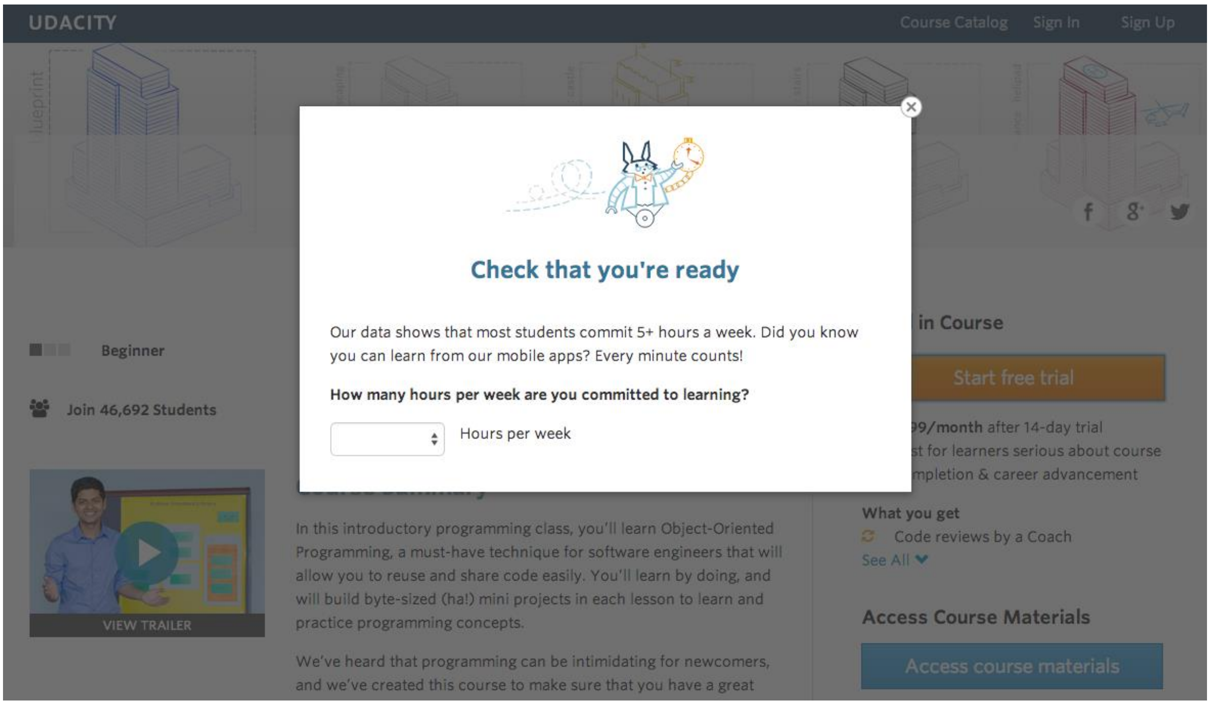
A/B testing of free trial screener

Experiment Overview: Free Trial Screener

At the time of this experiment, Udacity courses currently have two options on the course overview page: "start free trial", and "access course materials". If the student clicks "start free trial", they will be asked to enter their credit card information, and then they will be enrolled in a free trial for the paid version of the course. After 14 days, they will automatically be charged unless they cancel first. If the student clicks "access course materials", they will be able to view the videos and take the quizzes for free, but they will not receive coaching support or a verified certificate, and they will not submit their final project for feedback.
In the experiment, Udacity tested a change where if the student clicked "start free trial", they were asked how much time they had available to devote to the course. If the student indicated 5 or more hours per week, they would be taken through the checkout process as usual. If they indicated fewer than 5 hours per week, a message would appear indicating that Udacity courses usually require a greater time commitment for successful completion, and suggesting that the student might like to access the course materials for free. At this point, the student would have the option to continue enrolling in the free trial, or access the course materials for free instead. This screenshot shows what the experiment looks like.



The hypothesis was that this might set clearer expectations for students upfront, thus reducing the number of frustrated students who left the free trial because they didn't have enough time—without significantly reducing the number of students to continue past the free trial and eventually complete the course. If this hypothesis held true, Udacity could improve the overall student experience and improve coaches' capacity to support students who are likely to complete the course.

The unit of diversion is a cookie, although if the student enrolls in the free trial, they are tracked by user-id from that point forward. The same user-id cannot enroll in the free trial twice. For users that do not enroll, their user-id is not tracked in the experiment, even if they were signed in when they visited the course overview page.
The null hypothesis: the free trial screener will not impact on evaluation metrics.
The alternative hypothesis: the free trial screener will impact the evaluation metrics.

Metric Choice

Which of the following metrics would you choose to measure for this experiment and why? For each metric you choose, indicate whether you would use it as an invariant metric or an evaluation metric. The practical significance boundary for each metric, that is, the difference that would have to be observed before that was a meaningful change for the business, is given in parentheses. All practical significance boundaries are given as absolute changes. Any place where "unique cookies" are mentioned, the uniqueness is determined by day. (That is, the same cookie visiting on different days would be counted twice.) User-ids are automatically unique since the site does not allow the same user-id to enroll twice.

