Experiments and Analytics

Why do we conduct experiments?

- Limited knowledge/understanding
- Limited time
- Limited resources

How to conduct experiments

- The scientific method: a method of research in which a problem is identified, relevant data are gathered, a hypothesis is formulated from these data, and the hypothesis is empirically tested.
    - observe → construct a hypothesis/prediction → test → draw conclusions/generalize

How to conduct experiments

The basic principles of designing experiments:

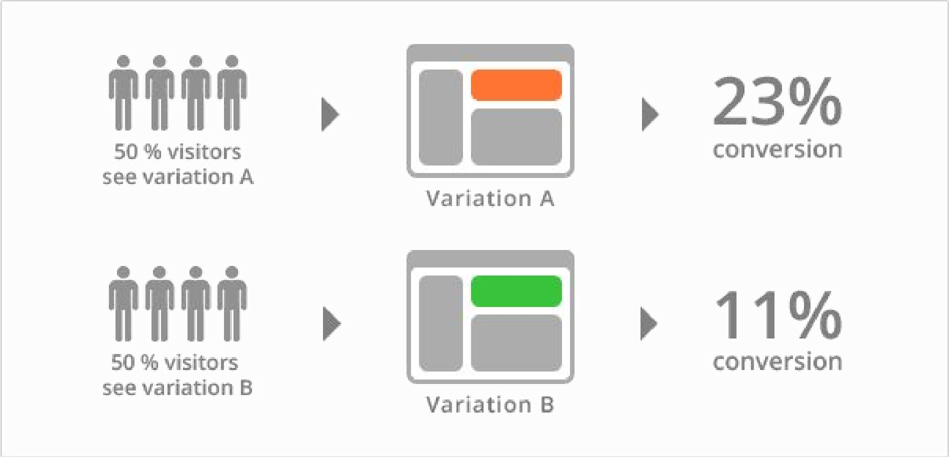
- Randomization: You need to have a deliberate process to eliminate potential biases from the conclusions through random assignment.
- Replication: We want to estimate or control the uncertainty in our results; i.e., reduce the error in our estimates.
- Blocking: We need to control for factors in our experiment which contribute to undesirable variation in data; i.e., reduce the impact of uncontrolled variations on the outcome.
    - e.g.: experiment on different days of the week
- Confounding: We need to watch out for confounding variables (the situation where the effect of two factors cannot be separated from each other)

A/B Testing

- **A/B testing** emerged originally as a method of comparing two versions of a webpage against each other to determine which one performs better.
- In general, it is a term used for a randomized experiment with two variants, A and B, which are the control and variation in the controlled experiment.
  - Also known as two-sample hypothesis testing in the field of statistics
- Now widely applied for website optimization for content, creative, messaging as well as offer testing.

A/B Testing

- By keeping all other elements exactly the same, you can determine the relative effectiveness of the two test variants.



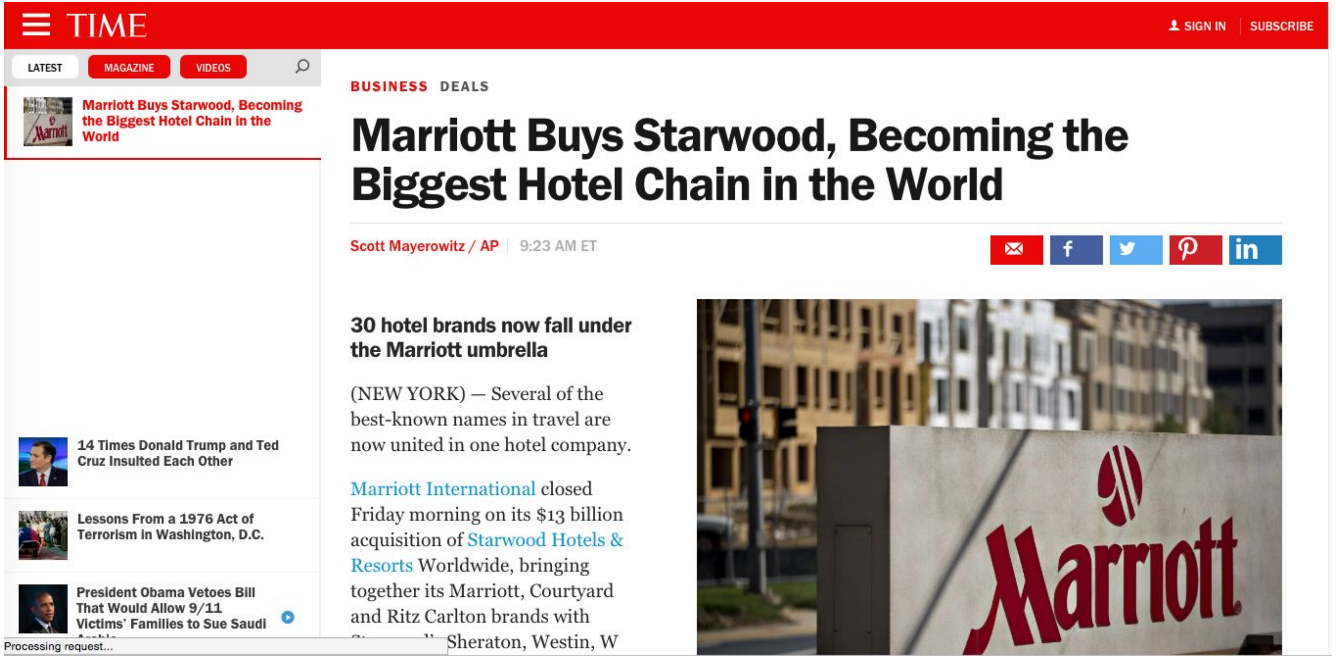
What to test:

