---|---|---|---|---|
| 20 | 0.1429 | 0.02 | 0.0509 | 0.5534 | 590.44 | 40.42 | 1067.68 |
| | 0.2294 | 0.05 | 0.0503 | 0.7912 | 446.91 | 139.74 | 414.85 |
| | 0.3333 | 0.1 | 0.0517 | 0.8068 | 242.66 | 117.20 | 196.10 |
| | 0.5 | 0.2 | 0.0522 | 0.8068 | 109.30 | 57.01 | 87.23 |
| | 1 | 0.5 | 0.0535 | 0.8392 | 29.29 | 12.71 | 21.79 |
| 50 | 0.1429 | 0.02 | 0.0521 | 0.5502 | 598.99 | 9.99 | 1067.68 |
| | 0.2294 | 0.05 | 0.0515 | 0.8037 | 437.41 | 107.55 | 414.85 |
| | 0.3333 | 0.1 | 0.0522 | 0.8033 | 213.53 | 63.86 | 196.10 |
| | 0.5 | 0.2 | 0.0517 | 0.8129 | 95.45 | 28.44 | 87.23 |
| | 1 | 0.5 | 0.0515 | 0.9732 | 50.02 | 0.46 | 21.79 |
| 100 | 0.1429 | 0.02 | 0.0494 | 0.5543 | 599.94 | 1.99 | 1067.68 |
| | 0.2294 | 0.05 | 0.0492 | 0.8053 | 428.91 | 82.76 | 414.85 |
| | 0.3333 | 0.1 | 0.0508 | 0.8051 | 204.99 | 42.03 | 196.10 |
| | 0.5 | 0.2 | 0.0509 | 0.8556 | 104.20 | 9.44 | 87.23 |
| | 1 | 0.5 | 0.0493 | 0.9997 | 100 | 0 | 21.79 |

Table 33: Linear regression with two independent binary predictors, $N_{max} = 600$.

| n | $\beta_1$ | $R^2$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *Target SS* |
|---|---|---|---|---|---|---|---|
| 20 | 0.1429 | 0.02 | 0.0512 | 0.7967 | 417.00 | 124.32 | 384.36 |
| | 0.2294 | 0.05 | 0.0516 | 0.8007 | 168.73 | 62.59 | 149.15 |
| | 0.3333 | 0.1 | 0.0540 | 0.8017 | 80.14 | 29.86 | 70.65 |
| | 0.5 | 0.2 | 0.0595 | 0.8030 | 36.26 | 13.12 | 31.40 |
| | 1 | 0.5 | 0.0506 | 0.9785 | 20.03 | 0.55 | 7.85 |
| 50 | 0.1429 | 0.02 | 0.0503 | 0.8063 | 401.24 | 82.34 | 384.36 |
| | 0.2294 | 0.05 | 0.0510 | 0.7995 | 156.26 | 32.82 | 149.15 |
| | 0.3333 | 0.1 | 0.0541 | 0.7971 | 74.43 | 15.26 | 70.65 |
| | 0.5 | 0.2 | 0.0508 | 0.9198 | 50.05 | 0.58 | 31.40 |
| | 1 | 0.5 | 0.0491 | 1 | 50 | 0 | 7.85 |
| 100 | 0.1429 | 0.02 | 0.0514 | 0.8066 | 393.08 | 57.08 | 384.36 |
| | 0.2294 | 0.05 | 0.0523 | 0.7935 | 152.89 | 22.01 | 149.15 |
| | 0.3333 | 0.1 | 0.0482 | 0.9058 | 100.04 | 0.65 | 70.65 |
| | 0.5 | 0.2 | 0.0497 | 0.9980 | 100 | 0 | 31.40 |
| | 1 | 0.5 | 0.0501 | 1 | 100 | 0 | 7.85 |

Table 34: Linear regression with two binary predictors, Pearson correlation between the predictors = 0.4 (tetrachoric correlation = 0.5878), $N_{max} = 600$.

| n | $\beta_1$ | $R^2$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *Target SS* |
|---|---|---|---|---|---|---|---|
| 20 | 0.1429 | 0.02 | 0.0506 | 0.7843 | 468.85 | 126.30 | 457.61 |
| | 0.2294 | 0.05 | 0.0521 | 0.8026 | 210.80 | 97.72 | 177.69 |
| | 0.3333 | 0.1 | 0.0552 | 0.8011 | 101.12 | 53.08 | 84.18 |
| | 0.5 | 0.2 | 0.0561 | 0.8098 | 45.66 | 28.21 | 37.40 |
| | 1 | 0.5 | 0.0512 | 0.9490 | 20.89 | 15.82 | 9.35 |
| 50 | 0.1429 | 0.02 | 0.0502 | 0.7947 | 470.98 | 94.03 | 457.61 |
| | 0.2294 | 0.05 | 0.0514 | 0.8076 | 189.06 | 48.73 | 177.69 |
| | 0.3333 | 0.1 | 0.0520 | 0.8020 | 90.03 | 23.23 | 84.18 |
| | 0.5 | 0.2 | 0.0499 | 0.8763 | 51.22 | 4.62 | 37.40 |
| | 1 | 0.5 | 0.0492 | 0.9999 | 50 | 0 | 9.35 |
| 100 | 0.1429 | 0.02 | 0.0503 | 0.8040 | 467.85 | 73.70 | 457.61 |
| | 0.2294 | 0.05 | 0.0511 | 0.8013 | 182.93 | 31.53 | 177.69 |
| | 0.3333 | 0.1 | 0.0513 | 0.8559 | 101.96 | 5.92 | 84.18 |
| | 0.5 | 0.2 | 0.0502 | 0.9927 | 100 | 0 | 37.40 |
| | 1 | 0.5 | 0.0502 | 1 | 100 | 0 | 9.35 |

Table 35: Linear regression with two binary predictors, Pearson correlation between the predictors = 0.8 (tetrachoric correlation = 0.9511), $N_{max} = 600$.

| n | $\beta_1$ | $R^2$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *Target SS* |
|---|---|---|---|---|---|---|---|
| 20 | 0.1429 | 0.02 | 0.0488 | 0.5356 | 580.94 | 59.14 | 1066.30 |
| | 0.2294 | 0.05 | 0.0494 | 0.7755 | 438.90 | 159.44 | 415.22 |
| | 0.3333 | 0.1 | 0.0516 | 0.8008 | 279.93 | 174.06 | 196.77 |
| | 0.5 | 0.2 | 0.0550 | 0.8075 | 164.96 | 171.06 | 87.35 |
| | 1 | 0.5 | 0.0537 | 0.8681 | 99.02 | 187.21 | 21.83 |
| 50 | 0.1429 | 0.02 | 0.0489 | 0.5516 | 596.15 | 21.37 | 1066.30 |
| | 0.2294 | 0.05 | 0.0506 | 0.7886 | 437.39 | 131.00 | 415.22 |
| | 0.3333 | 0.1 | 0.0510 | 0.8110 | 240.39 | 122.88 | 196.77 |
| | 0.5 | 0.2 | 0.0513 | 0.8178 | 112.82 | 76.79 | 87.35 |
| | 1 | 0.5 | 0.0506 | 0.9613 | 54.76 | 40.10 | 21.83 |
| 100 | 0.1429 | 0.02 | 0.0506 | 0.5543 | 599.43 | 6.79 | 1066.30 |
| | 0.2294 | 0.05 | 0.0510 | 0.7983 | 435.02 | 109.04 | 415.22 |
| | 0.3333 | 0.1 | 0.0509 | 0.8118 | 217.72 | 78.99 | 196.77 |
| | 0.5 | 0.2 | 0.0505 | 0.8623 | 112.13 | 29.27 | 87.35 |
| | 1 | 0.5 | 0.0506 | 0.9982 | 100.06 | 3.61 | 21.83 |

Table 36: Linear regression with two independent predictors (binary and continuous), the predictor of interest is binary, $N_{max} = 600$.

| n | $\beta_1$ | $R^2$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *Target SS* |
|---|---|---|---|---|---|---|---|
| 20 | 0.1429 | 0.02 | 0.0509 | 0.8013 | 416.54 | 125.08 | 384.36 |
| | 0.2294 | 0.05 | 0.0522 | 0.8019 | 168.55 | 62.07 | 149.15 |
| | 0.3333 | 0.1 | 0.0541 | 0.8054 | 80.06 | 29.52 | 70.65 |
| | 0.5 | 0.2 | 0.0594 | 0.8045 | 36.12 | 12.98 | 31.40 |
| | 1 | 0.5 | 0.0483 | 0.9769 | 20.02 | 0.33 | 7.85 |
| 50 | 0.1429 | 0.02 | 0.0509 | 0.8018 | 400.53 | 82.43 | 384.36 |
| | 0.2294 | 0.05 | 0.0514 | 0.7943 | 156.20 | 33.01 | 149.15 |
| | 0.3333 | 0.1 | 0.0521 | 0.8144 | 74.35 | 15.21 | 70.65 |
| | 0.5 | 0.2 | 0.0498 | 0.9225 | 50.05 | 0.61 | 31.40 |
| | 1 | 0.5 | 0.0505 | 1 | 50 | 0 | 7.85 |
| 100 | 0.1429 | 0.02 | 0.0514 | 0.7984 | 392.88 | 56.97 | 384.36 |
| | 0.2294 | 0.05 | 0.0526 | 0.8035 | 152.72 | 22.08 | 149.15 |
| | 0.3333 | 0.1 | 0.0513 | 0.9030 | 100.03 | 0.54 | 70.65 |
| | 0.5 | 0.2 | 0.0510 | 0.9977 | 100 | 0 | 31.40 |
| | 1 | 0.5 | 0.0513 | 1 | 100 | 0 | 7.85 |

Table 37: Linear regression with two predictors (binary and continuous), the predictor of interest is binary, Pearson correlation between the predictors = 0.4 (original correlation = 0.5037), $N_{max} = 600$.

| n | $\beta_1$ | $R^2$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *Target SS* |
|---|---|---|---|---|---|---|---|
| 20 | 0.1429 | 0.02 | 0.0500 | 0.7803 | 467.44 | 124.51 | 458.25 |
| | 0.2294 | 0.05 | 0.0510 | 0.7954 | 204.42 | 85.60 | 177.87 |
| | 0.3333 | 0.1 | 0.0551 | 0.8000 | 96.91 | 40.78 | 84.21 |
| | 0.5 | 0.2 | 0.0590 | 0.8003 | 43.49 | 18.02 | 37.43 |
| | 1 | 0.5 | 0.0512 | 0.9483 | 20.20 | 1.34 | 9.37 |
| 50 | 0.1429 | 0.02 | 0.0510 | 0.7975 | 469.85 | 92.97 | 458.25 |
| | 0.2294 | 0.05 | 0.0507 | 0.8068 | 187.32 | 45.22 | 177.87 |
| | 0.3333 | 0.1 | 0.0547 | 0.8020 | 89.01 | 21.48 | 84.21 |
| | 0.5 | 0.2 | 0.0503 | 0.8845 | 50.92 | 3.33 | 37.43 |
| | 1 | 0.5 | 0.0509 | 1 | 50 | 0 | 9.37 |
| 100 | 0.1429 | 0.02 | 0.0494 | 0.7993 | 468.17 | 71.44 | 458.25 |
| | 0.2294 | 0.05 | 0.0507 | 0.7966 | 182.48 | 30.17 | 177.87 |
| | 0.3333 | 0.1 | 0.0493 | 0.8565 | 101.61 | 5.00 | 84.21 |
| | 0.5 | 0.2 | 0.0516 | 0.9947 | 100 | 0 | 37.43 |
| | 1 | 0.5 | 0.0501 | 1 | 100 | 0 | 9.37 |

Table 38: Linear regression with two predictors (binary and continuous), the predictor of interest is binary, Pearson correlation between the predictors = 0.8 (original correlation = 0.9991), $N_{max} = 600$.