- **Number of cookies:** That is, number of unique cookies to view the course overview page d_{min}=3000.
- **Number of user-ids:** That is, number of users who enroll in the free trial d_{min}=50.
- **Number of clicks:** That is, number of unique cookies to click the "Start free trial" button (which happens before the free trial screener is triggered) d_{min}=240.
- **Click-through-probability:** That is, number of unique cookies to click the "Start free trial" button divided by number of unique cookies to view the course overview page (d_{min}=0.01).
- **Gross conversion:** That is, number of user-ids to complete checkout and enroll in the free trial divided by number of unique cookies to click the "Start free trial" button (d_{min}=0.01).
- **Retention:** That is, number of user-ids to remain enrolled past the 14-day boundary (and thus make at least one payment) divided by number of user-ids to complete checkout (d_{min}=0.01).
- **Net conversion:** That is, number of user-ids to remain enrolled past the 14-day boundary (and thus make at least one payment) divided by the number of unique cookies to click the "Start free trial" button (d_{min}=0.0075).

Choosing Invariant Metrics

In an A/B test, invariant metrics are used to check if the randomization process worked correctly, ensuring that there are no significant differences between the control and treatment groups before the experiment starts. These metrics should not be affected by the experiment itself. For the given A/B test, the best invariant metrics would be those that measure aspects of user behavior before users interact with the part of the page affected by the experiment (in this case, the screener that shows up after clicking "Start free trial").

Recommended Invariant Metrics:

Number of Cookies
Description: The number of unique cookies to view the course overview page.
Why It's a Good Invariant Metric: The experiment does not affect who visits the course overview page (before clicking anything), so the number of cookies viewing the page should remain constant between the control and treatment groups. If this metric changes significantly, it could indicate an issue with randomization or user behavior not related to the experiment d_{min}: 3000 (The difference that would need to be observed for it to be meaningful.)

Click-through Probability
Description: The ratio of unique cookies that click the "Start free trial" button divided by the number of unique cookies to view the course overview page.
Why It's a Good Invariant Metric: This metric measures the baseline interest in the free trial before the screener is triggered. Since the experiment modifies what happens after the click (the screener), the click-through probability should remain stable between the control and experiment groups. If this metric changes, it could indicate an issue with how the users interact with the page or some unintended effect of the experiment on user behavior
d_{min} : 0.01

Reasoning:
These metrics capture pre-experiment behavior. They measure users' natural behavior before encountering the experiment's impact (the screener), ensuring that both groups are balanced and randomization was successful. If these invariant metrics stay constant across both groups, it suggests that the results of the evaluation metrics can be attributed to the changes introduced by the experiment rather than to random variations or other factors.

Choosing Evaluation Metrics

Evaluation metrics are the key indicators that show whether the experiment had the intended effect. These metrics will be directly impacted by the screener and will help determine whether the change improves student experience and outcomes without significantly reducing the number of students who complete the course.

Recommended Evaluation Metrics:
Gross Conversion
Description: The ratio of the number of user-ids that complete the checkout process and enroll in the free trial to the number of unique cookies that click the "Start free trial" button.
Why It's a Good Evaluation Metric: The experiment's screener might deter students who don't have enough time to commit to the course. Therefore, gross conversion will likely change as a result of the screener. If the screener works as intended, we might see a small drop in gross conversion, filtering out students who are less likely to succeed. This is critical because Udacity wants to reduce trial enrollments by students who are likely to drop out d_{min}: 0.01 (minimum practical change for business significance)

Retention
Description: The ratio of the number of user-ids who remain enrolled past the 14-day free trial period (and thus make at least one payment) to the number of user-ids that complete the checkout process.
Why It's a Good Evaluation Metric: The main goal of the screener is to ensure that only students who have enough time and commitment proceed with the trial. By screening for time availability, we expect retention to improve—more students who enroll should stay past the trial period and make a payment. An increase in retention would suggest that the screener is effective at attracting students with a higher likelihood of completing the course
d_{min}: 0.01

Net Conversion
Description: The ratio of the number of user-ids that remain enrolled past the 14-day free trial period to the number of unique cookies that clicked the "Start free trial" button.
Why It's a Good Evaluation Metric: Net conversion is a combination of gross conversion and retention. It reflects the overall effectiveness of the screener in converting interested students into paying customers. If the screener helps to improve student retention without causing a significant drop in the number of initial enrollments, net conversion should either stay stable or improve. This metric is crucial because it balances the trade-off between filtering out students (which might lower gross conversion) and retaining committed students (which should increase retention) d_{min}: 0.0075

Calculating Standard Deviation

For each evaluation metric we will make an analytic estimate of its standard deviation, given a sample size of 5000 cookies visiting the course overview page.

Baseline values	
Unique cookies to view course overview page per day:	40000
Unique cookies to click "Start free trial" per day:	3200
Enrollments per day:	660
Click-through-probability on "Start free trial":	0.08
Probability of enrolling, given click:	0.20625
Probability of payment, given enroll:	0.53
Probability of payment, given click	0.1093125

First of all I will figure out how many units of analysis will correspond to 5000 pageviews for each evaluation metric.

