This document expands on the test power, minimum sample size, and minimum detectable effect calculation under the stacked incrementality test framework.

1 Test Power

Recall the Welch's t-statistic in use:

$$t = \frac{\overline{D} - 0}{\sqrt{\sum_{g \in G} \frac{s_g^2}{n_g}}} \tag{1}$$

We begin by observing H_0 is rejected if $t > t_{\nu,1-\alpha}$, the $1-\alpha$ quantile for t-distribution with ν degrees of freedom. Under a *specific* alternate hypothesis $\mu_D = \theta$, the test power $1 - \beta_\theta$ is specified as follow:

$$1 - \beta_\theta = Pr\left(\frac{\overline{D} - 0}{\sqrt{\sum_{g \in G} \frac{s_g^2}{n_g}}} > t_{\nu,1-\alpha} \,\middle|\, \mu_D = \theta \right). \tag{2}$$

Introducing a pair of self-cancelling θ into the inequality of the probability expression, and doing some basic algebraic manipulation, we have from Equation (2):

$$\begin{aligned} 1 - \beta_\theta &= Pr\left(\frac{\overline{D} - \theta + \theta}{\sqrt{\sum_{g \in G} \frac{s_g^2}{n_g}}} > t_{\nu,1-\alpha} \,\middle|\, \mu_D = \theta \right) && (3) \\ &= Pr\left(\frac{\overline{D} - \theta}{\sqrt{\sum_{g \in G} \frac{s_g^2}{n_g}}} + \frac{\theta}{\sqrt{\sum_{g \in G} \frac{s_g^2}{n_g}}} > t_{\nu,1-\alpha} \,\middle|\, \mu_D = \theta \right) && (4) \\ &= Pr\left(\frac{\overline{D} - \theta}{\sqrt{\sum_{g \in G} \frac{s_g^2}{n_g}}} > t_{\nu,1-\alpha} - \frac{\theta}{\sqrt{\sum_{g \in G} \frac{s_g^2}{n_g}}} \,\middle|\, \mu_D = \theta \right) && (5) \end{aligned}$$

Notice by basic probability theory, Expression (5) is equivalent to the following:

$$1 - \beta_\theta = 1 - Pr\left(\frac{\overline{D} - \theta}{\sqrt{\sum_{g \in G} \frac{s_g^2}{n_g}}} < t_{\nu,1-\alpha} - \frac{\theta}{\sqrt{\sum_{g \in G} \frac{s_g^2}{n_g}}} \,\middle|\, \mu_D = \theta \right), \tag{6}$$

which the LHS of the second term approximately following the Student's t distribution with ν degrees of freedom:

$$1 - \beta_\theta = 1 - T_\nu \left(t_{\nu,1-\alpha} - \frac{\theta}{\sqrt{\sum_{g \in G} \frac{s_g^2}{n_g}}} \right), \tag{7}$$

where T_ν denotes the cumulative density function (CDF) of the Student's t-distribution with ν degrees of freedom.

2 Minimum Sample Size Required

To achieve a (pre-specified) minimum test power $\pi_{\min} \in [0, 1]$, we require, with some algebraic manipulations:

$$1 - \beta_\theta > \pi_{\min} \tag{8}$$

$$\iff 1 - T_\nu \left(t_{\nu,1-\alpha} - \frac{\theta}{\sqrt{\sum_{g \in G} \frac{s_g^2}{n_g}}} \right) > \pi_{\min} \tag{9}$$

$$\iff 1 - \pi_{\min} > T_\nu \left(t_{\nu,1-\alpha} - \frac{\theta}{\sqrt{\sum_{g \in G} \frac{s_g^2}{n_g}}} \right). \tag{10}$$

Applying the inverse of the CDF of the t-distribution, which is simply the quantile function, to both sides of Inequality (10)[1], we obtain:

$$t_{\nu,1-\pi_{\min}} > t_{\nu,1-\alpha} - \frac{\theta}{\sqrt{\sum_{g \in G} \frac{s_g^2}{n_g}}} \tag{11}$$

$$\iff \frac{\theta}{\sqrt{\sum_{g \in G} \frac{s_g^2}{n_g}}} > t_{\nu,1-\alpha} - t_{\nu,1-\pi_{\min}}. \tag{12}$$

Assuming $\alpha < \pi_{\min}$, which implies $t_{\nu,1-\alpha} > t_{\nu,1-\pi_{\min}}$, we have from inequality (12):

$$\frac{\theta}{t_{\nu,1-\alpha} - t_{\nu,1-\pi_{\min}}} > \sqrt{\sum_{g \in G} \frac{s_g^2}{n_g}} \tag{13}$$

$$\implies \left(\frac{\theta}{t_{\nu,1-\alpha} - t_{\nu,1-\pi_{\min}}} \right)^2 > \sum_{g \in G} \frac{s_g^2}{n_g} \tag{14}$$

While it is possible to find the optimal n_g for each group g based on the measured / expected sample variances s_g^2. In the test design stage, we can make some simplifying assumptions for common scenarios.