| n | $\beta_1$ | $R^2$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *Target SS* |
|---|---|---|---|---|---|---|---|
| 20 | 0.1429 | 0.02 | 0.0499 | 0.5597 | 592.57 | 34.48 | 1052.06 |
| | 0.2294 | 0.05 | 0.0502 | 0.7986 | 439.55 | 132.44 | 408.97 |
| | 0.3333 | 0.1 | 0.0532 | 0.8059 | 227.40 | 96.76 | 193.72 |
| | 0.5 | 0.2 | 0.0544 | 0.8062 | 101.77 | 44.21 | 86.16 |
| | 1 | 0.5 | 0.0535 | 0.8262 | 27.27 | 9.64 | 21.54 |
| 50 | 0.1429 | 0.02 | 0.0508 | 0.5616 | 599.47 | 6.95 | 1052.06 |
| | 0.2294 | 0.05 | 0.0504 | 0.8030 | 429.53 | 96.27 | 408.97 |
| | 0.3333 | 0.1 | 0.0509 | 0.8034 | 206.24 | 52.01 | 193.72 |
| | 0.5 | 0.2 | 0.0540 | 0.7980 | 91.78 | 22.78 | 86.16 |
| | 1 | 0.5 | 0.0502 | 0.9807 | 50.00 | 0.17 | 21.54 |
| 100 | 0.1429 | 0.02 | 0.0496 | 0.5531 | 599.98 | 1.22 | 1052.06 |
| | 0.2294 | 0.05 | 0.0509 | 0.8003 | 420.95 | 71.04 | 408.97 |
| | 0.3333 | 0.1 | 0.0509 | 0.8046 | 199.91 | 34.65 | 193.72 |
| | 0.5 | 0.2 | 0.0508 | 0.8505 | 102.42 | 6.38 | 86.16 |
| | 1 | 0.5 | 0.0507 | 0.9999 | 100 | 0 | 21.54 |

Table 39: Linear regression with independent predictors (binary and continuous), the predictor of interest is continuous, $N_{max} = 600$.

| n | $\beta_1$ | $R^2$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *Target SS* |
|---|---|---|---|---|---|---|---|
| 20 | 0.1429 | 0.02 | 0.0499 | 0.7926 | 425.89 | 143.32 | 384.02 |
| | 0.2294 | 0.05 | 0.0519 | 0.8044 | 185.99 | 94.33 | 149.09 |
| | 0.3333 | 0.1 | 0.0544 | 0.8128 | 89.04 | 46.87 | 70.56 |
| | 0.5 | 0.2 | 0.0569 | 0.8174 | 40.33 | 20.26 | 31.34 |
| | 1 | 0.5 | 0.0495 | 0.9599 | 20.26 | 1.71 | 7.85 |
| 50 | 0.1429 | 0.02 | 0.0510 | 0.8077 | 411.45 | 107.29 | 384.02 |
| | 0.2294 | 0.05 | 0.0526 | 0.8087 | 162.59 | 48.58 | 149.09 |
| | 0.3333 | 0.1 | 0.0535 | 0.8173 | 77.73 | 22.27 | 70.56 |
| | 0.5 | 0.2 | 0.0521 | 0.9152 | 50.61 | 2.96 | 31.34 |
| | 1 | 0.5 | 0.0494 | 1 | 50 | 0 | 7.85 |
| 100 | 0.1429 | 0.02 | 0.0512 | 0.8056 | 400.12 | 79.69 | 384.02 |
| | 0.2294 | 0.05 | 0.0508 | 0.8029 | 156.00 | 31.58 | 149.09 |
| | 0.3333 | 0.1 | 0.0494 | 0.8996 | 100.52 | 2.93 | 70.56 |
| | 0.5 | 0.2 | 0.0510 | 0.9973 | 100 | 0 | 31.34 |
| | 1 | 0.5 | 0.0473 | 1 | 100 | 0 | 7.85 |

Table 40: Linear regression with two predictors (continuous and binary), the predictor of interest is continuous, Pearson correlation between the predictors = 0.4 (original correlation = 0.5037), $N_{max} = 600$.

| n | $\beta_1$ | $R^2$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *Target SS* |
|---|---|---|---|---|---|---|---|
| 20 | 0.1429 | 0.02 | 0.0498 | 0.7779 | 473.51 | 133.43 | 458.75 |
| | 0.2294 | 0.05 | 0.0500 | 0.8054 | 221.86 | 110.63 | 178.24 |
| | 0.3333 | 0.1 | 0.0543 | 0.8147 | 105.73 | 55.50 | 84.22 |
| | 0.5 | 0.2 | 0.0578 | 0.8163 | 47.54 | 24.33 | 37.45 |
| | 1 | 0.5 | 0.0492 | 0.9405 | 20.58 | 2.70 | 9.37 |
| 50 | 0.1429 | 0.02 | 0.0497 | 0.7912 | 473.49 | 103.89 | 458.75 |
| | 0.2294 | 0.05 | 0.0506 | 0.8035 | 193.86 | 57.85 | 178.24 |
| | 0.3333 | 0.1 | 0.0544 | 0.8156 | 92.30 | 27.29 | 84.22 |
| | 0.5 | 0.2 | 0.0509 | 0.8818 | 52.01 | 5.72 | 37.45 |
| | 1 | 0.5 | 0.0499 | 0.9998 | 50 | 0 | 9.37 |
| 100 | 0.1429 | 0.02 | 0.0510 | 0.8016 | 471.30 | 83.59 | 458.75 |
| | 0.2294 | 0.05 | 0.0513 | 0.8095 | 185.78 | 38.04 | 178.24 |
| | 0.3333 | 0.1 | 0.0506 | 0.8584 | 103.07 | 7.94 | 84.22 |
| | 0.5 | 0.2 | 0.0515 | 0.9917 | 100 | 0 | 37.45 |
| | 1 | 0.5 | 0.0494 | 1 | 100 | 0 | 9.37 |

Table 41: Linear regression with two predictors (continuous and binary), the predictor of interest is continuous, Pearson correlation between the predictors = 0.8 (original correlation = 0.9991), $N_{max} = 600$.

| n | $\beta_1$ | $R^2$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *Target SS* |
|---|---|---|---|---|---|---|---|
| 20 | 0.1429 | 0.02 | 0.0503 | 0.5393 | 587.54 | 47.79 | 1052.62 |
| | 0.2294 | 0.05 | 0.0500 | 0.7845 | 444.78 | 144.96 | 408.51 |
| | 0.3333 | 0.1 | 0.0511 | 0.8112 | 248.33 | 127.12 | 193.89 |
| | 0.5 | 0.2 | 0.0532 | 0.8135 | 112.36 | 63.87 | 86.19 |
| | 1 | 0.5 | 0.0544 | 0.8454 | 30.35 | 14.85 | 21.53 |
| 50 | 0.1429 | 0.02 | 0.0503 | 0.5561 | 598.17 | 14.05 | 1052.62 |
| | 0.2294 | 0.05 | 0.0479 | 0.8041 | 435.23 | 113.23 | 408.51 |
| | 0.3333 | 0.1 | 0.0512 | 0.8024 | 214.37 | 70.36 | 193.89 |
| | 0.5 | 0.2 | 0.0527 | 0.8167 | 96.16 | 31.40 | 86.19 |
| | 1 | 0.5 | 0.0496 | 0.9715 | 50.05 | 0.82 | 21.53 |
| 100 | 0.1429 | 0.02 | 0.0516 | 0.5500 | 599.83 | 3.37 | 1052.62 |
| | 0.2294 | 0.05 | 0.0498 | 0.8056 | 427.37 | 89.19 | 408.51 |
| | 0.3333 | 0.1 | 0.0510 | 0.8036 | 204.12 | 46.11 | 193.89 |
| | 0.5 | 0.2 | 0.0515 | 0.8523 | 104.67 | 10.64 | 86.19 |
| | 1 | 0.5 | 0.0495 | 1 | 100 | 0 | 21.53 |

Table 42: Linear regression with 10 independent continuous predictors, $N_{max} = 600$.

| n | $\beta_1$ | $R^2$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *Target SS* |
|---|---|---|---|---|---|---|---|
| 20 | 0.1429 | 0.02 | 0.0484 | 0.8736 | 528.34 | 121.98 | 384.49 |
| | 0.2294 | 0.05 | 0.0514 | 0.9072 | 327.58 | 168.39 | 149.09 |
| | 0.3333 | 0.1 | 0.0527 | 0.8976 | 173.25 | 118.97 | 70.71 |
| | 0.5 | 0.2 | 0.0578 | 0.8867 | 79.36 | 60.32 | 31.39 |
| | 1 | 0.5 | 0.0562 | 0.8956 | 25.19 | 12.33 | 7.86 |
| 50 | 0.1429 | 0.02 | 0.0523 | 0.8466 | 472.97 | 109.11 | 384.49 |
| | 0.2294 | 0.05 | 0.0529 | 0.8510 | 197.32 | 65.35 | 149.09 |
| | 0.3333 | 0.1 | 0.0520 | 0.8288 | 93.65 | 30.40 | 70.71 |
| | 0.5 | 0.2 | 0.0515 | 0.8882 | 52.59 | 6.93 | 31.39 |
| | 1 | 0.5 | 0.0513 | 0.9997 | 50 | 0 | 7.86 |
| 100 | 0.1429 | 0.02 | 0.0507 | 0.8287 | 434.37 | 85.86 | 384.49 |
| | 0.2294 | 0.05 | 0.0540 | 0.8243 | 169.94 | 36.50 | 149.09 |
| | 0.3333 | 0.1 | 0.0502 | 0.8768 | 101.58 | 5.72 | 70.71 |
| | 0.5 | 0.2 | 0.0516 | 0.9956 | 100 | 0 | 31.39 |
| | 1 | 0.5 | 0.0517 | 1 | 100 | 0 | 7.86 |

Table 43: Linear regression with 10 continuous predictors, Pearson correlation among the predictors = 0.4, $N_{max} = 600$.

| n | $\beta_1$ | $R^2$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *Target SS* |
|---|---|---|---|---|---|---|---|
| 20 | 0.1429 | 0.02 | 0.0495 | 0.7802 | 572.13 | 0.91 | 584.82 |
| | 0.2294 | 0.05 | 0.0529 | 0.9077 | 429.29 | 164.76 | 227.17 |
| | 0.3333 | 0.1 | 0.0516 | 0.9082 | 253.38 | 152.89 | 107.53 |
| | 0.5 | 0.2 | 0.0581 | 0.8898 | 119.05 | 88.63 | 47.72 |
| | 1 | 0.5 | 0.0617 | 0.8777 | 32.99 | 21.82 | 11.94 |
| 50 | 0.1429 | 0.02 | 0.0495 | 0.7800 | 572,32 | 59.97 | 584.82 |
| | 0.2294 | 0.05 | 0.0511 | 0.8562 | 297.90 | 95.60 | 227.17 |
| | 0.3333 | 0.1 | 0.0535 | 0.8450 | 142.12 | 46.91 | 107.53 |
| | 0.5 | 0.2 | 0.0536 | 0.8332 | 65.92 | 18.54 | 47.72 |
| | 1 | 0.5 | 0.0493 | 0.9963 | 50.00 | 0.10 | 11.94 |
| 100 | 0.1429 | 0.02 | 0.0526 | 0.7817 | 573.08 | 49.41 | 584.82 |
| | 0.2294 | 0.05 | 0.0501 | 0.8235 | 258.38 | 55.40 | 227.17 |
| | 0.3333 | 0.1 | 0.0514 | 0.8148 | 124.76 | 23.63 | 107.53 |
| | 0.5 | 0.2 | 0.0503 | 0.9628 | 100.01 | 0.36 | 47.72 |
| | 1 | 0.5 | 0.0493 | 1 | 100 | 0 | 11.94 |

Table 44: Linear regression with 10 continuous predictors, Pearson correlation among the predictors = 0.8, $N_{max} = 600$.