Baseline values		Given pageviews
Unique cookies to view course overview page per day:	40000	5000

Unique cookies to click "Start free trial" per day:	3200	400
Enrollments per day:	660	82

Given probabilities:
- Probability of enrolling, given a click: 0.20625 (Gross Conversion)
- Probability of making a payment, given an enroll: 0.53 (Retention)
- Probability of payment, given click: 0.1093125 (Net Conversion)

1. Standard deviation formula for Gross Conversion:

$$\sigma_{gross} = \sqrt{\frac{(\rho_{gross} \times (1 - \rho_{gross})}{n_{clicks}}}$$

   Result of calculation:

$$\sigma_{gross} = \sqrt{\frac{0.20625 \times (1 - 0.20625)}{400}} = 0.02023$$

2. Standard deviation formula for Retention:

$$\sigma_{retention} = \sqrt{\frac{\rho_{retention} \times (1 - \rho_{retention})}{n_{enrolments}}}$$

   Result of calculation:

$$\sigma_{retention} = \sqrt{\frac{0.53 \times (1 - 0.53)}{82}} = 0.0551$$

3. Standard deviation formula for Net Conversion:

$$\sigma_{net} = \sqrt{\frac{\rho_{net} \times (1 - \rho_{net})}{n_{clicks}}}$$

   Result of calculation:

$$\sigma_{net} = \sqrt{\frac{0.1093125 \times (1 - 0.1093125)}{400}} = 0.0156$$

Calculating Number of Pageviews

Should I use the Bonferroni Correction?

For this A/B test the Bonferroni Correction is not strictly necessary, because I'm concerned about detecting true effects and my metrics are related (e.g., Net Conversion depends on Retention and Gross Conversion, or improvements in Gross Conversion may lead to improvements in Retention).

To calculate the required sample size (number of pageviews) for this A/B test with significance level $\alpha=0.05$ and statistical power $\beta=0.2$ (which corresponds to 80% power), we need to use following factors:

1. Significance level (α): 0.05 the Probability of a Type I error (false positive).
2. Power ($1-\beta$): 0.80 the Probability of detecting true effect (that is avoiding a Type II error).
3. Effect size: The minimum detectable difference (the practical significance boundary) for each metric. Gross Conversion d_{min}: 0.01, Retention d_{min}: 0.01, Net Conversion d_{min}: 0.0075
4. Baseline conversion rates: Gross Conversion=0.20625, Retention=0.53, Net Conversion=0.1093125.

For A/B tests, the sample size can be estimated using standard formulas, based on two-proportion z-tests, and I will adjust it for each metric's baseline conversion rate and minimum detectable effect size.

$$n = \frac{(Z_{\alpha/2}+Z_{\beta})^{2} \times (\rho_{1} \times (1-\rho_{1}) + \rho_{2} \times (1-\rho_{2}))}{(\rho_{1} - \rho_{2})^{2}}$$

Where:
$Z_{\alpha/2}$: The z-value corresponding to the two-tailed significance level (α).
Z_{β} : The z-value corresponding to desired power.
ρ_1 and ρ_2 : The baseline conversion rate and the expected conversion rate after the change.
(ρ_1 - ρ_2) : The effect size (the practical significance boundary or minimum detectable effect)

In this case
$Z_{\alpha/2}$ =1.96 (for a 95% confidence level)
Z_{β} =0.84 (for 80% power)

1. Gross Conversion:
    - baseline probability (ρ_1): Probability of enrolling, given click = 0.20625
    - minimum detectable effect d_{min}=0.01 (1% absolute difference)
    - ρ_2 expected rate after the change = 0.20625 ± 0.01 (depending on whether we expect increase or decrease)

Let's calculate for increase (ρ_2=0.20625 + 0.01=0.21625)

$$n = \frac{(1.96 + 0.84)^{2} \times (0.20625 \times (1 - 0.20625) + 0.21625 \times (1 - 0.21625))}{(0.20625 - 0.21625)^{2}}$$

$$n = \frac{7.84 \times (0.16371 + 0.16948)}{0.01^{2}}$$

$$n = \frac{7.84 \times 0.33319}{0.0001} \sim 26122$$

Required sample size for Gross Conversion (per group): 26122 clicks on "Start free trial".
Since only 8% of pageviews lead to clicks, I will need pageviews in total:
$n = 100 \times 26122/8 = 326526$ pageviews in total for each group

2. Retention:
- baseline probability (ρ_1): Probability of making a payment, given enroll = 0.53
- minimum detectable effect d_{min}=0.01 (1% absolute difference)

- ρ_2 expected rate after the change = 0.53 ± 0.01 (depending on whether we expect increase or decrease)

Let's calculate for increase (ρ_2=0.53 + 0.01=0.54)

$$n = \frac{(1.96 + 0.84)^2 \times (0.53 \times (1 - 0.53) + 0.54 \times (1 - 0.54))}{(0.53 - 0.54)^2}$$

$$n = \frac{7.84 \times (0.2491 + 0.2484)}{0.01^2}$$

$$n = \frac{7.84 \times 0.4975}{0.0001} = 39004$$

Required sample size for Retention (per group): 39004. With a Gross Conversion Rate of 20.625%, I'd need:

$\frac{39004}{0.20625} \sim 189110$ pageviews in total for each group

3. Net Conversion:
    - baseline probability (ρ_1): probability of making payment given a click ρ_1=0.1093125
    - minimum detectable effect (d_{min}): 0.0075 (0.75% absolute difference)
    - ρ_2 (expected rate after the change) = 0.1093125+0.0075=0.1168125

$$n = \frac{(1.96 + 0.84)^2 \times (0.1093125 \times (1 - 0.1093125) + 0.1168125 \times (1 - 0.1168125))}{(0.1093125 - 0.1168125)^2}$$

$$n = \frac{7.84 \times (0.0973633 + 0.10316734)}{0.0075^2}$$

$$n = \frac{7.84 \times 0.20053064}{0.00005625} = \frac{1.57216}{0.00005625} \sim 27949$$

Required sample size for Net Conversion (per group): 27949 clicks. With 8% click-through rate, this corresponds to

$n = 100 \times 27949/8 \sim 349362$ pageviews in total for each group.

Choosing duration and exposure

Recap the sample size estimates per group from the previous calculations:

- Gross Conversion requires 326526 pageviews per group, totalling 653052 pageviews for experiment.
- Retention requires 189110 pageviews per group, totalling 378220 pageviews for experiment.
- Net Conversion requires 349362 pageviews per group, totalling 698724 pageviews for the experiment.

The largest sample requirement is Net Conversion with 698000 pageviews total (across control and experiment groups) to achieve sufficient statistical power.

Unique cookies to view the course overview page per day 40000.

I would divert the 50% fraction of traffic to this experiment. It will allow me to gather data quickly enough to detect statistically significant results in a reasonable timeframe and will help to minimize exposure to potential risk by not diverting 100% of users.

If the webpage receives 40000 pageviews per day, the experiment would collect:
$0.5 \times 40000 = 20000$ pageviews per day
Then total duration of the test would be:

$$days\ required = \frac{698000}{20000} \sim 35\ days$$

These results suggest the experiment would take over five weeks if traffic averages 20000 pageviews per day, which may be too long.

Sanity Checks

For invariant metrics Number of cookies assigned to the control group and click-through-probability on "Start Free Trial" I will compute a 95% confidence interval for the values I expect to observe.
To compute a 95% confidence interval:

1. For the Number of Cookies
   - calculate the fraction assigned to the control group
     $fraction_{cont} = \frac{control\ group\ pageviews}{total\ pageviews}$
     control group pageviews 345543
     experiment group pageviews 344660

$$fraction_{cont} = \frac{345543}{345543 + 344660} = 0.5006$$

   - calculate a 95% confidence interval on proportions

$$CI = fraction_{cont} \pm m$$

$$m = Z - score \times \sqrt{\frac{0.5 \times (1 - 0.5)}{total\ pageviews}}$$

     For 95% confidence level $Z - score \sim 1.96$

$$m = 1.96 \times \sqrt{\frac{0.5 \times 0.5}{345543 + 344660}}$$

$$m = 1.96 \times 0.0006 = 0.001176$$

     Lower bound of 95% confidence interval

$$CI_{lower} = 0.5 - 0.001176 = 0.498824$$

     Upper bound of 95% confidence interval

$$CI_{upper} = 0.5 + 0.001176 = 0.501176$$

     Sanity check for Number of Cookies is passed
     0.498824<0.5006<0.501176

2. Click-through-probability for "Start free Trial"
   - calculate the observed click-through-probability ρ_{pool}

$$\rho_{pool} = \frac{X_{cont} + X_{exp}}{N_{cont} + N_{exp}}$$

     control group clicks X_{cont} 28378
     experiment group clicks X_{exp} 28325
     control group pageviews N_{cont} 345543
     experiment group pageviews N_{exp} 344660

$$\rho_{pool} = \frac{28378 + 28325}{345543 + 344660} = \frac{56703}{690203} = 0.08215$$

- calculate standard error SE

$$SE_{pool} = \sqrt{\frac{\rho_{pool} \times (1 - \rho_{pool})}{N_{cont} + N_{exp}}}$$

$$SE_{pool} = \sqrt{\frac{0.08215 \times (1 - 0.08215)}{345543 + 344660}}$$

$SE_{pool} = \sqrt{\frac{0.0754}{690203}}$ =0.0003316

- calculate margin of error m_{pool}

  $m_{pool} = SE \times z - score$

  For 95% confidence level $Z - score \sim 1.96$

  $m_{pool} = 0.0003316 \times 1.96 = 0.0006499$

- calculate the click-through-probability for control and experimental group individually

$$\rho_{cont} = \frac{28378}{345543} = 0.08212$$

$$\rho_{exp} = \frac{28325}{344660} = 0.08218$$

- calculate the difference between click-through-probabilities of experimental and control groups

$$d = \rho_{exp} - \rho_{cont}$$

$$d = 0.082182 - 0.082125 = 0.000057$$

- calculate confidence interval

$$CI = d \pm m_{pool}$$

$$CI_{lower} = 0.000057 - 0.0006499 = -0.0005929$$

$$CI_{upper} = 0.000057 + 0.0006499 = 0.0007069$$

Outcome: since m_{pool}>d (0.0006499>0.000057), confidence interval includes 0, which means that d is not significant.