2.1 Equal Sample Size In All Groups

Our first scenario assume that the sample size in all groups are equal, i.e. $n_g = n_{\min} \, \forall g \in G$. Inequality (14) then becomes, under basic algebraic ma-

[1]As both CDF and quantile functions are non-decreasing functions, we do not need to worry about flipping the signs within the inequalities.

nipulations:

$$\left(\frac{\theta}{t_{\nu,1-\alpha}-t_{\nu,1-\pi_{\min}}}\right)^2 > \sum_{g\in G}\frac{s_g^2}{n_{\min}} \tag{15}$$

$$\iff \left(\frac{\theta}{t_{\nu,1-\alpha}-t_{\nu,1-\pi_{\min}}}\right)^2 > \frac{1}{n_{\min}}\sum_{g\in G} s_g^2 \tag{16}$$

$$\iff n_{\min} > \left(\frac{t_{\nu,1-\alpha}-t_{\nu,1-\pi_{\min}}}{\theta}\right)^2 \sum_{g\in G} s_g^2. \tag{17}$$

As ν depends on the value of $n_{\min}$ in the Welch-Satterthwaite Equation, it provides little benefit to derive Inequality (17) in full analytical form. Instead we can find a value quickly in an iterative manner, using the quantile value of the normal distribution as a first estimation:

Algorithm 1 An algorithm that finds a solution of $n_{\min}$ to Inequality (17) quickly. The algorithm iteratively updates n_min and ν until a solution is found.

$n_{\min} \leftarrow \left\lceil \mathrm{RHS}\left(z_{1-\alpha}, z_{1-\pi_{\min}}, \theta, [s_g^2]_{g\in G}\right)\right\rceil$
$\nu \leftarrow \mathrm{WSE}\left([s_g^2]_{g\in G}, [n_{\min}, n_{\min}, ..., n_{\min}]\right)$
while $n_{\min} \leq \mathrm{RHS}\left(t_{\nu,1-\alpha}, t_{\nu,1-\pi_{\min}}, \theta, [s_g^2]_{g\in G}\right)$ **do**
  $n_{\min} \leftarrow \left\lceil \mathrm{RHS}\left(t_{\nu,1-\alpha}, t_{\nu,1-\pi_{\min}}, \theta, [s_g^2]_{g\in G}\right)\right\rceil$
  $\nu \leftarrow \mathrm{WSE}\left([s_g^2]_{g\in G}, [n_{\min}, n_{\min}, ..., n_{\min}]\right)$
end while
return $n_{\min}$

where z_q is the q quantile of the standard normal distribution, $\lceil\cdot\rceil$ is the ceiling function, and the functions $\mathrm{RHS}(\cdot)$ (representing the RHS expression in Inequality (17)) and $\mathrm{WSE}(\cdot)$ (the generalised Welch-Satterthwaite Equation) are defined as follow:

$$\mathrm{RHS}\left(t, t', \theta, [s_g^2]_{g\in G}\right) = \left(\frac{t-t'}{\theta}\right)^2 \sum_{g\in G} s_g^2, \tag{18}$$

$$\mathrm{WSE}\left([s_g^2]_{g\in G}, [n_g]_{g\in G}\right) = \frac{\left(\sum_{g\in G}\frac{s_g^2}{n_g}\right)^2}{\sum_{g\in G}\frac{s_g^4}{(n_g-1)n_g^2}}. \tag{19}$$

Note we intentionally formulate Inequality (17) and Algorithm 1 in such a way that sample size calculation can be generalised to as many groups with different estimated / measured sample variance as one like[2].

[2]The inequality do assumes that the overall test statistic is a linear combination of each group's test statistic with weighting $\{+1, -1\}$. To apply a non unity weighting k_g to group g, simply replace the s_g^2 term in Inequality (17) with $k_g^2 s_g^2$.