| n | $\beta_1$ | $R^2$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *Target SS* |
|---|---|---|---|---|---|---|---|
| 20 | 0.1429 | 0.02 | 0.0512 | 0.3728 | 598.99 | 14.73 | 1732.42 |
| | 0.2294 | 0.05 | 0.0491 | 0.7231 | 580.66 | 65.67 | 672.42 |
| | 0.3333 | 0.1 | 0.0496 | 0.8916 | 496.99 | 141.12 | 318.95 |
| | 0.5 | 0.2 | 0.0510 | 0.9069 | 314.68 | 166.36 | 141.67 |
| | 1 | 0.5 | 0.0592 | 0.8886 | 88.91 | 67.60 | 35.28 |
| 50 | 0.1429 | 0.02 | 0.0495 | 0.3684 | 599.98 | 0.42 | 1732.42 |
| | 0.2294 | 0.05 | 0.0512 | 0.7257 | 587.30 | 40.10 | 672.42 |
| | 0.3333 | 0.1 | 0.0513 | 0.8534 | 408.87 | 113.98 | 318.95 |
| | 0.5 | 0.2 | 0.0523 | 0.8449 | 186.81 | 61.55 | 141.67 |
| | 1 | 0.5 | 0.0502 | 0.8596 | 54.72 | 9.65 | 35.28 |
| 100 | 0.1429 | 0.02 | 0.0497 | 0.3737 | 600 | 0 | 1732.42 |
| | 0.2294 | 0.05 | 0.0516 | 0.7388 | 590.78 | 28.75 | 672.42 |
| | 0.3333 | 0.1 | 0.0504 | 0.8199 | 362.12 | 76.39 | 318.95 |
| | 0.5 | 0.2 | 0.0521 | 0.8179 | 161.73 | 34.67 | 141.67 |
| | 1 | 0.5 | 0.0486 | 0.9901 | 100 | 0 | 35.28 |

Table 45: Linear regression with 10 independent binary predictors, $N_{max} = 600$.

| n | $\beta_1$ | $R^2$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *Target SS* |
|---|---|---|---|---|---|---|---|
| 20 | 0.1429 | 0.02 | 0.0488 | 0.8750 | 531.06 | 116.53 | 384.37 |
| | 0.2294 | 0.05 | 0.0499 | 0.9081 | 314.08 | 155.09 | 149.15 |
| | 0.3333 | 0.1 | 0.0551 | 0.8944 | 159.13 | 100.13 | 70.65 |
| | 0.5 | 0.2 | 0.0637 | 0.8776 | 71.85 | 41.79 | 31.40 |
| | 1 | 0.5 | 0.0536 | 0.8957 | 23.63 | 10.96 | 7.85 |
| 50 | 0.1429 | 0.02 | 0.0503 | 0.8524 | 472.04 | 95.59 | 384.37 |
| | 0.2294 | 0.05 | 0.0524 | 0.8462 | 188.96 | 47.64 | 149.15 |
| | 0.3333 | 0.1 | 0.0542 | 0.8319 | 89.70 | 22.42 | 70.65 |
| | 0.5 | 0.2 | 0.0488 | 0.8795 | 51.10 | 3.67 | 31.40 |
| | 1 | 0.5 | 0.0499 | 0.9999 | 50 | 0 | 7.85 |
| 100 | 0.1429 | 0.02 | 0.0520 | 0.8253 | 428.85 | 66.77 | 384.37 |
| | 0.2294 | 0.05 | 0.0516 | 0.8152 | 166.48 | 26.20 | 149.15 |
| | 0.3333 | 0.1 | 0.0489 | 0.8762 | 100.38 | 2.15 | 70.65 |
| | 0.5 | 0.2 | 0.0527 | 0.9973 | 100 | 0 | 31.40 |
| | 1 | 0.5 | 0.0512 | 1 | 100 | 0 | 7.85 |

Table 46: Linear regression with 10 binary predictors, Pairwise Pearson correlations = 0.4 (pairwise tetrachoric correlations = 0.5878), $N_{max} = 600$.

| n | $\beta_1$ | $R^2$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *Target SS* |
|---|---|---|---|---|---|---|---|
| 20 | 0.1429 | 0.02 | 0.0498 | 0.7826 | 579.39 | 68.25 | 584.24 |
| | 0.2294 | 0.05 | 0.0515 | 0.9273 | 463.30 | 159.96 | 227.21 |
| | 0.3333 | 0.1 | 0.0510 | 0.9219 | 309.42 | 182.23 | 107.63 |
| | 0.5 | 0.2 | 0.0555 | 0.9163 | 174.20 | 158.27 | 47.75 |
| | 1 | 0.5 | 0.0589 | 0.8981 | 68.03 | 121.83 | 11.93 |
| 50 | 0.1429 | 0.02 | 0.0489 | 0.7764 | 578.54 | 51.31 | 584.24 |
| | 0.2294 | 0.05 | 0.0517 | 0.8627 | 313.36 | 103.96 | 227.21 |
| | 0.3333 | 0.1 | 0.0536 | 0.8550 | 150.03 | 53.89 | 107.63 |
| | 0.5 | 0.2 | 0.0526 | 0.8431 | 68.89 | 22.11 | 47.75 |
| | 1 | 0.5 | 0.0514 | 0.9945 | 50.01 | 0.53 | 11.93 |
| 100 | 0.1429 | 0.02 | 0.0488 | 0.7766 | 577.12 | 45.08 | 584.24 |
| | 0.2294 | 0.05 | 0.0514 | 0.8228 | 263.98 | 56.73 | 227.21 |
| | 0.3333 | 0.1 | 0.0516 | 0.8184 | 126.83 | 24.71 | 107.63 |
| | 0.5 | 0.2 | 0.0505 | 0.9554 | 100.03 | 0.67 | 47.75 |
| | 1 | 0.5 | 0.0500 | 1 | 100 | 0 | 11.93 |

Table 47: Linear regression with 10 binary predictors, Pairwise Pearson correlations = 0.8 (pairwise tetrachoric correlations = 0.9511), $N_{max} = 600$.

| n | $\beta_1$ | $R^2$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *Target SS* |
|---|---|---|---|---|---|---|---|
| 20 | 0.1429 | 0.02 | 0.0499 | 0.3601 | 599.96 | 2.85 | 1736.86 |
| | 0.2294 | 0.05 | 0.0506 | 0.7263 | 599.27 | 14.03 | 672.07 |
| | 0.3333 | 0.1 | 0.0500 | 0.9541 | 597.54 | 28.33 | 318.79 |
| | 0.5 | 0.2 | 0.0483 | 0.9981 | 594.24 | 47.93 | 141.61 |
| | 1 | 0.5 | 0.0503 | 0.9981 | 589.47 | 72.06 | 35.47 |
| 50 | 0.1429 | 0.02 | 0.0488 | 0.3694 | 599.98 | 1.59 | 1736.86 |
| | 0.2294 | 0.05 | 0.0506 | 0.7201 | 593.12 | 33.73 | 672.07 |
| | 0.3333 | 0.1 | 0.0521 | 0.9191 | 531.88 | 115.56 | 318.79 |
| | 0.5 | 0.2 | 0.0505 | 0.9391 | 399.27 | 187.85 | 141.61 |
| | 1 | 0.5 | 0.0515 | 0.9291 | 223.57 | 220.59 | 35.47 |
| 100 | 0.1429 | 0.02 | 0.0492 | 0.3619 | 600 | 0 | 1736.86 |
| | 0.2294 | 0.05 | 0.0508 | 0.7235 | 591.46 | 31.96 | 672.07 |
| | 0.3333 | 0.1 | 0.0514 | 0.8729 | 454.91 | 124.73 | 318.79 |
| | 0.5 | 0.2 | 0.0521 | 0.8680 | 234.45 | 112.54 | 141.61 |
| | 1 | 0.5 | 0.0505 | 0.9566 | 104.80 | 28.53 | 35.47 |

Table 48: Linear regression with 10 independent predictors, the predictor of interest is binary, other 9 predictors are continuous, $N_{max} = 600$.

| n | $\beta_1$ | $R^2$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *Target SS* |
|---|---|---|---|---|---|---|---|
| 20 | 0.1429 | 0.02 | 0.0494 | 0.8826 | 531.53 | 116.32 | 384.37 |
| | 0.2294 | 0.05 | 0.0522 | 0.9099 | 313.67 | 155.33 | 149.15 |
| | 0.3333 | 0.1 | 0.0530 | 0.9019 | 158.30 | 97.95 | 70.65 |
| | 0.5 | 0.2 | 0.0630 | 0.8823 | 71.56 | 46.96 | 31.40 |
| | 1 | 0.5 | 0.0563 | 0.8939 | 23.51 | 8.97 | 7.85 |
| 50 | 0.1429 | 0.02 | 0.0520 | 0.8526 | 471.90 | 96.14 | 384.37 |
| | 0.2294 | 0.05 | 0.0529 | 0.8415 | 189.08 | 47.65 | 149.15 |
| | 0.3333 | 0.1 | 0.0552 | 0.8281 | 89.78 | 22.39 | 70.65 |
| | 0.5 | 0.2 | 0.0485 | 0.8821 | 51.07 | 3.58 | 31.40 |
| | 1 | 0.5 | 0.0504 | 1 | 50 | 0 | 7.85 |
| 100 | 0.1429 | 0.02 | 0.0505 | 0.8212 | 428.06 | 66.46 | 384.37 |
| | 0.2294 | 0.05 | 0.0509 | 0.8112 | 166.60 | 26.37 | 149.15 |
| | 0.3333 | 0.1 | 0.0496 | 0.8830 | 100.37 | 2.11 | 70.65 |
| | 0.5 | 0.2 | 0.0509 | 0.9971 | 100 | 0 | 31.40 |
| | 1 | 0.5 | 0.0489 | 1 | 100 | 0 | 7.85 |

Table 49: Linear regression with 10 predictors, the predictor of interest is binary, other 9 predictors are continuous, pairwise Pearson correlation = 0.4 (original pairwise correlation = 0.5037), $N_{max} = 600$.

| n | $\beta_1$ | $R^2$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *Target SS* |
|---|---|---|---|---|---|---|---|
| 20 | 0.1429 | 0.02 | 0.0507 | 0.7801 | 573.16 | 76.29 | 587.90 |
| | 0.2294 | 0.05 | 0.0496 | 0.9034 | 420.60 | 162.25 | 228.31 |
| | 0.3333 | 0.1 | 0.0501 | 0.8996 | 240.03 | 142.39 | 108.14 |
| | 0.5 | 0.2 | 0.0572 | 0.8932 | 112.35 | 78.99 | 48.04 |
| | 1 | 0.5 | 0.0606 | 0.8726 | 30.87 | 18.65 | 12.00 |
| 50 | 0.1429 | 0.02 | 0.0505 | 0.7913 | 575.95 | 53.68 | 587.90 |
| | 0.2294 | 0.05 | 0.0533 | 0.8498 | 292.03 | 85.03 | 228.31 |
| | 0.3333 | 0.1 | 0.0528 | 0.8467 | 138.21 | 40.74 | 108.14 |
| | 0.5 | 0.2 | 0.0544 | 0.8337 | 63.88 | 15.81 | 48.04 |
| | 1 | 0.5 | 0.0497 | 0.9977 | 50.00 | 0.03 | 12.00 |
| 100 | 0.1429 | 0.02 | 0.0493 | 0.7841 | 577.05 | 43.63 | 587.90 |
| | 0.2294 | 0.05 | 0.0516 | 0.8244 | 255.84 | 48.56 | 228.31 |
| | 0.3333 | 0.1 | 0.0521 | 0.8156 | 123.22 | 20.78 | 108.14 |
| | 0.5 | 0.2 | 0.0490 | 0.9642 | 100.00 | 0.22 | 48.04 |
| | 1 | 0.5 | 0.0508 | 1 | 100 | 0 | 12.00 |

Table 50: Linear regression with 10 predictors, the predictor of interest is binary, other 9 predictors are continuous, pairwise Pearson correlation = 0.72 (original pairwise correlation = 0.9), $N_{max} = 600$.