Sanity checks results

invariant metric	*lower bound*	*upper bound*	*observed*	*passes*
Number of cookies	0.498824	0.501176	0.5006	yes
Click-through-probability “Start free trial”	-0.0005929	0.0007069	0.000057	yes

Effect Size Check

For each of evaluation metrics Gross Conversion, Retention, Net Conversion, compute a 95% confidence interval around the difference.
To compute a 95% confidence intervals:

1. Gross Conversion
   - compute Gross Conversion for control and experiment groups

     $$GC = \frac{E}{C}$$

     enrollments in experimental group E_{exp} 3423
     clicks “Start Free Trial” in experimental group C_{exp} 17260
     enrolments in control group E_{cont} 3785
     clicks “Start Free Trial” in control group C_{cont} 17293

     $$GC_{exp} = \frac{3423}{17260} = 0.1983$$
     $$GC_{cont} = \frac{3785}{17293} = 0.2188$$

   - compute a 95% confidence interval for Gross Conversion in the experiment group

     $CI_{exp} = GC_{exp} \pm m_{exp}$

     $$m_{exp} = Z \times \sqrt{\frac{GC_{exp} \times (1 - GC_{exp})}{C_{exp}}}$$

     where:
     - ❖ CI_{exp} confidence interval for Gross Conversion in the experiment group
     - ❖ m_{exp} is margin of error for experiment group
     - ❖ GC_{exp} is the observed proportion of the metric for experiment group
     - ❖ Z is z-score for 95% confidence interval Z=1.96
     - ❖ C_{exp} is the sample size for experiment group (the total number of clicks “Start Free Trial”)

     $m_{exp} = 1.96 \times \sqrt{\frac{0.1983\times(1-0.1983)}{17260}} = 0.0059$
     $CI_{GC\,exp} = 0.1983 + 0.0059 = 0.2042$
     $CI_{GC\,exp} = 0.1983 - 0.0059 = 0.1924$
     So the 95% confidence interval for Gross Conversion in experiment group is [0.1924, 0.2042].

   - compute a 95% confidence interval for Gross Conversion for control group

     $CI_{GC\,cont} = GC_{cont} \pm m_{cont}$

     $$m_{GC\,cont} = Z \times \sqrt{\frac{GC_{cont} \times (1 - GC_{cont})}{C_{cont}}}$$

     where:
     - ❖ $CI_{GC\,cont}$ confidence interval for Gross Conversion in the experiment group
     - ❖ $m_{GC\,cont}$ is margin of error for control group

- ❖ $GC_{GC\ ont}$ is the observed proportion of the metric for control group
- ❖ Z is z-score for 95% confidence interval Z=1.96
- ❖ $C_{GC\ cont}$ is the sample size for control group (the total number of clicks "Start Free Trial")

$m_{GC\ cont} = 1.96 \times \sqrt{\frac{0.2188\times(1-0.2188)}{17293}} = 0.0061$
$CI_{GC\ cont} = 0.2188 + 0.0061 = 0.2249$
$CI_{GC\ cont} = 0.2188 - 0.0061 = 0.2127$
So the 95% confidence interval for Gross Conversion in control group is [0.2127, 0.2249].
The confidence intervals for experiment [0.1924, 0.2042] and control [0.2127, 0.2249] groups are not intersected, therefore the Gross Conversion values are statistically significant.

- compute a 95% confidence interval around the difference d_{GC}

$$d_{GC} = GC_{exp} - GC_{cont}$$
$$d_{GC} = 0.1983 - 0.2188 = |-0.0205|$$

The difference between GC is 0.02, which is practically significant (d_{min}=0.01<0.02).

$$GC_{pool} = \frac{3423 + 3785}{17260 + 17293} = \frac{7208}{34553} = 0.2086$$

$SE_{GC\ pool} = \sqrt{\frac{GC_{pool}(1-GC_{pool})}{C_{exp}+C_{cont}}}$
$SE_{GC\ pool} = \sqrt{\frac{0.2086\times(1-0.2086)}{17260+17293}} = 0.0022$
$m_{GC\ pool} = 1.96 \times 0.0022 = 0.0043$
$CI_{GC\ pool} = 0.2086 + 0.0043 = 0.2129$
$CI_{GC\ pool} = 0.2086 - 0.0043 = 0.2043$
So the 95% confidence interval around difference for Gross Conversion is [0.2043, 0.2129].
A 95% confidence interval does not include 0, thus Gross Conversion in the experiment group is statistically significant.

