The Data Generating Process (DGP):

Customer C Spend on Order $X = t * c * r * d + \epsilon$

$t =$ Whether or not the customer recieved the experimental treatment $t \in [1, 1.01]$.

$c =$ Customer's mean order amount in dollars where $c \sim \mathcal{N}(\$100, \$25)$.

$r =$ Scalar corresponding to mean retailer order amount where $r \sim \mathcal{N}(1.0, 0.05)$.

$d =$ Scalar corresponding to mean order amount for that day of the week where $d \in [0.7, 1.3]$.

$\epsilon =$ Random noise where $\epsilon \sim \mathcal{N}(\$25, \$1)$.

Let the number of orders N placed by Customer $C \ \sim exp(\lambda = 1)$.

The Model:

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4$$

...where X_1 is the treatment, X_2 is the retailer's mean order amount prior to treatment, and X_3 is the mean order amount for that day of the week prior to treatment.

	Method I	**Method II**	**Method III**	**Method IV**	**Method V**	**Method VI**	**Method VII**	**Method VIII**	**Method IX**
Points Used	3	5	7	3	5	7	3	5	7
Simulations Per Point	100	100	100	200	200	200	300	300	300
Verification Simulations	500	500	500	500	500	500	500	500	500
Total Simulations Required	800	1,000	1,200	1,100	1,500	1,900	1,400	2,000	2,600
Mean Run Time (Min.)	10.3	10.2	14.9	14.5	17.8	25.0	18.7	23.1	35.0
Mean Error: Est. Power - Desired Power	0.012	0.048	0.046	0.043	0.033	0.041	0.025	0.026	0.042
Std. of Error Est. Power - Desired Power	0.094	0.057	0.043	0.07	0.052	0.037	0.084	0.034	0.028

Simulations were conducted using a single thread (Intel Xeon CPU E5-2670).

If we were implementing a difference of means t-test, then we could simply apply the formula below to get the appropriate sample size.

$$n = 2\left(\frac{Z_{1-(\alpha/2)} + Z_{1-\beta}}{ES}\right)^2 \text{ where } ES = \frac{|\mu_1 - \mu_2|}{\sigma}$$

Here, α is the probability of a Type I error (false rejection of the NULL hypothesus, i.e. a false positive) and β is the probability of a Type II error (i.e. a false negative). Observe how if we want to make our test more rigorous and reduce the false positive rate, the the left side of the numerator increase and the resulting recommended sample size - n, also increases. Similarly, if we want to reduce the probability for failing to detect the difference across groups, then we will need to increase the right-hand side of the numerator - also increasing the recommended sample size.

$$y = \beta_0 + \epsilon$$

$$y = \beta_0 + X_1\beta_1 + \epsilon$$

$$y = \beta_0 + X_1\beta_1 + X_2\beta_2 + \epsilon$$

$$y = \beta_0 + X_1\beta_1 + X_2\beta_2 + X_3\beta_3 + \epsilon$$

$$y = \beta_0 + ... + \beta_n X_n + \epsilon \quad \text{where} \quad \hat{\epsilon} = y - \hat{y}$$

In our ongoing effort to predict the amount spent on an individual order, let's try the following...

$$y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \epsilon$$

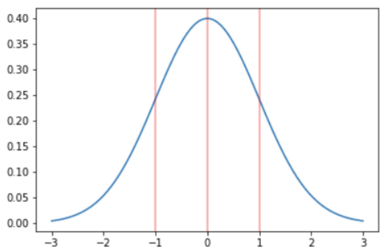
where X_1 is whether the customer was assigned the treatment...

X_2 is the customer's mean order amount prior to treatment...

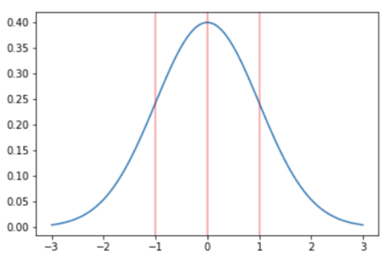
X_3 is the retailer's mean order amount prior to treatment...

X_4 is the mean order amount for that day of the week prior to treatment...