| n | $\beta_1$ | $R^2$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *Target SS* |
|---|---|---|---|---|---|---|---|
| 20 | 0.1429 | 0.02 | 0.0509 | 0.5523 | 595.02 | 32.82 | 1030.71 |
| | 0.2294 | 0.05 | 0.0503 | 0.8642 | 534.99 | 115.03 | 400.25 |
| | 0.3333 | 0.1 | 0.0527 | 0.9072 | 379.71 | 164.81 | 189.36 |
| | 0.5 | 0.2 | 0.0551 | 0.9016 | 194.16 | 122.53 | 84.07 |
| | 1 | 0.5 | 0.0648 | 0.8718 | 50.51 | 34.79 | 21.07 |
| 50 | 0.1429 | 0.02 | 0.0503 | 0.5554 | 599.75 | 4.92 | 1030.71 |
| | 0.2294 | 0.05 | 0.0509 | 0.8446 | 485.32 | 100.09 | 400.25 |
| | 0.3333 | 0.1 | 0.0540 | 0.8500 | 243.47 | 70.03 | 189.36 |
| | 0.5 | 0.2 | 0.0529 | 0.8365 | 108.56 | 31.28 | 84.07 |
| | 1 | 0.5 | 0.0508 | 0.9585 | 50.07 | 0.88 | 21.07 |
| 100 | 0.1429 | 0.02 | 0.0504 | 0.5611 | 599.99 | 0.80 | 1030.71 |
| | 0.2294 | 0.05 | 0.0518 | 0.8236 | 447.41 | 77.50 | 400.25 |
| | 0.3333 | 0.1 | 0.0508 | 0.8170 | 213.44 | 39.33 | 189.36 |
| | 0.5 | 0.2 | 0.0519 | 0.8378 | 104.75 | 9.46 | 84.07 |
| | 1 | 0.5 | 0.0506 | 0.9997 | 100 | 0 | 21.07 |

Table 51: Linear regression with 10 independent predictors, the predictor of interest is continuous, other 9 predictors are binary, $N_{max} = 600$.

| n | $\beta_1$ | $R^2$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *Target SS* |
|---|---|---|---|---|---|---|---|
| 20 | 0.1429 | 0.02 | 0.0509 | 0.8720 | 527.24 | 123.41 | 384.34 |
| | 0.2294 | 0.05 | 0.0518 | 0.9040 | 328.27 | 168.64 | 149.16 |
| | 0.3333 | 0.1 | 0.0540 | 0.9002 | 173.47 | 119.59 | 70.62 |
| | 0.5 | 0.2 | 0.0604 | 0.8896 | 79.74 | 61.89 | 31.41 |
| | 1 | 0.5 | 0.0550 | 0.8943 | 25.32 | 13.85 | 7.86 |
| 50 | 0.1429 | 0.02 | 0.0501 | 0.8489 | 474.01 | 109.31 | 384.34 |
| | 0.2294 | 0.05 | 0.0511 | 0.8448 | 196.78 | 64.94 | 149.16 |
| | 0.3333 | 0.1 | 0.0537 | 0.8409 | 93.55 | 30.65 | 70.62 |
| | 0.5 | 0.2 | 0.0501 | 0.8790 | 52.59 | 6.94 | 31.41 |
| | 1 | 0.5 | 0.0503 | 0.9997 | 50 | 0 | 7.86 |
| 100 | 0.1429 | 0.02 | 0.0527 | 0.8265 | 434.67 | 85.94 | 384.34 |
| | 0.2294 | 0.05 | 0.0503 | 0.8171 | 170.31 | 36.30 | 149.16 |
| | 0.3333 | 0.1 | 0.0511 | 0.8799 | 101.58 | 5.73 | 70.62 |
| | 0.5 | 0.2 | 0.0495 | 0.9963 | 100 | 0 | 31.41 |
| | 1 | 0.5 | 0.0496 | 1 | 100 | 0 | 7.86 |

Table 52: Linear regression with 10 predictors, the predictor of interest is continuous, other 9 predictors are binary, pairwise Pearson correlation = 0.4 (original pairwise correlation = 0.5037), $N_{max} = 600$.

| n | $\beta_1$ | $R^2$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *Target SS* |
|---|---|---|---|---|---|---|---|
| 20 | 0.1429 | 0.02 | 0.0502 | 0.7753 | 572.42 | 78.53 | 587.73 |
| | 0.2294 | 0.05 | 0.0517 | 0.9037 | 430.10 | 165.92 | 227.98 |
| | 0.3333 | 0.1 | 0.0514 | 0.9049 | 262.79 | 161.82 | 108.06 |
| | 0.5 | 0.2 | 0.0574 | 0.8915 | 133.52 | 121.65 | 48.09 |
| | 1 | 0.5 | 0.0605 | 0.8758 | 51.28 | 103.57 | 11.99 |
| 50 | 0.1429 | 0.02 | 0.0516 | 0.7754 | 572.63 | 59.13 | 587.73 |
| | 0.2294 | 0.05 | 0.0504 | 0.8517 | 298.46 | 95.88 | 227.98 |
| | 0.3333 | 0.1 | 0.0516 | 0.8393 | 141.81 | 47.06 | 108.06 |
| | 0.5 | 0.2 | 0.0516 | 0.8411 | 65.72 | 18.33 | 48.09 |
| | 1 | 0.5 | 0.0517 | 0.9953 | 50.00 | 0.08 | 11.99 |
| 100 | 0.1429 | 0.02 | 0.0506 | 0.7816 | 573.44 | 48.94 | 587.73 |
| | 0.2294 | 0.05 | 0.0516 | 0.8191 | 259.30 | 55.86 | 227.98 |
| | 0.3333 | 0.1 | 0.0534 | 0.8194 | 125.15 | 23.75 | 108.06 |
| | 0.5 | 0.2 | 0.0515 | 0.9604 | 100.01 | 0.42 | 48.09 |
| | 1 | 0.5 | 0.0475 | 1 | 100 | 0 | 11.99 |

Table 53: Linear regression with 10 predictors, the predictor of interest is continuous, other 9 predictors are binary, pairwise Pearson correlation = 0.72 (original pairwise correlation = 0.9), $N_{max} = 600$.

| n | $\beta_1$ | $R^2$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *Target SS* |
|---|---|---|---|---|---|---|---|
| 20 | 0.1429 | 0.02 | 0.0492 | 0.5516 | 599.62 | 9.72 | 1030.10 |
| | 0.2294 | 0.05 | 0.0503 | 0.9150 | 596.37 | 31.70 | 399.13 |
| | 0.3333 | 0.1 | 0.0493 | 0.9932 | 589.05 | 61.72 | 189.62 |
| | 0.5 | 0.2 | 0.0494 | 0.9953 | 579.90 | 92.72 | 84.09 |
| | 1 | 0.5 | 0.0492 | 0.9919 | 573.76 | 117.32 | 21.05 |
| 50 | 0.1429 | 0.02 | 0.0493 | 0.5607 | 598.99 | 10.75 | 1030.10 |
| | 0.2294 | 0.05 | 0.0506 | 0.8451 | 492.09 | 114.31 | 399.13 |
| | 0.3333 | 0.1 | 0.0512 | 0.8689 | 300.58 | 154.38 | 189.62 |
| | 0.5 | 0.2 | 0.0531 | 0.8634 | 189.49 | 186.64 | 84.09 |
| | 1 | 0.5 | 0.0513 | 0.9637 | 139.77 | 203.07 | 21.05 |
| 100 | 0.1429 | 0.02 | 0.0495 | 0.5593 | 599.87 | 3.05 | 1030.10 |
| | 0.2294 | 0.05 | 0.0508 | 0.8231 | 440.37 | 90.82 | 399.13 |
| | 0.3333 | 0.1 | 0.0526 | 0.8173 | 211.45 | 49.89 | 189.62 |
| | 0.5 | 0.2 | 0.0506 | 0.8454 | 106.63 | 19.72 | 84.09 |
| | 1 | 0.5 | 0.0509 | 0.9998 | 100.49 | 15.64 | 21.05 |

Table 54: Logistic regression with a single continuous predictor, $N_{max} = 600$.

| n | $\beta_1$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *E($R^2$)* | *SD($R^2$)* | *Target SS* |
|---|---|---|---|---|---|---|---|---|
| 20 | 0.291 | 0.0488 | 0.8644 | 478.90 | 114.14 | 0.0224 | 0.0130 | 392.36 |
| | 0.469 | 0.0457 | 0.8660 | 214.96 | 101.14 | 0.0524 | 0.0285 | 166.54 |
| | 0.702 | 0.0391 | 0.8554 | 98.83 | 57.40 | 0.1001 | 0.0488 | 86.83 |
| | 1.127 | 0.0292 | 0.8439 | 39.21 | 26.90 | 0.1891 | 0.0791 | 47.90 |
| 50 | 0.291 | 0.0490 | 0.8308 | 425.29 | 87.56 | 0.0219 | 0.0132 | 392.36 |
| | 0.469 | 0.0446 | 0.8363 | 166.38 | 39.80 | 0.0521 | 0.0288 | 166.54 |
| | 0.702 | 0.0384 | 0.8161 | 74.41 | 17.41 | 0.0990 | 0.0514 | 86.83 |
| | 1.127 | 0.0421 | 0.9176 | 50.07 | 0.95 | 0.1946 | 0.0782 | 47.90 |
| 100 | 0.291 | 0.0467 | 0.8133 | 398.50 | 60.38 | 0.0221 | 0.0135 | 392.36 |
| | 0.469 | 0.0448 | 0.8021 | 153.97 | 23.63 | 0.0516 | 0.0300 | 166.54 |
| | 0.702 | 0.0473 | 0.9061 | 100.05 | 0.89 | 0.1039 | 0.0500 | 86.83 |
| | 1.127 | 0.0444 | 0.9984 | 100 | 0 | 0.2057 | 0.0656 | 47.90 |

Table 55: Logistic regression with a single binary predictor, $N_{max} = 600$.

| n | $\beta_1$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *E($R^2$)* | *SD($R^2$)* | *Target SS* |
|---|---|---|---|---|---|---|---|---|
| 20 | 0.286 | 0.0499 | 0.8545 | 459.82 | 63.38 | 0.0218 | 0.0131 | 391.74 |
| | 0.459 | 0.0469 | 0.8548 | 188.42 | 59.70 | 0.0528 | 0.0300 | 157.01 |
| | 0.667 | 0.0465 | 0.8566 | 93.59 | 59.72 | 0.1024 | 0.0545 | 78.72 |
| | 1.003 | 0.0370 | 0.8365 | 45.18 | 60.91 | 0.1900 | 0.0881 | 39.74 |
| 50 | 0.286 | 0.0497 | 0.8244 | 411.33 | 26.50 | 0.0223 | 0.0138 | 391.74 |
| | 0.459 | 0.0477 | 0.8267 | 160.18 | 11.36 | 0.0541 | 0.0316 | 157.01 |
| | 0.667 | 0.0446 | 0.8186 | 76.11 | 5.42 | 0.1032 | 0.0570 | 78.72 |
| | 1.003 | 0.0471 | 0.9222 | 50.02 | 2.50 | 0.2033 | 0.0877 | 39.74 |
| 100 | 0.286 | 0.0486 | 0.8104 | 396.69 | 10.95 | 0.0221 | 0.0139 | 391.74 |
| | 0.459 | 0.0495 | 0.8094 | 154.33 | 4.29 | 0.0541 | 0.0326 | 157.01 |
| | 0.667 | 0.0520 | 0.9076 | 100.00 | 0.04 | 0.1066 | 0.0552 | 78.72 |
| | 1.003 | 0.0516 | 0.9974 | 100 | 0 | 0.2043 | 0.0707 | 39.74 |

Table 56: Logistic regression with two independent continuous predictors, $N_{max} = 600$.