2. Retention
   - compute Retention for control and experiment groups

$$R = \frac{P}{E}$$

payments in the experiment group P_{exp}=1945
enrollments in the experiment group E_{exp}=3423
payments in the control group P_{cont}=2033
enrollments in the experiment group E_{cont}=3785

$$R_{exp} = \frac{1945}{3423} = 0.5682$$

$$R_{cont} = \frac{2033}{3785} = 0.5371$$

- compute a 95% confidence interval for Retention in the experiment group

$$CI_{R\,exp} = R_{exp} \pm m_{exp}$$
$$m_{R\,exp} = Z \times \sqrt{\frac{R_{experiment} \times (1 - R_{experiment})}{E_{experiment}}}$$

where:

  - ❖ CI $_{R\,exp}$ confidence interval for Retention in experiment group
  - ❖ m $_{R\,exp}$ is margin of error for Retention in experiment group
  - ❖ R $_{exp}$ is the observed proportion of the retention
  - ❖ Z is z-score for 95% confidence interval Z=1.96
  - ❖ E $_{exp}$ is the sample size in experiment group (the total number of payments)

$m_{exp} = 1.96 \times \sqrt{\frac{0.5682 \times (1 - 0.5682)}{3423}} = 0.0166$
$CI_{R\,exp} = 0.5682 + 0.0166 = 0.5848$
$CI_{GC\,exp} = 0.5682 - 0.0166 = 0.5516$
So the 95% confidence interval for Retention in experiment group is [0.5516, 0.5848].

- compute a 95% confidence interval for Retention in the control group

$$R_{cont} = R_{cont} \pm m_{cont}$$
$$m_{R\,cont} = Z \times \sqrt{\frac{R_{cont} \times (1 - R_{cont})}{E_{cont}}}$$

where:

  - ❖ m $_{R\,cont}$ is margin of error for Retention in control group
  - ❖ R $_{cont}$ is the observed proportion of the Retention in control group
  - ❖ Z is z-score for 95% confidence interval Z=1.96
  - ❖ E $_{cont}$ is the sample size on control group (the total number of payments)

$m_{R\,cont} = 1.96 \times \sqrt{\frac{0.5371 \times (1 - 0.5371)}{3785}} = 0.0159$
$CI_{R\,cont} = 0.5371 + 0.0159 = 0.553$
$CI_{R\,cont} = 0.5371 - 0.0159 = 0.5212$
So the 95% confidence interval for Retention in control group is [0.5212, 0.553].

- compute a 95% confidence interval around d $_R$

$$d_R = R_{exp} - R_{cont}$$
$$d_R = 0.5682 - 0.5371 = |0.0311|$$

The difference between GC is 0.0311, which is practically significant (d_{min}=0.01<0.0311).

$$R_{pool} = \frac{1945 + 2033}{3423 + 3785} = \frac{3978}{7208} = 0.5518$$

$SE_{R\,pool} = \sqrt{\frac{R_{pool}(1-R_{pool})}{E_{exp}+E_{cont}}}$

$SE_{R\,pool} = \sqrt{\frac{0.5518\times(1-0.5518)}{3423+3785}} = 0.0058$

$m_{R\,pool} = 1.96 \times 0.0058 = 0.0114$

$CI_{R\,pool} = 0.5518 + 0.0114 = 0.5632$

$CI_{R\,pool} = 0.5518 - 0.0114 = 0.5404$

So the 95% confidence interval around d_R for Retention is [0.5404, 0.5632].

A 95% confidence interval does not include 0, thus Retention in the experiment group is statistically significant.

3. Net Conversion
   - compute Net Conversion for control and experiment groups

$$NC = GC \times R$$

Gross Conversion for experiment group GC_{expe}=0.198319
Retention for experiment group R_{exp}=0.5682
Gross Conversion for control group GC_{cont}=0.218875
Retention for control group R_{cont}=0.5371

$$NC_{exp} = 0.1983 \times 0.5682 = 0.1126$$

$$NC_{cont} = 0.2188 \times 0.5371 = 0.1175$$

   - compute a 95% confidence interval for Net Conversion in experiment group

Compute variance for Net Conversion in experiment group.
For the product of two random variables, the variance is approximately

$$\upsilon^2_{NC\,exp} \sim (GC^2_{exp} \times \upsilon^2_{R\,esp}) + (R^2_{exp} \times \upsilon^2_{GC\,exp})$$

Where

$$\upsilon^2_{GC\,exp} = \frac{GC_{exp} \times (1 - GC_{exp})}{C_{exp}}$$

$$\upsilon^2_{R\,exp} = \frac{R_{exp} \times (1 - R_{exp})}{E_{exp}}$$

$$\upsilon^2_{GC\,exp} = \frac{0.1983 \times (1 - 0.1983)}{17260} = 0.000009211$$

$$\upsilon^2_{R\,exp} = \frac{0.5682 \times (1 - 0.5682)}{3423} = 0.00007167$$

The variance is

$$\upsilon^2_{NC\,exp} \sim (0.198319^2 \times 0.00007167) + (0.5682^2 \times 0.000009211)$$

$$\upsilon^2_{NC\,exp} \sim (0.0393 \times 0.00007167) + (0.3228 \times 0.000009211)$$

$$\upsilon^2_{NC\,exp} \sim 0.000002816 + 0.000002973 = 0.000005789$$

The formula for standard error is

$$SE_{NC\,exp} = \sqrt{\upsilon^2_{NC\,exp}}$$

$$SE_{NC\,exp} = \sqrt{0.000005789} = 0.0024$$

The formula for margin of error is

$$m_{NC\,exp} = SE_{NC\,exp} \times Z - score$$
$$m_{NC\,exp} = 0.0024 \times 1.96 = 0.0047$$