2.2 Fixed Sample Size Ratio Between Groups

The second scenario assumes there is a fixed ratio in sample size between various groups. This is particularly useful if one has an abundance of test individuals, and would like to limit the proportion of individuals who are in the control / exposed group(s) due to various business considerations (e.g. withholding certain promotional marketing emails to individuals in control group(s) are sometimes necessary to measure the value of those emails, though it is considered harmful from a customer engagement perspective).

In the simplest form we can partition all the groups G into two sets of groups G_1 and G_2, and specify the sample size ratio between the group(s) within sets G_1 and G_2 as $k_1 : k_2$ respectively. Inequality (14) then becomes, again with basic algebraic manipulations:

$$\left(\frac{\theta}{t_{\nu,1-\alpha} - t_{\nu,1-\pi_{\min}}}\right)^2 > \sum_{g \in G_1} \frac{s_g^2}{n_{\min}} + \sum_{g \in G_2} \frac{s_g^2}{n_g} \tag{20}$$

$$\iff \left(\frac{\theta}{t_{\nu,1-\alpha} - t_{\nu,1-\pi_{\min}}}\right)^2 > \sum_{g \in G_1} \frac{s_g^2}{n_{\min}} + \sum_{g \in G_2} \frac{s_g^2}{\frac{k_2}{k_1} n_{\min}} \tag{21}$$

$$\iff \left(\frac{\theta}{t_{\nu,1-\alpha} - t_{\nu,1-\pi_{\min}}}\right)^2 > \frac{1}{n_{\min}} \left(\sum_{g \in G_1} s_g^2 + \sum_{g \in G_2} \frac{k_1 s_g^2}{k_2}\right) \tag{22}$$

$$\iff n_{\min} > \left(\frac{t_{\nu,1-\alpha} - t_{\nu,1-\pi_{\min}}}{\theta}\right)^2 \left(\sum_{g \in G_1} s_g^2 + \sum_{g \in G_2} \frac{k_1 s_g^2}{k_2}\right), \tag{23}$$

which a solution can be found by slightly modifying the RHS$(\cdot)$ function in Algorithm 1. The above only covers the case where the groups are partitioned into two, though it can be easily extended to cover cases where there are multiple partitions, as long as the ratio between each partition is specified.

2.3 Fixed Sample Size for Some Groups

Our last scenario assumes some group's sample size is specified and hence fixed. This might be the case when one has a limited number of individuals in some groups, and can only resort in varying the number of test individuals in other groups to achieve the necessary statistical significance and power.

We formulate the scenario as follow. The groups G are partitioned into two: the groups which sample size is specified (denoted G_s, i.e. group $g \in G_s$ has a fixed sample size of n_g), and the groups which sample size is to be determined (denoted G_d). We further assume we only need to find one minimum sample size $n_{\min}$ for all the groups in G_d. Inequality (14) then becomes, again with basic algebraic manipulations:

$$\left(\frac{\theta}{t_{\nu,1-\alpha} - t_{\nu,1-\pi_{\min}}}\right)^2 > \sum_{g\in G_d} \frac{s_g^2}{n_{\min}} + \sum_{g\in G_s} \frac{s_g^2}{n_g} \tag{24}$$

$$\iff \left(\frac{\theta}{t_{\nu,1-\alpha} - t_{\nu,1-\pi_{\min}}}\right)^2 - \sum_{g\in G_s} \frac{s_g^2}{n_g} > \frac{1}{n_{\min}} \sum_{g\in G_d} s_g^2 \tag{25}$$

$$\iff n_{\min} > \frac{\sum_{g\in G_d} s_g^2}{\left(\frac{\theta}{t_{\nu,1-\alpha} - t_{\nu,1-\pi_{\min}}}\right)^2 - \sum_{g\in G_s} \frac{s_g^2}{n_g}}, \tag{26}$$

which a solution can be found by modifying modifying the RHS($\cdot$) function in Algorithm 1. This formulation is intended to support an arbitrary number (up to the total number of groups minus one) of groups that require a fixed sample size.

3 Minimum Detectable Effect

In a *post-hoc* analysis, an useful metric is the minimum effect the test will be able to detect (i.e. the effect specified by the alternate hypothesis θ under a test with power of at least $\pi_{\min}$. This can be done by rearranging inequality (12) and fixing all s_g^2 and n_g, leading to:

$$\theta > (t_{\nu,1-\alpha} - t_{\nu,1-\pi_{\min}}) \sqrt{\sum_{g\in G} \frac{s_g^2}{n_g}}. \tag{27}$$