

Experiment Design

Metric Choice

Métricas Invariantes

Number of cookies

Number of clicks

Click-through-probability

Métricas de Avaliação

Retention

Gross Conversion

Net Conversion

Métricas invariantes são métricas que deveriam ser a mesma tanto no grupo de controle quanto no grupo do experimento. As métricas escolhidas ocorrem em estágios no site antes da mudança (mensagem de texto alertando sobre tempo mínimo de dedicação) e portanto deveriam permanecer a mesma no grupo de controle e no grupo de experimento. Isso permitirá checar se os grupos são equivalentes. Não vou utilizar a métrica Number user-id pois ela mostra a quantidade de pessoas que se inscreveram na free trial e esperamos que esse numero diminua após o experimento.

As métricas de avaliação são as métricas que deveriam variar entre os grupos de controle e de experimento mostrando a efetividade (ou não) do experimento. Escolhi a métrica Retention pois a quantidade de estudantes que permanecem após a free trial deveria aumentar. Também escolhi as métricas Gross Conversion e Net Conversion pois elas deveriam mostrar que existem menos estudantes se inscrevendo na free trial mas a quantidade de estudantes que permanecem após experimento deveria permanecer o mesmo.

Measuring Standard Deviation

Dado que todas as métricas de avaliação são probabilidades assumi que a distribuição delas é uma distribuição de Bernoulli e utilizei a formula abaixo:

$$RAIZ (p * (1 - p) / N)$$

p -> probabilidade

N -> população

Retention

N -> (tamanho amostra cookies * matriculas por dia / cookies exclusivos para visualizar paginas por dia)

$N \rightarrow 5000 * 660 / 40000 = 82,5$

$p \rightarrow 0,53$

$sd \rightarrow 0,054949$

Gross Conversion

$N \rightarrow (\text{tamanho amostra cookies}) * \text{probabilidade de click-through}$

$N \rightarrow 5000 * 0,08 = 400$

$p \rightarrow 0,20625$

$sd \rightarrow 0,20231$

Net Conversion

$N \rightarrow (\text{tamanho amostra cookies}) * \text{probabilidade de click-through}$

$N \rightarrow 5000 * 0,08 = 400$

$p \rightarrow 0,1093125$

$sd \rightarrow 0,015602$

Para Gross Conversion e Net Conversion, a unidade de desvio é igual à unidade de análise. Assim, a estimativa analítica do desvio padrão tende a ser uma estimativa empírica do desvio padrão. Depois de calcular as visualizações de página, duração e exposição, se decidimos manter a retenção, temos que calcular a variabilidade empírica de retenção.

Sizing

Number of Samples vs. Power

Eu não utilizei a correção de Bonferroni, as métricas não são independentes e a correlação de Bonferroni seria muito conservadora para este caso. Utilizei a calculadora online disponível no link <http://www.evanmiller.org/ab-testing/sample-size.html>.

Retention

Baseline conversion rate: 0,53

dmin: 0,01

α : 0,05

$1 - \beta$: 0,8

Sample size per variation: 39.115

Sample size: $39115 * 2 = 78230$

Pageviews: $78230 / 0,8 = 4739879$

Gross Conversion

Baseline conversion rate: 0,20625

dmin: 0,01

α : 0,05

$1 - \beta$: 0,8

Sample size per variation: 25.835

Sample size: $25.835 * 2 = 51670$

Pageviews: $51670 / 0,8 = 642475$

Net Conversion

Baseline conversion rate: 0,1093125

dmin: 0,0075

α : 0,05

$1 - \beta$: 0,8

Sample size per variation: 24.413

Sample size: $24.413 * 2 = 54826$

Pageviews: $54826 / 0,8 = 679300$

A métrica retention exige um número muito alto de page views: 4.739.879 o que faz com que ela não seja uma métrica muito acessível já que o tempo e custo disso poderia ser muito alto. Dessa forma mantereí as métricas Net Conversion e Gross conversion. Como Net conversion é a métrica que dentre as duas mais exige page view, vou manter esse valor como o número necessário de page views.

Duration vs. Exposure

Não acredito que tenha nenhum risco para a Udacity como perda de informações sensíveis ou aumento no tráfico de dados. Também não identifiquei nenhum risco de algum usuário se machucar ou se prejudicar com esse experimento e dessa forma acho que poderia ser coletado 100% do trafico de dados que é de 40000 page views por dia. Dessa forma precisaríamos de 17 dias para coletar a quantidade necessárias de páginas.

Experiment Analysis

Sanity Checks

Number of cookies

Tcont: 345543

Texp: 344660

se -> $\text{raiz}((0,5 * 0,5) / (345543 + 344660)) = 0,0006$

m -> $0,0006 * 1,96 = 0,0012$

CI -> $0,5 - 0,0012 = 0,4988$

$0,5 + 0,0012 = 0,5012$
 $p \rightarrow 345543 / (345543 + 344660) = 0,5006$
Passou: Sim

Number of clicks

Tcont: 28378
Texp: 28325

$se \rightarrow \text{raiz}((0,5 * 0,5) / (28378 + 28325)) = 0,0021$
 $m \rightarrow 0,00180536 * 1,96 = 0,0041$
 $CI \rightarrow 0,5 - 0,003538506 = 0,4959$
 $0,5 + 0,003538506 = 0,5041$
 $p \rightarrow 28378 / (28378 + 28325) = 0,5005$
Passou: Sim