A 95% confidence interval is

$$CI_{NC\,exp} = 0.1126 + 0.0047 = 0.1173$$
$$CI_{NC\,exp} = 0.1126 - 0.0047 = 0.1079$$

So the 95% confidence interval for Net conversion in experiment group is [0.1079, 0.1173]

- compute a 95% confidence interval for Net Conversion in control group
  Compute variance for Net Conversion in control group.
  For the product of two random variables, the variance is approximately

$$v^2_{NC\,cont} \sim (GC^2_{cont} \times v^2_{R\,cont}) + (R^2_{cont} \times v^2_{GC\,cont})$$

Where

$$v^2_{GC\,cont} = \frac{GC_{cont} \times (1 - GC_{cont})}{C_{cont}}$$
$$v^2_{R\,cont} = \frac{R_{cont} \times (1 - R_{cont})}{E_{cont}}$$
$$v^2_{GC\,cont} = \frac{0.2188 \times (1 - 0.2188)}{17293} = \frac{0.1709}{17293} = 0.000009884$$
$$v^2_{R\,cont} = \frac{0.5371 \times (1 - 0.5371)}{3785} = \frac{0.2486}{3785} = 0.00006568$$

The variance is

$$v^2_{NC\,cont} \sim (0.2188^2 \times 0.00006568) + (0.5371^2 \times 0.000009886)$$
$$v^2_{NC\,cont} \sim (0.0479 \times 0.00006568) + (0.2884 \times 0.000009886)$$
$$v^2_{NC\,cont} \sim 0.000003146 + 0.000002851 \sim 0.000005997$$

The formula for standard error is

$$SE_{NC\,cont} = \sqrt{v^2_{NC\,cont}}$$
$$SE_{NC\,cont} = \sqrt{0.000005997} = 0.002448$$

The formula for margin of error is

$$m_{NC\,cont} = SE_{NC\,cont} \times Z - score$$
$$m_{NC\,cont} = 0.002448 \times 1.96 = 0.0048$$

A 95% confidence interval is

$$CI_{NC\,cont} = 0.1175 + 0.0048 = 0.1223$$
$$CI_{NC\,cont} = 0.1175 - 0.0048 = 0.1127$$

So the 95% confidence interval for Net Conversion in control group is [0.1127, 0.1223]
The 95% confidence intervals of the control group [0.1127; 0.1223] and the experiment group [0.1079; 0.1173] do not intersect, thus the results are statistically significant.

- compute a 95% confidence interval around difference d_{NC}

$$d_{NC} = NC_{exp} - NC_{cont}$$
$$d_{NC} = 0.1126 - 0.1175 = |-0.0049|$$

The difference between NC is 0.0049, which is not practically significant (d_{min}=0.0075>0.0049).
Compute pooled GC using the formula

$$GC_{pool} = \frac{E_{cont} + E_{exp}}{C_{cont} + C_{exp}}$$

$$GC_{pool} = \frac{3785 + 3423}{17293 + 17260} = \frac{7208}{34553} = 0.2086$$

Compute pooled R using the formula

$$R_{pool} = \frac{P_{cont} + P_{exp}}{E_{cont} + E_{exp}}$$

$$R_{pool} = \frac{2033 + 1945}{3785 + 3423} = \frac{3978}{7208} = 0.5518$$

Compute pooled NC using the formula

$$NC_{pool} = GC_{pool} \times R_{pool}$$

$NC_{pool} = 0.2086 \times 0.5518 = 0.1151$

Compute variance for pooled Gross Conversion

$$v^2_{pool} = \frac{GC_{pool} \times (1 - GC_{pool})}{C_{cont} + C_{exp}}$$

$$v^2_{GC\,pool} = \frac{0.2086 \times (1 - 0.2086)}{17293 + 17260} = \frac{0.165}{34553} = 0.000004775$$

Compute variance for pooled Retention

$$v^2_{R\,pool} = \frac{R_{pool} \times (1 - R_{pool})}{E_{cont} + E_{exp}}$$

$$v^2_{R\,pool} = \frac{0.5518 \times (1 - 0.5518)}{3785 + 3423} = \frac{0.2473}{7208} = 0.00003431$$