- Headline
- Call to Action
- Images
- Video
- Copy
- Colors
- Forms
- Confirmation pages

- Testing two elements is the minimum, but you could test three elements at the same time. If, for example, you are testing three lists, you would have an A/B/C split.

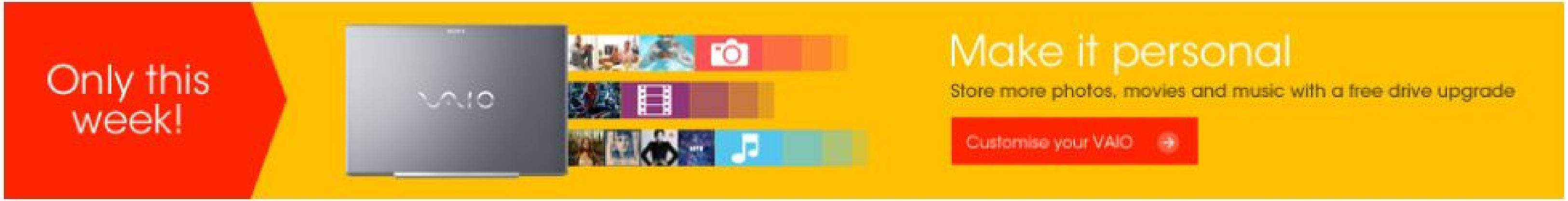
A/B Testing Use Case

- A media company wants to increase the amount of time readers spend on their site and promote their articles with social sharing.



A/B Testing Case Study

- Banner ads promoting customizable laptops on Sony's homepage and product pages were underperforming – clickthroughs and purchases from the banners were extremely low.
- Two different calls-to-action: one advertising the customizable Sony Vaio notebooks and one promoting a recent sitewide offer for a free memory upgrade.



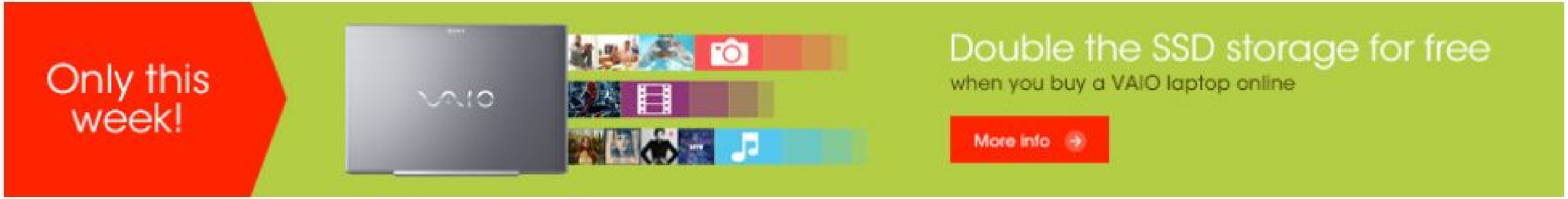
The original banner displayed messaging around the personalization campaign and the promotion.

A/B Testing Case Study

- Hypothesis: dual call-to-action messages was confusing and overwhelming to site visitors, making them less likely to click through on the banner.
- Test: A/B/C test with two different variations of the banner against the original targeted only to site visitors from the Netherlands and the United Kingdom – two of Sony's largest markets for the customizable laptop campaign.
    - One variation focused entirely on the customizable laptops, while the other foc the sitewide offer.

A/B Testing Case Study

Variation 1 focused entirely on the idea of customization. The banner included a “Vaio by you” logo, and a call-to-action to create and customize a laptop. The banner did not reference the sitewide offer.



Variation 2 focused on the promotional aspect, showing specific messaging to create urgency around the offer. This variation made no reference to personalization.

A/B Testing Case Study

- What is your interpretation of the test results?

Population	Impressions	Clicks	Added to Cart
Control	39,010	468	140
Variation 1	38,978	496	181
Variation 2	39,004	477	139

A/B Testing Case Study

- How confident are you in your conclusions???

Statistical Significance

- Statistical significance is the likelihood that the difference in conversion rates between a given variation and the baseline is not due to random chance.
- A result of an experiment is statistically significant if it is likely not caused by chance for a given statistical significance level.
- Your statistical significance level reflects your risk tolerance and confidence level. For example, if you run an A/B testing experiment with a significance level of 95%, this means that if you determine a winner, you can be 95% confident that the observed results are real and not caused by randomness. It also means that there is a 5% chance that you could be wrong.