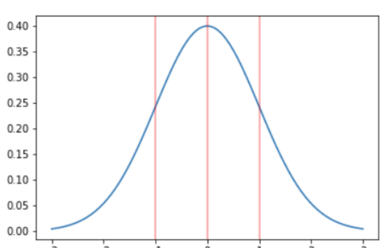
ϵ is unaccounted for variation in y, and $\hat{\epsilon} = y - \hat{y}$



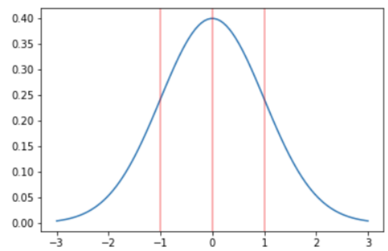
(a) μ +/- [1]σ
with 500 iterations.



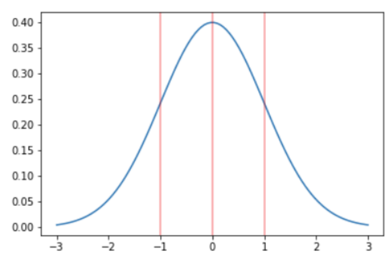
(b) μ +/- [1,2]σ
with 500 iterations.



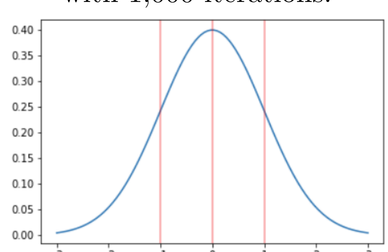
(c) μ +/- [1,2,3]σ
with 500 iterations.



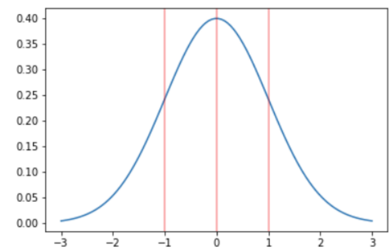
(d) μ +/- [1]σ
with 1,000 iterations.



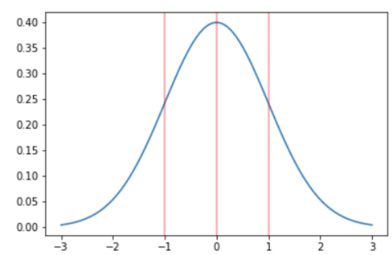
(e) μ +/- [1,2]σ
with 1,000 iterations.



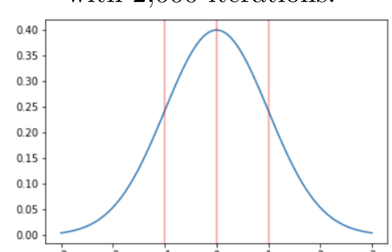
(f) μ +/- [1,2,3]σ
with 1,000 iterations.



(g) μ +/- [1]σ
with 2,000 iterations.



(h) μ +/- [1,2]σ
with 2,000 iterations.



(i) μ +/- [1,2,3]σ
with 2,000 iterations.

	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV	XV	XVI
Diff. in Estimated Power (Mean)	-0.723	-0.683	-.541	-0.387	0.198	0.199	0.199	0.143	0.199	0.199	0.199	0.105	0.189	0.192	0.185	-0.124
Diff. in Estimated Power (Std.)	0.011	0.014	0.018	-0.022	0.002	0.0009	0.001	0.01	0.0009	0.0007	0.0009	0.01	0.004	0.003	0.005	0.023
Mean Estimated Sample Size	31	123	689	3,836	3,836	14,017	83,711	460,231	3,847	14,055	83,954	461,663	4,093	14,904	88,932	491,528
Std. of Estimated Sample Sizes	31	123	689	3,836	3,836	14,017	83,711	460,231	3,847	14,055	83,954	461,663	4,093	14,904	88,932	491,528
Mean Adjusted R^2	0.99	0.96	0.94	0.93	0.49	0.41	0.38	0.38	0.36	0.36	0.36	0.36	0.004	0.001	0.0	0.0
Verification Simulations	500	500	500	500	500	500	500	500	500	500	500	500	500	500	500	500