| n | $\beta_1$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *E($R^2$)* | *SD($R^2$)* | *Target SS* |
|---|---|---|---|---|---|---|---|---|
| 20 | 0.291 | 0.0492 | 0.8807 | 508.73 | 106.29 | 0.0237 | 0.0127 | 396.85 |
| | 0.469 | 0.0459 | 0.9013 | 248.82 | 119.86 | 0.0558 | 0.0264 | 166.12 |
| | 0.702 | 0.0422 | 0.8859 | 116.55 | 76.50 | 0.1067 | 0.0464 | 86.69 |
| | 1.127 | 0.0319 | 0.8756 | 46.62 | 40.16 | 0.2016 | 0.0727 | 47.49 |
| 50 | 0.291 | 0.0464 | 0.8443 | 442.11 | 89.96 | 0.0241 | 0.0133 | 396.85 |
| | 0.469 | 0.0447 | 0.8386 | 173.53 | 42.43 | 0.0570 | 0.0297 | 166.12 |
| | 0.702 | 0.0388 | 0.8274 | 78.05 | 19.09 | 0.1084 | 0.0511 | 86.69 |
| | 1.127 | 0.0427 | 0.9237 | 50.13 | 1.52 | 0.2079 | 0.0747 | 47.49 |
| 100 | 0.291 | 0.0493 | 0.8245 | 407.12 | 61.93 | 0.0244 | 0.0136 | 396.85 |
| | 0.469 | 0.0435 | 0.8176 | 157.11 | 24.23 | 0.0563 | 0.0304 | 166.12 |
| | 0.702 | 0.0479 | 0.8974 | 100.08 | 1.09 | 0.1106 | 0.0503 | 86.69 |
| | 1.127 | 0.0471 | 0.9981 | 100 | 0 | 0.2153 | 0.0666 | 47.49 |

Table 57: Logistic regression with two continuous predictors, Pearson correlation between the predictors = 0.4, $N_{max} = 600$.

| n | $\beta_1$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *E($R^2$)* | *SD($R^2$)* | *Target SS* |
|---|---|---|---|---|---|---|---|---|
| 20 | 0.291 | 0.0480 | 0.8557 | 547.81 | 82.76 | 0.0237 | 0.0123 | 464.24 |
| | 0.469 | 0.0477 | 0.8981 | 289.54 | 126.37 | 0.0550 | 0.0248 | 194.38 |
| | 0.702 | 0.0428 | 0.8863 | 136.58 | 83.34 | 0.1061 | 0.0437 | 100.99 |
| | 1.127 | 0.0348 | 0.8803 | 55.32 | 45.66 | 0.2017 | 0.0695 | 54.01 |
| 50 | 0.291 | 0.0494 | 0.8341 | 508.49 | 81.13 | 0.0239 | 0.0125 | 464.24 |
| | 0.469 | 0.0464 | 0.8355 | 206.83 | 50.21 | 0.0559 | 0.0275 | 194.38 |
| | 0.702 | 0.0419 | 0.8307 | 92.76 | 22.70 | 0.1076 | 0.0482 | 100.99 |
| | 1.127 | 0.0408 | 0.8772 | 50.51 | 2.97 | 0.2061 | 0.0746 | 54.01 |
| 100 | 0.291 | 0.0466 | 0.8262 | 481.18 | 66.65 | 0.0240 | 0.0127 | 464.24 |
| | 0.469 | 0.0435 | 0.8208 | 186.77 | 28.82 | 0.0564 | 0.0274 | 194.38 |
| | 0.702 | 0.0454 | 0.8474 | 100.95 | 3.94 | 0.1087 | 0.0479 | 100.99 |
| | 1.127 | 0.0460 | 0.9945 | 100 | 0 | 0.2128 | 0.0646 | 54.01 |

Table 58: Logistic regression with two continuous predictors, Pearson correlation between the predictors = 0.8, $N_{max} = 600$.

| n | $\beta_1$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *E($R^2$)* | *SD($R^2$)* | *Target SS* |
|---|---|---|---|---|---|---|---|---|
| 20 | 0.291 | 0.0498 | 0.5715 | 599.88 | 3.46 | 0.0237 | 0.0117 | 1062.13 |
| | 0.469 | 0.0471 | 0.8734 | 525.05 | 97.51 | 0.0532 | 0.0185 | 433.57 |
| | 0.702 | 0.0487 | 0.8924 | 300.21 | 129.10 | 0.1030 | 0.0313 | 213.53 |
| | 1.127 | 0.0406 | 0.8922 | 124.32 | 78.21 | 0.2022 | 0.0536 | 103.81 |
| 50 | 0.291 | 0.0493 | 0.5533 | 600.00 | 0.3191 | 0.0235 | 0.0117 | 1062.13 |
| | 0.469 | 0.0475 | 0.8321 | 469.13 | 88.06 | 0.0529 | 0.0192 | 433.57 |
| | 0.702 | 0.0468 | 0.8449 | 215.08 | 52.58 | 0.1044 | 0.0349 | 213.53 |
| | 1.127 | 0.0397 | 0.8351 | 83.98 | 20.72 | 0.2040 | 0.0592 | 103.81 |
| 100 | 0.291 | 0.0483 | 0.5522 | 600 | 0 | 0.0234 | 0.0118 | 1062.13 |
| | 0.469 | 0.0485 | 0.8262 | 434.66 | 64.92 | 0.0532 | 0.0195 | 433.57 |
| | 0.702 | 0.0469 | 0.8203 | 194.63 | 30.11 | 0.1043 | 0.0352 | 213.53 |
| | 1.127 | 0.0470 | 0.8533 | 100.26 | 2.00 | 0.2059 | 0.0587 | 103.81 |

Table 59: Logistic regression with two independent binary predictors, $N_{max} = 600$.

| n | $\beta_1$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *E($R^2$)* | *SD($R^2$)* | *Target SS* |
|---|---|---|---|---|---|---|---|---|
| 20 | 0.286 | 0.0493 | 0.8905 | 507.22 | 67.53 | 0.0238 | 0.0127 | 391.75 |
| | 0.459 | 0.0489 | 0.8930 | 214.95 | 67.46 | 0.0571 | 0.0290 | 157.01 |
| | 0.667 | 0.0477 | 0.8888 | 106.20 | 60.97 | 0.1098 | 0.0515 | 78.71 |
| | 1.003 | 0.0418 | 0.8782 | 51.19 | 61.59 | 0.2060 | 0.0844 | 39.74 |
| 50 | 0.286 | 0.0497 | 0.8321 | 431.31 | 35.73 | 0.0242 | 0.0138 | 391.75 |
| | 0.459 | 0.0500 | 0.8355 | 168.08 | 15.76 | 0.0583 | 0.0315 | 157.01 |
| | 0.667 | 0.0454 | 0.8272 | 79.79 | 7.44 | 0.1127 | 0.0561 | 78.71 |
| | 1.003 | 0.0490 | 0.9161 | 50.04 | 0.85 | 0.2163 | 0.0862 | 39.74 |
| 100 | 0.286 | 0.0506 | 0.8157 | 405.52 | 14.85 | 0.0243 | 0.0139 | 391.75 |
| | 0.459 | 0.0483 | 0.8144 | 157.76 | 5.81 | 0.0601 | 0.0333 | 157.01 |
| | 0.667 | 0.0501 | 0.9092 | 100.00 | 0.03 | 0.1162 | 0.0557 | 78.71 |
| | 1.003 | 0.0526 | 0.9982 | 100 | 0 | 0.2122 | 0.0702 | 39.74 |

Table 60: Logistic regression with two binary predictors, Pearson correlation between the predictors = 0.4 (tetrachoric correlation = 0.5878), $N_{max} = 600$.

| n | $\beta_1$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *E($R^2$)* | *SD($R^2$)* | *Target SS* |
|---|---|---|---|---|---|---|---|---|
| 20 | 0.286 | 0.0479 | 0.8546 | 544.52 | 64.15 | 0.0236 | 0.0125 | 465.48 |
| | 0.459 | 0.0502 | 0.8858 | 251.90 | 83.06 | 0.0564 | 0.0269 | 186.83 |
| | 0.667 | 0.0461 | 0.8806 | 123.85 | 66.24 | 0.1101 | 0.0497 | 93.79 |
| | 1.003 | 0.0416 | 0.8629 | 59.81 | 66.09 | 0.2075 | 0.0813 | 47.28 |
| 50 | 0.286 | 0.0490 | 0.8379 | 513.44 | 61.26 | 0.0238 | 0.0125 | 465.48 |
| | 0.459 | 0.0480 | 0.8412 | 209.22 | 49.24 | 0.0574 | 0.0289 | 186.83 |
| | 0.667 | 0.0440 | 0.8320 | 99.36 | 23.17 | 0.1099 | 0.0521 | 93.79 |
| | 1.003 | 0.0413 | 0.8649 | 51.78 | 6.87 | 0.2105 | 0.0837 | 47.28 |
| 100 | 0.286 | 0.0488 | 0.8268 | 489.12 | 48.26 | 0.0240 | 0.0128 | 465.48 |
| | 0.459 | 0.0470 | 0.8160 | 191.14 | 21.86 | 0.0592 | 0.0302 | 186.83 |
| | 0.667 | 0.0465 | 0.8478 | 101.45 | 5.36 | 0.1122 | 0.0522 | 93.79 |
| | 1.003 | 0.0500 | 0.9923 | 100.00 | 0.03 | 0.2125 | 0.0703 | 47.28 |

Table 61: Logistic regression with two binary predictors, Pearson correlation between the predictors = 0.8 (tetrachoric correlation = 0.9511), $N_{max} = 600$.

| n | $\beta_1$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *E($R^2$)* | *SD($R^2$)* | *Target SS* |
|---|---|---|---|---|---|---|---|---|
| 20 | 0.286 | 0.0484 | 0.5017 | 533.25 | 81.80 | 0.0239 | 0.0126 | 1073.74 |
| | 0.459 | 0.0438 | 0.6379 | 317.01 | 154.72 | 0.0568 | 0.0261 | 437.37 |
| | 0.667 | 0.0311 | 0.6210 | 192.22 | 167.95 | 0.1099 | 0.0478 | 218.31 |
| | 1.003 | 0.0129 | 0.5397 | 130.27 | 185.98 | 0.2115 | 0.0768 | 111.50 |
| 50 | 0.286 | 0.0465 | 0.5285 | 572.79 | 66.80 | 0.0235 | 0.0120 | 1073.74 |
| | 0.459 | 0.0480 | 0.7701 | 431.22 | 155.40 | 0.0555 | 0.0234 | 437.37 |
| | 0.667 | 0.0341 | 0.7414 | 226.80 | 106.71 | 0.1106 | 0.0435 | 218.31 |
| | 1.003 | 0.0201 | 0.6117 | 104.75 | 55.37 | 0.2188 | 0.0729 | 111.50 |
| 100 | 0.286 | 0.0494 | 0.5407 | 597.69 | 21.47 | 0.0231 | 0.0117 | 1073.74 |
| | 0.459 | 0.0470 | 0.8160 | 466.91 | 108.22 | 0.0541 | 0.0208 | 437.37 |
| | 0.667 | 0.0370 | 0.8070 | 247.41 | 102.48 | 0.1071 | 0.0389 | 218.31 |
| | 1.003 | 0.0281 | 0.7997 | 120.64 | 39.65 | 0.2087 | 0.0625 | 111.50 |

Table 62: Logistic regression with two independent predictors (binary and continuous), the predictor of interest is binary, $N_{max} = 600$.