Click-through-probability

Tcont: 0,082125814
Texp: 0,082182441

Pcont $\rightarrow 0,0821$
Pexp $\rightarrow 0,0822$
 $P_{\text{pool}} = (28378 + 28325) / (345543 + 344660) = 0,082154$
 $d = 0,0822 - 0,0821 = 0,0000566$
 $SE_{\text{pool}} = \text{raiz}(0,0821 * (1 - 0,0821) * (1/345543 + 1/344660)) = 0,00066$
 $m \rightarrow 0,00066 * 1,96 = 0,0013$
 $CI \rightarrow -0,0013 - 0,0013$
Passou: Sim

Result Analysis

Effect Size Tests

Gross Conversion

dmin = 0,01
ncontrol = 17293
enroll_cont = 3785
nexp = 17260
enroll_exp = 3423

pcont $\rightarrow 3785/172393 = 0,2189$

pexp -> $3423 / 17260 = 0,1983$
d = $0,1983 - 0,2189 = -0,0206$

ppool -> $(3423+3785)/(17293+17260) = 0,2086$
SEpool -> $\text{raiz}(0,2086 * (1 - 0,2086) * (1/172393 + 1/17260)) = 0,0044$
M -> $0,0044 * 1,96 = 0,0086$
CI -> $-0,0206 - 0,0086 = -0,0291$
 $-0,0206 + 0,0086 = -0,0120$
estatisticamente significativa: Sim
praticamente significativa: Sim

Net Conversion

dmin = 0.0075
ncontrol = 17293
enroll_cont = 2033
nexp = 17260
enroll_exp = 1945

pcont -> $2033/17293 = 0,1176$
pexp -> $1945 / 17260 = 0,1127$
d = $0,1176 - 0,1127 = -0,0049$

ppool -> $(2033+1945)/(17293+17260) = 0,1151$
SEpool -> $\text{raiz}(0,1151 * (1 - 0,1151) * (1/172393 + 1/17260)) = 0,0034$
M -> $0,0034 * 1,96 = 0,0067$
CI -> $-0,0049 - 0,0067 = -0,0116$
 $-0,0049 + 0,0067 = 0,0019$
estatisticamente significativa: Não
praticamente significativa: Não

Sign Tests

Gross Conversion

Alpha = 0,05
Success: 4
Trial: 23
p = 0,0026
estatisticamente significativa: Sim

Net Conversion

Alpha = 0,05

Success: 10

Trial: 23

p = 0,6776

estatisticamente significativa: Não

Summary

O experimento foi conduzido para identificar se exibir a pergunta no site de quantas horas os usuários poderiam disponibilizar por semana para conclusão do curso antes de clicar no botão de matrícula irá melhorar a experiência do aluno e capacidade dos treinadores para apoiar os alunos que são susceptíveis de completar o curso, sem reduzir significativamente o número de alunos que se inscrevem. As métricas invariantes são Número de cookies, Número de cliques e Probabilidade de cliques. Conversão bruta e conversão líquida são as métricas de avaliação. A hipótese nula é que não há diferença entre controle e grupo experimental.

A correção de Bonferroni é usada apenas para reduzir as chances de identificar pelo menos um resultado significativo devido ao acaso na presença de métricas múltiplas. Como meus experimentos requerem duas métricas de avaliação, conversão bruta e conversão líquida, para serem válidas para o lançamento, nenhuma correção de Bonferroni foi usada.

Os resultados sob o teste de sinal e teste de tamanho de efeito são consistentes entre si, ambos mostrando que a conversão bruta é estatisticamente significativa, enquanto a conversão líquida não

Recommendation

Para lançar o experimento seria esperado dos resultados que após a inserção da pergunta menos pessoas estariam se inscrevendo para o teste gratuito mas isso não mudaria a quantidade de alunos que permaneciam no curso após 14 dias. Ao analisar as métricas percebemos que a conversão bruta é menor no grupo de experimentos, o que significa que menos pessoas estão se inscrevendo para o teste gratuito como o esperado.

No entanto não é possível afirmar que isso não reduziu de forma significativa o número de alunos que permaneceram após 14 dias (após free trial) já que conversão líquida não é estatisticamente significativa e o intervalo de confiança de conversão líquida inclui o valor negativo do limite de significância prática. Isso indica que eu não posso estar confiante sobre o tamanho desse efeito no nível de confiança de 95%. Com esse resultado é possível inferir uma perda no total de alunos que estão continuando após 14 dias e por isso não recomendo o lançamento do experimento.

Follow-Up Experiment

Como próximo experimento, eu adicionaria a caixa que pergunta quantas horas o aluno possui disponível para o estudo, uma lista de pré-requisitos para curso bem como uma lista de cursos prévios que seriam interessante o aluno fazer antes de iniciar o curso (como por exemplo estatística). Isso poderia evitar que alunos percebessem que não possuem todo o background necessário antes de clicar no botão iniciar.

As métricas que eu utilizaria seriam:

Métricas Invariantes

Number of cookies

Number of clicks

Click-through-probability

Métricas de Avaliação

Retention

Gross Conversion

Net Conversion

A unidade de desvio seria cookie para alunos não cadastrados e user-id uma vez que estivessem cadastrados.

Referencias

<https://discussions.udacity.com/t/sanity-checks-confidence-intervals-observed-values/27175/17>