Apply the Delta Method for the product's variance

$$v^2_{NC\,pool} \sim (GC^2_{pool} \times v^2_{R\,pool}) + (R^2_{pool} \times v^2_{GC\,pool})$$

$$v^2_{NC\,pool} \sim (0.2086^2 \times 0.00003431) + (0.5518^2 \times 0.000004775)$$

$$v^2_{NC\,pool} \sim (0.0435 \times 0.00003431) + (0.3044 \times 0.000004775)$$

$$v^2_{NC\,pool} \sim 0.000001492 + 0.00000006326 = 0.000001555$$

Pooled Standard Error for the pooled Net Conversion is

$$SE_{NC\,pool} = \sqrt{v^2_{NC\,pool}}$$

$$SE_{NC\,pool} = \sqrt{0.000001555} = 0.001247$$

Margin of error is

$$m_{NC\,pool} = SE_{NC\,pool} \times Z - score$$

$$m_{NC\,pool} = 0.001247 \times 1.96 = 0.0024$$

A 95% confidence interval for NC_{pool} is

$$CI_{NC\,pool} = 0.1151 + 0.0024 = 0.1175$$

$$CI_{NC\,pool} = 0.1151 - 0.0024 = 0.1127$$

So the 95% confidence interval for pooled Net Conversion is [0.1127, 0.1175] and does not include 0, thus the results are statistically significant.

Effect Size Results

evaluation metric	lower bound	upper bound	statistical significance	practical significance
Gross conversion	0.204316	0.212884	yes	yes
Retention	0.54032	0.5632	yes	yes
Net Conversion	0.1127	0.1175	yes	no

Sign Tests

Compute p-value for evaluation metrics Gross Conversion, Retention, Net condersion.

- compute the p-value for Gross Conversion using two-proportion z-test formula

$$z = \frac{(GC_{control} - GC_{experiment})}{\sqrt{GC_{pooled} \times (1 - GC_{pooled}) \times (\frac{1}{C_{control}} + \frac{1}{C_{experiment}})}}$$

$$z = \frac{0.2188 - 0.1983}{\sqrt{0.2086 \times (1 - 0.2086) \times (\frac{1}{17260} + \frac{1}{17293})}}$$

$$z = \frac{0.0205}{\sqrt{0.165 \times (0.00005794 + 0.00005783)}} = \frac{0.0205}{\sqrt{0.165 \times 0.00011577}} = 4.691$$

$$p - value = 2 \times (1 - \Phi(|z|))$$

Since $\Phi(|4.691|)$ is very close to 1, these results in a p-value close to 0, indicating strong evidence against the null hypothesis.

- compute the p-value for Retention using two-proportion z-test formula

$$z = \frac{(R_{cont} - R_{exp})}{\sqrt{R_{pool} \times (1 - R_{pool}) \times (\frac{1}{E_{cont}} + \frac{1}{E_{expt}})}}$$

$$z = \frac{0.5371 - 0.5682}{\sqrt{0.5518 \times (1 - 0.5518) \times (\frac{1}{3423} + \frac{1}{3785})}}$$

$$z = \frac{-0.0311}{\sqrt{0.2473 \times (0.0002921 + 0.0002642)}} = \frac{-0.0311}{\sqrt{0.2473 \times 0.0005563}} = -2.6581$$

$$p - value = 2 \times (1 - \Phi(|z|))$$

$$\rho - value = 2 \times (1 - \Phi(|-2.6581|)$$

$$p - value = 2 \times (1 - 0.996) = 0.008$$

p-value less than α (0.008<0.05), thus the result of Retention in the experiment group is statistically significant.

- compute p-value for the Net Conversion

$$z = \frac{NC_{exp} - NC_{cont}}{SE_{NC\,pool}}$$

$$z = \frac{0.1126 - 0.1175}{0.0024} = \frac{-0.0049}{0.0024} = -2.0416$$

Convert Z-score to p-value

$$p - value = 2 \times (1 - \Phi(|-2.0416|)$$
$$p - value = 2 \times (1 - 0.9793) = 0.0414$$

p-value is approximately 0.0414, indicating statistical significance at the 5% level if the result below 0.05.

Sign Tests Results

evaluation metric	*p-value*	*statistical significance*
Gross conversion	0	yes
Retention	0.008	yes
Net Conversion	0.0414	yes

Conclusion

Let's recall our hypothesis:
The null hypothesis: the free trial screener will not impact on evaluation metrics, visitors do not see the start free trial screener.
The alternative hypothesis: the free trial screener will impact the evaluation metrics, visitors see the start free trial screener.
And look through the results

Results of testing evaluation metrics

evaluation metric	*control group*	*experiment group*	*difference*	*d min*	*p-value*	*practical significance*	*statistical significance*
Gross conversion	0.2188	0.1983	-0.0205	0.01	0	yes	yes
Retention	0.5371	0.5682	0.0311	0.01	0.008	yes	yes
Net Conversion	0.1175	0.1126	-0.0049	0.0075	0.0414	no	yes

First of all I want to point out that the results of calculations are statistically significant due to the correct set up of A/B testing. We see the predictably reduced for 2% Gross Conversion in the experiment group, because visitors became more aware of pushing the "Start Free Trial" button after the free trial screener demonstration. The Net Conversion reduced for

0.49% the same time which is not practically significant. The Retention raised for 3% which demonstrated that more enrolled students paid for courses in the experimental group compared with the control group.
The start free trial demonstration leads to better awareness of website visitors about courses and a more informed choice of paid courses, which in turn increases students’ willingness to enroll in paid courses and reduces students’ frustration with the quality of teaching in general. I would recommend implementing the start free trial screener.