Adobe Target
Calculations for A/B Testing

The following calculations are used to compute the A/B test statistics for conversion rates.

Each experience is compared to the "Control" experience. The conversion rate is computed either by visit or by visitor depending on which metric is selected. The inputs to the calculations are only the number of visits or visitors and the number of conversions.

The conversion rate for each experience is given by:

$$CR = \frac{C}{N}$$

where C is the number of conversions, and N is the number of visits or visitors.

The conversion rate is a random variable that depends on the particular sample. It would be slightly different from day to day just because of the finite number of visits/visitors. The standard error of the (mean) conversion rate is the "standard error of the mean" which decreases as the number of visits or visitors, N, increases. It is given by:

$$\sigma_{CR} = \sqrt{\frac{CR(1 - CR)}{N}}$$

This only applies if the (mean) conversion rate is distributed approximately normally. It is assumed that the mean is normally distributed so long as $n_c \geq K$ and $n_v - n_c \geq K$ where K ranges from 5 to 30. In Target K is set to 30.

The 95% confidence interval of the conversion rate is a range of possible conversion rates where there is a 95% chance that the true conversion rate is within this range. Adobe Target always reports a 95% confidence interval. Because the standard error of the mean is normally distributed the confidence interval is:

$$[CR - 1.96\,\sigma_{CR},\ CR + 1.96\,\sigma_{CR}].$$

The difference between the conversion rate of any experience a, compared to the control experience is

$$\Delta CR = (CR_a - CR_c)$$

It has a standard error of:

$$\sigma_{\Delta CR} = \sqrt{\sigma_a{}^2 + \sigma_c{}^2}$$

where σ_a is the standard error of the conversion rate of the test experience, a, and σ_c is the standard error of the conversion rate of the control experience, c.

The lift of experience a relative to the control is:

$$l = \frac{(CR_a - CR_c)}{CR_c} = \frac{CR_a}{CR_c} - 1$$

In the case where the errors in the conversion rates are small compared to the conversion rates for the control and test experience, the standard error of the lift is approximately:

$$\sigma_l \cong l \sqrt{\left(\frac{\sigma_a}{CR_a}\right)^2 + \left(\frac{\sigma_c}{CR_c}\right)^2}$$

This results in a confidence interval of the lift of:

$$[l - 1.96 \, \sigma_l, \; l + 1.96 \, \sigma_l].$$

T Tests

A t-test is performed between the test experience and the control experience to see if the data for the test and the control might have been caused just by statistical fluctuations. We start by assuming the "null hypothesis" which is that the test and control experiences are actually the same, and then find the "p-value" which represents the probability of observing the values we observed if the null hypothesis was correct. Note this has nothing to do with the lift. It just says whether or not the two experiences are different.

Calculating t

The t value is defined to be the difference of the means of any two independent random variables, x and y, divided by the standard error of this difference:

$$t \equiv \frac{\bar{x} - \bar{y}}{\sigma_{\bar{x} - \bar{y}}}$$

where $\bar{x}$ and $\bar{y}$ are the means of x and y, respectively, and $\sigma_{\bar{x} - \bar{y}}$ is the standard error of the difference between $\bar{x}$ and $\bar{y}$ given by:

$$\sigma_{\bar{x} - \bar{y}} = \sqrt{\sigma_{\bar{x}}^2 + \sigma_{\bar{y}}^2}$$

For conversion rates, $\sigma_{\bar{x}} = \sigma_a$ is the standard error of the test experience, a, and $\sigma_{\bar{y}} = \sigma_c$ is the standard error of the conversion rate of the control experience, c.

The standard error of the mean of any random variable x is given by:

$$\sigma_{\bar{x}}^2 = var(x)/N$$

where N is the number of values that were used to calculate $\bar{x}$, and the variance is given by considering the observed values $\{x_i\}$:

$$var(x) = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2 = \frac{1}{N-1}\left[\sum_{i=1}^{N}x_i^2 - N\bar{x}^2\right]$$

Note that in this formula for the variance we divide by $N-1$ because we don't know the true mean exactly so there is 1 free parameter to be fit (this is called 1 "degree of freedom").

For conversion rates each visit (or visitor), i, either results in a conversion or it does not. If it results in a conversion we let $x_i = 1$ and if it does not result in a conversion we let $x_i = 0$. It is easy to show that in the approximation of large N, where $N-1 \cong N$, the above expression for the variance reduces to

$$var(x) \cong \bar{x} * (1 - \bar{x})$$

or

$$var(CR) \cong CR(1 - CR)$$

which leads to:

$$\sigma_{CR} = \sqrt{\frac{CR(1 - CR)}{N}}$$

When we measure the (mean) RPV, x_i is the revenue for visit i, N is the number of visits, and $\bar{x}$ is the (mean) RPV. Note that in this case there are many visits that do not result in revenue and for these visits $x_i = 0$.

In the case where we are measuring the average order value (AOV), x_i is the revenue for visit i, and N is the number of *orders*. Note the only difference between AOV vs. RPV is that we only count visits that have orders for AOV.

Pooled Variance

When we do a t test, we assume that the two experiences are the same, and so we assume the two measured responses, x and y, come from the same distribution. In that case we can get a more accurate measure of the variance by "pooling" the samples.

Recall that for the standard error of the differences in the means we had:

$$\sigma_{\bar{x}-\bar{y}}^2 = \sigma_{\bar{x}}^2 + \sigma_{\bar{y}}^2 = \frac{var(x)}{N_x} + \frac{var(y)}{N_y}$$

The pooled variance is the variance that would be calculated if we combined the computations for the variance of x and y and is given by:

$$var_{pooled} = \frac{(N_x - 1) * var(x) + (N_y - 1) * var(y)}{N_x + N_y - 2}$$

This results in:

$$\sigma_{\bar{x}-\bar{y}}^2 = \frac{var(x)}{N_x} + \frac{var(y)}{N_y} = var_{pooled} * \left(\frac{1}{N_x} - \frac{1}{N_y}\right)$$

therefore

$$t = \frac{\bar{x} - \bar{y}}{\sqrt{var_{pooled} * \left(\frac{1}{N_x} - \frac{1}{N_y}\right)}}$$

Often in the literature it is assumed that N_x and N_y are much larger than 1 so

$$var_{pooled} \approx \frac{N_x * var(x) + N_y * var(y)}{N_x + N_y}$$

Calculating p and the Confidence

Once we have the value of t, we can calculate the probability that the experiences were really the same, but appeared to be different (p). The p-value comes from the integral of the t-distribution. Since we are only testing the hypothesis that the two experiences were the same we don't care whether or not t > 0 or t < 0 so we use a 2-sided t-test to compute p. This computation is done numerically by various math libraries, but it requires a value, v, which represents the number of degrees of freedom that we are fitting. For the pooled variance this is just:

$$v = N_x + N_y - 2$$

The p-value can be computed from the two-tailed t-test in excel:

$$p = t.dist.2t(|t|, v)$$

In Target we compute it using an apache commons statistics library. The confidence, c, we report is:

$$c = 1 - p$$