| n | $\beta_1$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *E($R^2$)* | *SD($R^2$)* | *Target SS* |
|---|---|---|---|---|---|---|---|---|
| 20 | 0.286 | 0.0494 | 0.8866 | 509.00 | 68.09 | 0.0235 | 0.0126 | 391.74 |
| | 0.459 | 0.0489 | 0.8957 | 219.25 | 78.61 | 0.0569 | 0.0286 | 157.01 |
| | 0.667 | 0.0474 | 0.8908 | 108.95 | 67.96 | 0.1091 | 0.0508 | 78.71 |
| | 1.003 | 0.0419 | 0.8812 | 52.84 | 65.30 | 0.2037 | 0.0807 | 39.74 |
| 50 | 0.286 | 0.0509 | 0.8383 | 430.88 | 35.14 | 0.0240 | 0.0136 | 391.74 |
| | 0.459 | 0.0490 | 0.8350 | 167.59 | 14.77 | 0.0582 | 0.0313 | 157.01 |
| | 0.667 | 0.0457 | 0.8331 | 79.66 | 6.97 | 0.1130 | 0.0558 | 78.71 |
| | 1.003 | 0.0502 | 0.9190 | 50.03 | 0.59 | 0.2137 | 0.0838 | 39.74 |
| 100 | 0.286 | 0.0490 | 0.8203 | 405.28 | 14.65 | 0.0245 | 0.0139 | 391.74 |
| | 0.459 | 0.0489 | 0.8131 | 157.69 | 5.69 | 0.0596 | 0.0327 | 157.01 |
| | 0.667 | 0.0505 | 0.9093 | 100.00 | 0.01 | 0.1162 | 0.0557 | 78.71 |
| | 1.003 | 0.0502 | 0.9981 | 100 | 0 | 0.2116 | 0.0698 | 39.74 |

Table 63: Logistic regression with two predictors (binary and continuous), the predictor of interest is binary, Pearson correlation between the predictors = 0.4 (original correlation = 0.5037), $N_{max} = 600$.

| n | $\beta_1$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *E($R^2$)* | *SD($R^2$)* | *Target SS* |
|---|---|---|---|---|---|---|---|---|
| 20 | 0.286 | 0.0489 | 0.8648 | 554.50 | 58.52 | 0.0235 | 0.0121 | 467.00 |
| | 0.459 | 0.0476 | 0.8910 | 264.90 | 98.79 | 0.0561 | 0.0264 | 187.54 |
| | 0.667 | 0.0462 | 0.8834 | 132.19 | 78.02 | 0.1077 | 0.0473 | 93.60 |
| | 1.003 | 0.0409 | 0.8780 | 63.41 | 67.92 | 0.2028 | 0.0775 | 47.51 |
| 50 | 0.286 | 0.0494 | 0.8365 | 511.07 | 56.89 | 0.0235 | 0.0125 | 467.00 |
| | 0.459 | 0.0470 | 0.8301 | 202.87 | 31.56 | 0.0571 | 0.0291 | 187.54 |
| | 0.667 | 0.0447 | 0.8318 | 96.11 | 14.78 | 0.1114 | 0.0525 | 93.60 |
| | 1.003 | 0.0438 | 0.8675 | 50.69 | 3.84 | 0.2097 | 0.0925 | 47.51 |
| 100 | 0.286 | 0.0491 | 0.8221 | 485.62 | 42.02 | 0.0238 | 0.0127 | 467.00 |
| | 0.459 | 0.0480 | 0.8114 | 189.05 | 16.93 | 0.0580 | 0.0297 | 187.54 |
| | 0.667 | 0.0479 | 0.8436 | 100.63 | 2.66 | 0.1125 | 0.0534 | 93.60 |
| | 1.003 | 0.0491 | 0.9914 | 100 | 0 | 0.2126 | 0.0702 | 47.51 |

Table 64: Logistic regression with two predictors (binary and continuous), the predictor of interest is binary, Pearson correlation between the predictors = 0.8 (original correlation = 0.9991), $N_{max} = 600$.

| n | $\beta_1$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *E($R^2$)* | *SD($R^2$)* | *Target SS* |
|---|---|---|---|---|---|---|---|---|
| 20 | 0.286 | 0.0507 | 0.5537 | 600.00 | 0.65 | 0.0231 | 0.0116 | 1073.74 |
| | 0.459 | 0.0509 | 0.8695 | 531.41 | 80.69 | 0.0535 | 0.0185 | 428.96 |
| | 0.667 | 0.0480 | 0.8919 | 298.50 | 108.22 | 0.1036 | 0.0321 | 215.89 |
| | 1.003 | 0.0454 | 0.8892 | 141.19 | 84.00 | 0.2023 | 0.0537 | 108.67 |
| 50 | 0.286 | 0.0514 | 0.5563 | 600 | 0 | 0.0233 | 0.0117 | 1073.74 |
| | 0.459 | 0.0488 | 0.8401 | 469.82 | 67.86 | 0.0534 | 0.0193 | 428.96 |
| | 0.667 | 0.0488 | 0.8317 | 224.60 | 37.77 | 0.1053 | 0.0354 | 215.89 |
| | 1.003 | 0.0436 | 0.8322 | 99.65 | 16.88 | 0.2037 | 0.0607 | 108.67 |
| 100 | 0.286 | 0.0486 | 0.5549 | 600 | 0 | 0.0234 | 0.0117 | 1073.74 |
| | 0.459 | 0.0489 | 0.8161 | 438.70 | 45.18 | 0.0533 | 0.0198 | 428.96 |
| | 0.667 | 0.0464 | 0.8219 | 207.97 | 21.56 | 0.1055 | 0.0369 | 215.89 |
| | 1.003 | 0.0487 | 0.8150 | 101.32 | 3.87 | 0.2051 | 0.0622 | 108.67 |

Table 65: Logistic regression with independent predictors (binary and continuous), the predictor of interest is continuous, $N_{max} = 600$.

| n | $\beta_1$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *E($R^2$)* | *SD($R^2$)* | *Target SS* |
|---|---|---|---|---|---|---|---|---|
| 20 | 0.291 | 0.0485 | 0.8846 | 509.75 | 105.03 | 0.0240 | 0.0130 | 392.44 |
| | 0.469 | 0.0483 | 0.8912 | 247.08 | 118.34 | 0.0554 | 0.0269 | 165.90 |
| | 0.702 | 0.0421 | 0.8888 | 115.87 | 75.07 | 0.1076 | 0.0467 | 86.91 |
| | 1.127 | 0.0318 | 0.8806 | 46.27 | 39.26 | 0.2019 | 0.0728 | 47.37 |
| 50 | 0.291 | 0.0480 | 0.8472 | 442.24 | 89.75 | 0.0243 | 0.0134 | 392.44 |
| | 0.469 | 0.0455 | 0.8400 | 173.85 | 42.78 | 0.0568 | 0.0297 | 165.90 |
| | 0.702 | 0.0402 | 0.8297 | 78.03 | 19.05 | 0.1073 | 0.0506 | 86.91 |
| | 1.127 | 0.0420 | 0.9181 | 50.13 | 1.59 | 0.2070 | 0.0752 | 47.37 |
| 100 | 0.291 | 0.0469 | 0.8187 | 407.13 | 61.71 | 0.0242 | 0.0135 | 392.44 |
| | 0.469 | 0.0440 | 0.8111 | 157.13 | 24.43 | 0.0570 | 0.0296 | 165.90 |
| | 0.702 | 0.0474 | 0.8992 | 100.08 | 1.11 | 0.1119 | 0.0506 | 86.91 |
| | 1.127 | 0.0482 | 0.9985 | 100 | 0 | 0.2135 | 0.0671 | 47.37 |

Table 66: Logistic regression with two predictors (continuous and binary), the predictor of interest is continuous, Pearson correlation between the predictors = 0.4 (original correlation = 0.5037), $N_{max} = 600$.

| n | $\beta_1$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *E($R^2$)* | *SD($R^2$)* | *Target SS* |
|---|---|---|---|---|---|---|---|---|
| 20 | 0.291 | 0.0478 | 0.8605 | 547.89 | 82.85 | 0.0237 | 0.0123 | 467.64 |
| | 0.469 | 0.0468 | 0.8987 | 290.81 | 127.43 | 0.0551 | 0.0253 | 194.36 |
| | 0.702 | 0.0437 | 0.8894 | 137.09 | 83.32 | 0.1059 | 0.0433 | 101.54 |
| | 1.127 | 0.0334 | 0.8747 | 55.38 | 45.61 | 0.2025 | 0.0710 | 54.04 |
| 50 | 0.291 | 0.0504 | 0.8316 | 508.91 | 81.19 | 0.0238 | 0.0126 | 467.64 |
| | 0.469 | 0.0460 | 0.8419 | 207.00 | 50.76 | 0.0559 | 0.0269 | 194.36 |
| | 0.702 | 0.0416 | 0.8331 | 92.87 | 22.91 | 0.1074 | 0.0474 | 101.54 |
| | 1.127 | 0.0410 | 0.8774 | 50.50 | 2.85 | 0.2056 | 0.0743 | 54.04 |
| 100 | 0.291 | 0.0480 | 0.8181 | 481.96 | 67.03 | 0.0238 | 0.0126 | 467.64 |
| | 0.469 | 0.0460 | 0.8248 | 187.54 | 29.11 | 0.0566 | 0.0279 | 194.36 |
| | 0.702 | 0.0449 | 0.8478 | 100.96 | 3.98 | 0.1091 | 0.0480 | 101.54 |
| | 1.127 | 0.0459 | 0.9932 | 100 | 0 | 0.2122 | 0.0659 | 54.04 |

Table 67: Logistic regression with two predictors (continuous and binary), the predictor of interest is continuous, Pearson correlation between the predictors = 0.8 (original correlation = 0.9991), $N_{max} = 600$.

| n | $\beta_1$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *E($R^2$)* | *SD($R^2$)* | *Target SS* |
|---|---|---|---|---|---|---|---|---|
| 20 | 0.291 | 0.0507 | 0.5539 | 599.52 | 7.19 | 0.0235 | 0.0117 | 1084.87 |
| | 0.469 | 0.0500 | 0.8590 | 523.37 | 104.62 | 0.0531 | 0.0183 | 451.97 |
| | 0.702 | 0.0459 | 0.8881 | 311.91 | 140.72 | 0.1028 | 0.0306 | 237.06 |
| | 1.127 | 0.0412 | 0.8751 | 132.03 | 89.97 | 0.2028 | 0.0488 | 129.16 |
| 50 | 0.291 | 0.0494 | 0.5523 | 599.95 | 1.65 | 0.0236 | 0.0117 | 1084.87 |
| | 0.469 | 0.0504 | 0.8285 | 469.61 | 97.07 | 0.0524 | 0.0187 | 451.97 |
| | 0.702 | 0.0441 | 0.8338 | 219.58 | 63.28 | 0.1037 | 0.0328 | 237.06 |
| | 1.127 | 0.0396 | 0.8262 | 85.55 | 24.40 | 0.2018 | 0.0532 | 129.16 |
| 100 | 0.291 | 0.0496 | 0.5517 | 600.00 | 0.34 | 0.0235 | 0.0117 | 1084.87 |
| | 0.469 | 0.0476 | 0.8180 | 434.85 | 74.89 | 0.0530 | 0.0191 | 451.97 |
| | 0.702 | 0.0440 | 0.8150 | 195.57 | 36.02 | 0.1038 | 0.0339 | 237.06 |
| | 1.127 | 0.0448 | 0.8217 | 100.52 | 3.06 | 0.2016 | 0.0543 | 129.16 |

Table 68: Logistic regression with 10 independent continuous predictors, $N_{max} = 600$.

| n | $\beta_1$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *E($R^2$)* | *SD($R^2$)* | *Target SS* |
|---|---|---|---|---|---|---|---|---|
| 20 | 0.291 | 0.0522 | 0.9408 | 597.34 | 19.76 | 0.0368 | 0.0133 | 394.94 |
| | 0.469 | 0.0515 | 0.9943 | 532.33 | 115.92 | 0.0679 | 0.0209 | 166.74 |
| | 0.702 | 0.0506 | 0.9921 | 394.69 | 183.77 | 0.1239 | 0.0339 | 87.00 |
| | 1.127 | 0.0510 | 0.9892 | 234.16 | 193.66 | 0.2363 | 0.0557 | 47.75 |
| 50 | 0.291 | 0.0499 | 0.9211 | 564.08 | 60.60 | 0.0374 | 0.0137 | 394.94 |
| | 0.469 | 0.0498 | 0.9403 | 271.46 | 89.95 | 0.0803 | 0.0269 | 166.74 |
| | 0.702 | 0.0482 | 0.9337 | 123.13 | 47.71 | 0.1492 | 0.0434 | 87.00 |
| | 1.127 | 0.0480 | 0.9378 | 55.49 | 16.02 | 0.2630 | 0.0574 | 47.75 |
| 100 | 0.291 | 0.0493 | 0.8740 | 482.99 | 71.47 | 0.0395 | 0.0146 | 394.94 |
| | 0.469 | 0.0506 | 0.8733 | 188.30 | 32.49 | 0.0909 | 0.0299 | 166.74 |
| | 0.702 | 0.0559 | 0.9011 | 101.39 | 5.15 | 0.1685 | 0.0472 | 87.00 |
| | 1.127 | 0.0563 | 0.9968 | 100 | 0 | 0.2747 | 0.0638 | 47.75 |

# References


[1] Coffey, C.S. and Muller, K.E. (1999). Exact test size and power of a Gaussian error linear model for an internal pilot study. *Statistics in Medicine* **18**, 1199–214.

