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 customer's mean order amount prior to treatment,

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

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

Required Sample Size Estimates Taken Via Interpolation

	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
Total Simulations Required	300	500	700	600	1,000	1,400	900	1,500	2,100
Mean Difference in Estimated Power (Std.)	0.04 (0.104)	0.034 (0.034)	-0.008 (0.047)	0.039 (0.089)	0.025 (0.024)	-0.022 (0.041)	0.046 (0.061)	0.025 (0.022)	-0.021 (0.033)
Mean Estimated Sample Size (Std.)	25,147 (5,208)	23,497 (1,836)	21,101 (2,434)	24,752 (4,698)	22,883 (1,237)	20,244 (2,036)	24,925 (3,690)	22,956 (1,054)	20,231 (1,491)
Mean Estimation Time in Minutes (Std.)	0.761 (0.625)	0.474 (0.1)	0.659 (0.081)	1.447 (1.116)	1.239 (0.28)	1.464 (0.295)	1.879 (1.409)	1.411 (0.196)	1.976 (0.198)
Verification Simulations	500	500	500	500	500	500	500	500	500

Simulations were conducted on dual Intel Xeon CPU E5-2670 processors.

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...

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

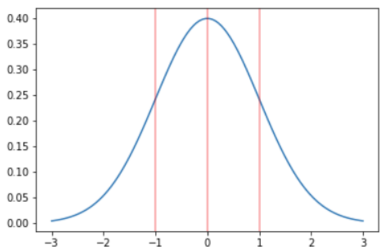
where X_1 is whether the customer was assigned the treatment...

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

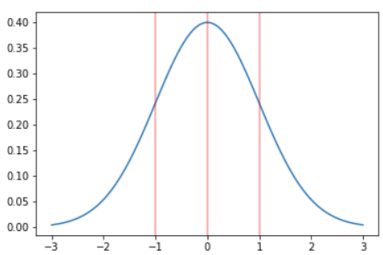
X_3 is the retailer's mean order amount prior to treatment...

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

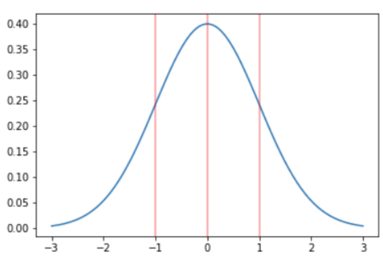
ϵ is unaccounted for variation in y, and $\hat{\epsilon} = y - \hat{y}$



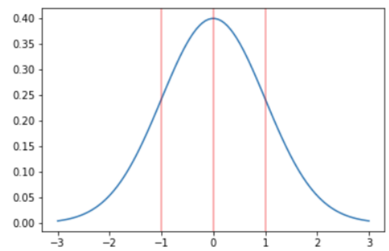
(a) μ +/- [1]σ with 500 iterations.



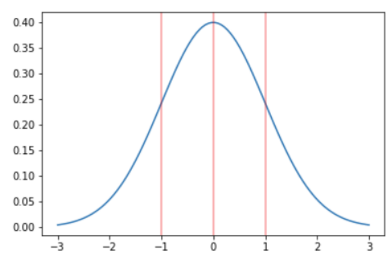
(b) μ +/- [1,2]σ with 500 iterations.



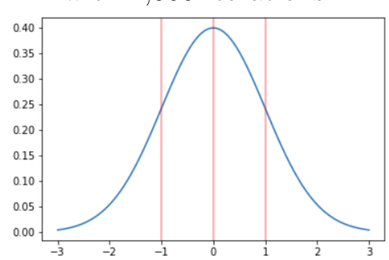
(c) μ +/- [1,2,3]σ with 500 iterations.



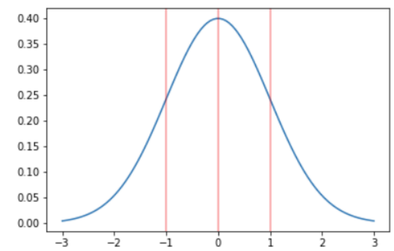
(d) μ +/- [1]σ with 1,000 iterations.



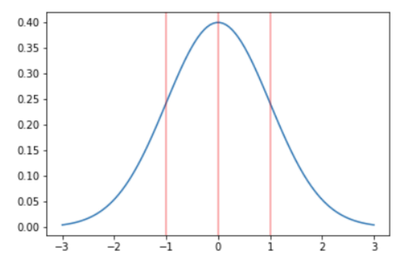
(e) μ +/- [1,2]σ with 1,000 iterations.



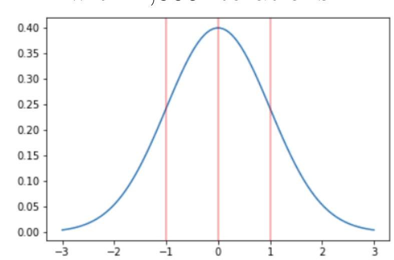
(f) μ +/- [1,2,3]σ with 1,000 iterations.



(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.

Required Sample Size Estimates Taken Via Analysis of Residual Variance

	I	**II**	**III**	**IV**	**V**	**VI**	**VII**	**VIII**	**IX**	**X**	**XI**	**XII**	**XIII**	**XIV**	**XV**	**XVI**
Mean Diff. in Estimated Power (Std.)	-0.723 (0.011)	-0.683 (0.014)	-0.541 (0.018)	-0.387 (-0.022)	0.198 (0.002)	0.199 (<0.01)	0.199 (<0.01)	0.143 (0.01)	0.199 (<0.01)	0.199 (<0.01)	0.199 (<0.01)	0.105 (0.01)	0.189 (<0.01)	0.192 (<0.01)	0.185 (<0.01)	-0.124 (0.023)
Mean Estimated Sample Size (Std.)	31 (0.43)	123 (1.2)	689 (5.7)	3,836 (36.2)	3,836 (19.2)	14,017 (68.0)	83,711 (435.7)	460,231 (2,183.2)	3,847 (23.6)	14,055 (77.07)	83,954 (512.8)	461,663 (2370.2)	4,093 (25.6)	14,904 (84.3)	88,932 (504.0)	491,528 (2,927.6)
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

Simulations were conducted on dual Intel Xeon CPU E5-2670 processors.

Let $Pr(p < 0.05) = 1 - e^{-\lambda x}$ where x is the number of observations sampled and the alternative hypothesis is true.