[2] Friede, T. and Kieser, M. (2006). Sample Size Recalculation in Internal Pilot Study Designs: A Review. *Biometrical Journal*, **48**, 537–555.

[3] Friede, T. and Kieser, M. (2011). Blinded sample size recalculation for clinical trials with normal data and baseline adjusted analysis. *Pharmaceutical statistics*, **10**, 8–13.

[4] Glueck, D.H. and Muller, K.E. (2003) Adjusting power for a baseline covariate in linear models. *Statistics in Medicine*, **22**, 2535–51.

[5] Gurka, J.M., Coffey, K.S., and Gurka K.K. (2010) Internal pilots for observational studies. *Biometrical journal*, **52**, 590–603.

[6] Haneuse S., Schildcrout J., Gillen D. (2012) A two-stage strategy to accomodate general patterns of counfouding in the design of observational studies. *Biostatistics*, **13**, 274–288.

[7] L.F. Leach, R.K. Henson, The use and impact of adjusted R2 effects in published regression research. *Multiple Linear Regression Viewpoints*, 33(1), 1–11, 2007

[8] N. Nagelkerke, A Note on a General Definition of the Coefficient of Determination, *Biometrika*, 78(3), 691-–92, 1991

[9] F.Y. Hsieh , D.A. Bloch, M.D. Larsen, A simple method of sample size calculation for linear and logistic regression,  Statist. Med. 17, 1623–1634 (1998)

[10] Proschan, M.A. (2005) Two-stage sample size re-estimation based on a nuisance parameter: a review. *Journal of Biopharmaceutical Statistics*, **15**, 559–74.

Table 69: Logistic regression with 10 continuous predictors, Pearson correlation among the predictors = 0.4, $N_{max} = 600$.

| n | $\beta_1$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *E($R^2$)* | *SD($R^2$)* | *Target SS* |
|---|---|---|---|---|---|---|---|---|
| 20 | 0.291 | 0.0492 | 0.8176 | 599.78 | 5.04 | 0.0368 | 0.0131 | 593.26 |
| | 0.469 | 0.0509 | 0.9870 | 573.48 | 70.97 | 0.0661 | 0.0190 | 243.22 |
| | 0.702 | 0.0516 | 0.9901 | 466.23 | 158.65 | 0.1198 | 0.0291 | 124.41 |
| | 1.127 | 0.0517 | 0.9868 | 292.68 | 196.01 | 0.2305 | 0.0481 | 64.02 |
| 50 | 0.291 | 0.0512 | 0.8095 | 598.96 | 8.80 | 0.0367 | 0.0133 | 593.26 |
| | 0.469 | 0.0485 | 0.9448 | 401.22 | 107.13 | 0.0714 | 0.0222 | 243.22 |
| | 0.702 | 0.0504 | 0.9376 | 185.60 | 66.53 | 0.1356 | 0.0373 | 124.41 |
| | 1.127 | 0.0475 | 0.9331 | 73.49 | 28.45 | 0.2497 | 0.0537 | 64.02 |
| 100 | 0.291 | 0.0508 | 0.8074 | 595.50 | 16.66 | 0.0368 | 0.0132 | 593.26 |
| | 0.469 | 0.0504 | 0.8793 | 286.08 | 49.66 | 0.0780 | 0.0250 | 243.22 |
| | 0.702 | 0.0490 | 0.8717 | 128.50 | 21.30 | 0.1497 | 0.0412 | 124.41 |
| | 1.127 | 0.0562 | 0.9745 | 100.00 | 0.15 | 0.2694 | 0.0590 | 64.02 |

Table 70: Logistic regression with 10 continuous predictors, Pearson correlation among the predictors = 0.8, $N_{max} = 600$.

| n | $\beta_1$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *E($R^2$)* | *SD($R^2$)* | *Target SS* |
|---|---|---|---|---|---|---|---|---|
| 20 | 0.291 | 0.0532 | 0.3741 | 600 | 0 | 0.0366 | 0.0133 | 1716.16 |
| | 0.469 | 0.0509 | 0.7452 | 599.89 | 4.02 | 0.0654 | 0.0182 | 688.07 |
| | 0.702 | 0.0512 | 0.9640 | 589.21 | 43.72 | 0.1150 | 0.0237 | 337.57 |
| | 1.127 | 0.0520 | 0.9892 | 487.98 | 147.31 | 0.2170 | 0.0327 | 156.66 |
| 50 | 0.291 | 0.0524 | 0.3880 | 600 | 0 | 0.0365 | 0.0132 | 1716.16 |
| | 0.469 | 0.0526 | 0.7318 | 599.82 | 3.44 | 0.0654 | 0.0183 | 688.07 |
| | 0.702 | 0.0488 | 0.9383 | 497.14 | 95.00 | 0.1163 | 0.0244 | 337.57 |
| | 1.127 | 0.0491 | 0.9479 | 213.27 | 74.93 | 0.2231 | 0.0379 | 156.66 |
| 100 | 0.291 | 0.0525 | 0.3851 | 600 | 0 | 0.0367 | 0.0133 | 1716.16 |
| | 0.469 | 0.0494 | 0.7506 | 599.45 | 5.37 | 0.0655 | 0.0179 | 688.07 |
| | 0.702 | 0.0493 | 0.8773 | 378.02 | 64.81 | 0.1192 | 0.0265 | 337.57 |
| | 1.127 | 0.0501 | 0.8719 | 147.23 | 25.34 | 0.2334 | 0.0444 | 156.66 |

Table 71: Logistic regression with 10 independent binary predictors, $N_{max} = 600$.

| n | $\beta_1$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *E($R^2$)* | *SD($R^2$)* | *Target SS* |
|---|---|---|---|---|---|---|---|---|
| 20 | 0.286 | 0.0511 | 0.9369 | 599.62 | 5.18 | 0.0363 | 0.0132 | 391.79 |
| | 0.459 | 0.0506 | 0.9950 | 515.07 | 112.61 | 0.0694 | 0.0213 | 157.02 |
| | 0.667 | 0.0554 | 0.9940 | 333.89 | 157.14 | 0.1310 | 0.0381 | 78.72 |
| | 1.003 | 0.0563 | 0.9917 | 174.35 | 130.49 | 0.2553 | 0.0686 | 39.74 |
| 50 | 0.286 | 0.0539 | 0.9263 | 581.64 | 32.53 | 0.0366 | 0.0134 | 391.79 |
| | 0.459 | 0.0509 | 0.9469 | 266.10 | 66.11 | 0.0821 | 0.0276 | 157.02 |
| | 0.667 | 0.0535 | 0.9419 | 126.84 | 36.04 | 0.1532 | 0.0447 | 78.72 |
| | 1.003 | 0.0511 | 0.9282 | 58.25 | 16.43 | 0.2669 | 0.0616 | 39.74 |
| 100 | 0.286 | 0.0507 | 0.8814 | 487.17 | 38.04 | 0.0395 | 0.0145 | 391.79 |
| | 0.459 | 0.0536 | 0.8700 | 189.67 | 15.54 | 0.0940 | 0.0313 | 157.02 |
| | 0.667 | 0.0573 | 0.8922 | 100.58 | 2.65 | 0.1757 | 0.0508 | 78.72 |
| | 1.003 | 0.0596 | 0.9958 | 100 | 0 | 0.2771 | 0.0701 | 39.74 |

Table 72: Logistic regression with 10 binary predictors, Pairwise Pearson correlations = 0.4 (pairwise tetrachoric correlations = 0.5878), $N_{max} = 600$.

| n | $\beta_1$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *E($R^2$)* | *SD($R^2$)* | *Target SS* |
|---|---|---|---|---|---|---|---|---|
| 20 | 0.286 | 0.0530 | 0.8097 | 599.85 | 3.28 | 0.0364 | 0.0131 | 593.40 |
| | 0.459 | 0.0540 | 0.9829 | 546.10 | 90.59 | 0.0681 | 0.0205 | 237.59 |
| | 0.667 | 0.0536 | 0.9835 | 366.61 | 150.23 | 0.1302 | 0.0370 | 120.56 |
| | 1.003 | 0.0574 | 0.9689 | 196.80 | 140.95 | 0.2570 | 0.0698 | 60.46 |
| 50 | 0.286 | 0.0514 | 0.7985 | 599.83 | 2.81 | 0.0361 | 0.0132 | 593.40 |
| | 0.459 | 0.0513 | 0.9595 | 453.58 | 113.56 | 0.0702 | 0.0213 | 237.59 |
| | 0.667 | 0.0510 | 0.9576 | 242.70 | 107.94 | 0.1315 | 0.0369 | 120.56 |
| | 1.003 | 0.0472 | 0.9563 | 110.43 | 60.86 | 0.2454 | 0.0561 | 60.46 |
| 100 | 0.286 | 0.0503 | 0.8017 | 598.89 | 7.11 | 0.0362 | 0.0133 | 593.40 |
| | 0.459 | 0.0503 | 0.8920 | 314.40 | 61.30 | 0.0776 | 0.0250 | 237.59 |
| | 0.667 | 0.0497 | 0.8902 | 149.46 | 30.09 | 0.1473 | 0.0425 | 120.56 |
| | 1.003 | 0.0588 | 0.9634 | 100.31 | 3.27 | 0.2645 | 0.0621 | 60.46 |

Table 73: Logistic regression with 10 binary predictors, Pairwise Pearson correlations = 0.8 (pairwise tetrachoric correlations = 0.9511), $N_{max} = 600$.

| n | $\beta_1$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *E($R^2$)* | *SD($R^2$)* | *Target SS* |
|---|---|---|---|---|---|---|---|---|
| 20 | 0.286 | 0.0525 | 0.3753 | 599.98 | 1.84 | 0.0366 | 0.0134 | 1762.09 |
| | 0.459 | 0.0514 | 0.7320 | 598.22 | 18.84 | 0.0663 | 0.0186 | 712.72 |
| | 0.667 | 0.0510 | 0.9520 | 593.21 | 47.99 | 0.1156 | 0.0246 | 354.25 |
| | 1.003 | 0.0523 | 0.9884 | 589.54 | 69.44 | 0.2151 | 0.0334 | 177.01 |
| 50 | 0.286 | 0.0528 | 0.3734 | 599.92 | 3.82 | 0.0370 | 0.0135 | 1762.09 |
| | 0.459 | 0.0532 | 0.7393 | 598.74 | 15.46 | 0.0663 | 0.0187 | 712.72 |
| | 0.667 | 0.0523 | 0.9357 | 556.26 | 83.04 | 0.1180 | 0.0281 | 354.25 |
| | 1.003 | 0.0543 | 0.9000 | 378.26 | 161.75 | 0.2323 | 0.0510 | 177.01 |
| 100 | 0.286 | 0.0532 | 0.3695 | 600 | 0 | 0.0365 | 0.0133 | 1762.09 |
| | 0.459 | 0.0535 | 0.7388 | 599.54 | 6.45 | 0.0661 | 0.0185 | 712.72 |
| | 0.667 | 0.0522 | 0.9367 | 556.67 | 80.80 | 0.1158 | 0.0247 | 354.25 |
| | 1.003 | 0.0468 | 0.9658 | 375.98 | 153.40 | 0.2199 | 0.0374 | 177.01 |

Table 74: Logistic regression with 10 independent predictors, the predictor of interest is binary, other 9 predictors are continuous, $N_{max} = 600$.

| n | $\beta_1$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *E($R^2$)* | *SD($R^2$)* | *Target SS* |
|---|---|---|---|---|---|---|---|---|
| 20 | 0.286 | 0.0520 | 0.9350 | 599.69 | 4.51 | 0.0365 | 0.0132 | 391.77 |
| | 0.459 | 0.0518 | 0.9963 | 534.27 | 105.22 | 0.0680 | 0.0202 | 157.01 |
| | 0.667 | 0.0523 | 0.9964 | 393.54 | 175.21 | 0.1254 | 0.0350 | 78.71 |
| | 1.003 | 0.0536 | 0.9941 | 250.82 | 191.27 | 0.2378 | 0.0572 | 39.74 |
| 50 | 0.286 | 0.0510 | 0.9274 | 581.53 | 32.61 | 0.0367 | 0.0133 | 391.77 |
| | 0.459 | 0.0535 | 0.9486 | 265.32 | 65.28 | 0.0821 | 0.0274 | 157.01 |
| | 0.667 | 0.0533 | 0.9361 | 126.53 | 35.85 | 0.1515 | 0.0446 | 78.71 |
| | 1.003 | 0.0501 | 0.9291 | 58.00 | 16.12 | 0.2687 | 0.0613 | 39.74 |
| 100 | 0.286 | 0.0505 | 0.8785 | 486.64 | 37.52 | 0.0396 | 0.0144 | 391.77 |
| | 0.459 | 0.0540 | 0.8682 | 189.63 | 15.45 | 0.0939 | 0.0318 | 157.01 |
| | 0.667 | 0.0572 | 0.8911 | 100.53 | 2.49 | 0.1758 | 0.0506 | 78.71 |
| | 1.003 | 0.0590 | 0.9970 | 100 | 0 | 0.2783 | 0.0701 | 39.74 |

Table 75: Logistic regression with 10 predictors, the predictor of interest is binary, other 9 predictors are continuous, pairwise Pearson correlation = 0.4 (original pairwise correlation = 0.5037), $N_{max} = 600$.

| n | $\beta_1$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *E($R^2$)* | *SD($R^2$)* | *Target SS* |
|---|---|---|---|---|---|---|---|---|
| 20 | 0.286 | 0.0492 | 0.7925 | 599.97 | 1.47 | 0.0363 | 0.0136 | 597.42 |
| | 0.459 | 0.0501 | 0.9901 | 575.61 | 63.98 | 0.0670 | 0.0194 | 240.41 |
| | 0.667 | 0.0507 | 0.9927 | 471.43 | 151.70 | 0.1207 | 0.0301 | 120.93 |
| | 1.003 | 0.0505 | 0.9917 | 319.03 | 193.86 | 0.2294 | 0.0496 | 60.78 |
| 50 | 0.286 | 0.0518 | 0.8005 | 599.89 | 2.22 | 0.0363 | 0.0133 | 597.42 |
| | 0.459 | 0.0508 | 0.9396 | 403.79 | 96.17 | 0.0721 | 0.0221 | 240.41 |
| | 0.667 | 0.0511 | 0.9383 | 196.77 | 61.94 | 0.1363 | 0.0391 | 120.93 |
| | 1.003 | 0.0489 | 0.9378 | 87.88 | 30.87 | 0.2494 | 0.0564 | 60.78 |
| 100 | 0.286 | 0.0521 | 0.8081 | 599.15 | 5.70 | 0.0364 | 0.0132 | 597.42 |
| | 0.459 | 0.0504 | 0.8741 | 292.75 | 40.37 | 0.0794 | 0.0258 | 240.41 |
| | 0.667 | 0.0493 | 0.8667 | 138.82 | 18.99 | 0.1520 | 0.0437 | 120.93 |
| | 1.003 | 0.0582 | 0.9620 | 100.01 | 0.27 | 0.2713 | 0.0643 | 60.78 |

Table 76: Logistic regression with 10 predictors, the predictor of interest is binary, other 9 predictors are continuous, pairwise Pearson correlation = 0.72 (original pairwise correlation = 0.9), $N_{max} = 600$.

| n | $\beta_1$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *E($R^2$)* | *SD($R^2$)* | *Target SS* |
|---|---|---|---|---|---|---|---|---|
| 20 | 0.286 | 0.0511 | 0.5629 | 600 | 0 | 0.0364 | 0.0132 | 1049.87 |
| | 0.459 | 0.0512 | 0.9167 | 598.97 | 10.98 | 0.0656 | 0.0184 | 420.83 |
| | 0.667 | 0.0516 | 0.9914 | 563.34 | 80.93 | 0.1163 | 0.0253 | 211.10 |
| | 1.003 | 0.0525 | 0.9900 | 431.60 | 168.27 | 0.2190 | 0.0374 | 106.01 |
| 50 | 0.286 | 0.0516 | 0.5584 | 600 | 0 | 0.0364 | 0.0133 | 1049.87 |
| | 0.459 | 0.0519 | 0.9085 | 581.73 | 38.50 | 0.0662 | 0.0187 | 420.83 |
| | 0.667 | 0.0513 | 0.9484 | 343.14 | 89.02 | 0.1218 | 0.0295 | 211.10 |
| | 1.003 | 0.0508 | 0.9451 | 154.28 | 48.83 | 0.2307 | 0.0465 | 106.01 |
| 100 | 0.286 | 0.0529 | 0.5651 | 600 | 0 | 0.0363 | 0.0132 | 1049.87 |
| | 0.459 | 0.0527 | 0.8753 | 510.27 | 56.14 | 0.0672 | 0.0194 | 420.83 |
| | 0.667 | 0.0498 | 0.8719 | 244.29 | 31.44 | 0.1301 | 0.0338 | 211.10 |
| | 1.003 | 0.0519 | 0.8684 | 110.24 | 11.85 | 0.2467 | 0.0528 | 106.01 |

Table 77: Logistic regression with 10 independent predictors, the predictor of interest is continuous, other 9 predictors are binary, $N_{max} = 600$.

| n | $\beta_1$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *E($R^2$)* | *SD($R^2$)* | *Target SS* |
|---|---|---|---|---|---|---|---|---|
| 20 | 0.291 | 0.0520 | 0.9401 | 597.79 | 17.65 | 0.0369 | 0.0134 | 378.70 |
| | 0.469 | 0.0510 | 0.9954 | 536.70 | 112.53 | 0.0677 | 0.0204 | 150.73 |
| | 0.702 | 0.0500 | 0.9943 | 404.40 | 184.56 | 0.1239 | 0.0331 | 71.90 |
| | 1.127 | 0.0506 | 0.9906 | 249.71 | 210.36 | 0.2350 | 0.0536 | 33.44 |
| 50 | 0.291 | 0.0512 | 0.9219 | 564.69 | 60.05 | 0.0373 | 0.0136 | 378.70 |
| | 0.469 | 0.0497 | 0.9411 | 272.22 | 90.32 | 0.0800 | 0.0268 | 150.73 |
| | 0.702 | 0.0483 | 0.9316 | 122.82 | 46.71 | 0.1485 | 0.0438 | 71.90 |
| | 1.127 | 0.0498 | 0.9365 | 55.63 | 16.49 | 0.2628 | 0.0577 | 33.44 |
| 100 | 0.291 | 0.0508 | 0.8797 | 482.27 | 71.47 | 0.0396 | 0.0146 | 378.70 |
| | 0.469 | 0.0488 | 0.8720 | 188.39 | 32.54 | 0.0912 | 0.0299 | 150.73 |
| | 0.702 | 0.0541 | 0.8966 | 101.43 | 5.29 | 0.1674 | 0.0467 | 71.90 |
| | 1.127 | 0.0572 | 0.9978 | 100.00 | 0.01 | 0.2737 | 0.0628 | 33.44 |

Table 78: Logistic regression with 10 predictors, the predictor of interest is continuous, other 9 predictors are binary, pairwise Pearson correlation = 0.4 (original pairwise correlation = 0.5037), $N_{max} = 600$.

| n | $\beta_1$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *E($R^2$)* | *SD($R^2$)* | *Target SS* |
|---|---|---|---|---|---|---|---|---|
| 20 | 0.291 | 0.0516 | 0.8112 | 599.88 | 3.61 | 0.0368 | 0.0132 | 510.82 |
| | 0.469 | 0.0529 | 0.9920 | 580.43 | 61.30 | 0.0657 | 0.0186 | 203.06 |
| | 0.702 | 0.0522 | 0.9940 | 492.54 | 149.35 | 0.1187 | 0.0283 | 97.02 |
| | 1.127 | 0.0508 | 0.9920 | 336.80 | 205.43 | 0.2279 | 0.0456 | 44.98 |
| 50 | 0.291 | 0.0510 | 0.8070 | 598.76 | 9.84 | 0.0366 | 0.0132 | 510.82 |
| | 0.469 | 0.0508 | 0.9415 | 398.41 | 108.03 | 0.0713 | 0.0219 | 203.06 |
| | 0.702 | 0.0489 | 0.9391 | 185.10 | 67.78 | 0.1350 | 0.0369 | 97.02 |
| | 1.127 | 0.0466 | 0.9324 | 73.15 | 28.42 | 0.2500 | 0.0544 | 44.98 |
| 100 | 0.291 | 0.0496 | 0.7993 | 595.18 | 17.50 | 0.0368 | 0.0133 | 510.82 |
| | 0.469 | 0.0502 | 0.8702 | 285.96 | 49.59 | 0.0783 | 0.0248 | 203.06 |
| | 0.702 | 0.0489 | 0.8682 | 128.52 | 21.51 | 0.1499 | 0.0411 | 97.02 |
| | 1.127 | 0.0568 | 0.9715 | 100.00 | 0.41 | 0.2697 | 0.0589 | 44.98 |

Table 79: Logistic regression with 10 predictors, the predictor of interest is continuous, other 9 predictors are binary, pairwise Pearson correlation = 0.72 (original pairwise correlation = 0.9), $N_{max} = 600$.

| n | $\beta_1$ | *Type I error* | *Power* | *E(N)* | *SD(N)* | *E($R^2$)* | *SD($R^2$)* | *Target SS* |
|---|---|---|---|---|---|---|---|---|
| 20 | 0.291 | 0.0497 | 0.5681 | 600.00 | 0.01 | 0.0370 | 0.0135 | 825.22 |
| | 0.469 | 0.0483 | 0.9130 | 599.56 | 8.98 | 0.0657 | 0.0185 | 326.66 |
| | 0.702 | 0.0510 | 0.9977 | 595.56 | 34.58 | 0.1150 | 0.0235 | 155.84 |
| | 1.127 | 0.0502 | 0.9988 | 585.79 | 74.43 | 0.2146 | 0.0291 | 72.23 |
| 50 | 0.291 | 0.0497 | 0.5637 | 599.99 | 0.78 | 0.0369 | 0.0133 | 825.22 |
| | 0.469 | 0.0508 | 0.8888 | 545.41 | 78.83 | 0.0669 | 0.0189 | 326.66 |
| | 0.702 | 0.0480 | 0.9155 | 330.44 | 145.61 | 0.1240 | 0.0301 | 155.84 |
| | 1.127 | 0.0437 | 0.9094 | 189.60 | 186.09 | 0.2376 | 0.0475 | 72.23 |
| 100 | 0.291 | 0.0486 | 0.5615 | 600 | 0 | 0.0368 | 0.0133 | 825.22 |
| | 0.469 | 0.0477 | 0.8555 | 475.59 | 77.85 | 0.0683 | 0.0194 | 326.66 |
| | 0.702 | 0.0455 | 0.8515 | 215.75 | 43.41 | 0.1309 | 0.0326 | 155.84 |
| | 1.127 | 0.0485 | 0.8361 | 102.35 | 17.92 | 0.2515 | 0.0501 | 72.23 |

[11] G. Shieh, On Power and Sample Size Calculations for Likelihood Ratio Tests in Generalized Linear Models, Biometrics 56, 1192–1196 (2000)