Statistical Significance

- Two key variables for determining statistical significance:
  - sample size
  - effect size
- **Sample size** refers to how large the sample for your experiment is. The larger your sample size, the more confident you can be in the result of the experiment (assuming that it is a randomized sample).
- Effect size is the difference in test outcome due to the experiment. If the effect is small (for example, 0.1% increase in conversion rate) you will need a very large sample size to determine whether that difference is significant or just due to chance.

Populations, Samples, Statistics

- A **population** is any large collection of objects or individuals, such as Americans, students, or trees about which information is desired.
- A **parameter** is any summary number, like an average or percentage, that describes the entire population.
- A **sample** is a representative group drawn from the population.
- A **statistic** is any summary number, like an average or percentage, that describes the sample.
    - The population mean μ vs the sample mean $\bar{x}$

Populations, Samples, Statistics

- How are samples and sample statistics useful?
  - E.g.: the average weight of a random sample of 100 middle-aged female Americans to estimate the average weight of all middle-aged female Americans.
  - Because samples are manageable in size, we can determine the actual value of any statistic.

    We use the known value of the sample statistic to learn about the unknown value of the population parameter.

Confidence Intervals

- NYU has a population of approximately 57,000 students. A research question is "what proportion of these students consume alcohol regularly?" A survey was administered to a sample of 1000 NYU students. 43 percent of the sampled students reported that they drink alcohol regularly.
- How confident can we be that 43% is close to the actual proportion of all NYU students who consume alcohol?
- A **confidence interval** gives an estimated range of values which is likely to include an unknown population parameter, the estimated range being calculated from a given set of sample data.

Sample estimate ± margin of error

Confidence Intervals

Sample estimate ± **margin of error**

Sample mean ± (t-multiplier × standard error)

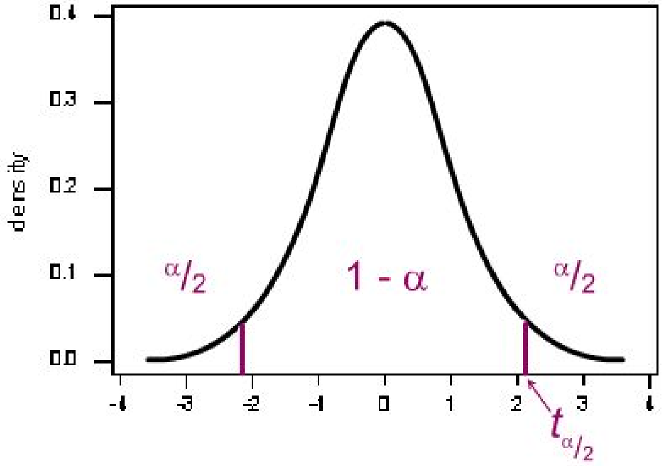
- t-interval for a population mean μ = t-multiplier × standard error
  - the margin of error in estimating a population mean μ is calculated by multiplying the t-multiplier by the standard error of the sample mean.
- The formula for the confidence interval in notation is:

$$\bar{x} \pm t_{\alpha/2,n-1}\left(\frac{s}{\sqrt{n}}\right)$$

- t-multiplier depends on the sample size (n-1) and confidence level (α/2)
- The "standard error," which is standard deviation/$\sqrt{n}$ quantifies how much the sample means vary from sample to sample.
- This formula is only appropriate if the data are **normally distributed**.
-

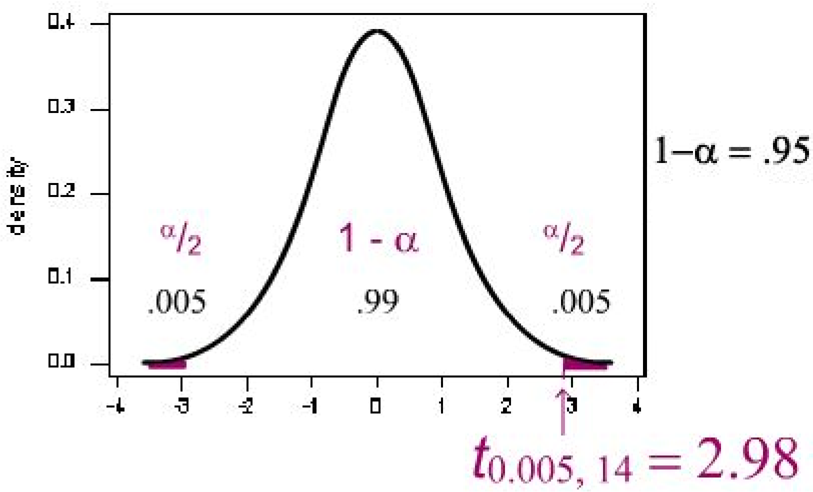
Confidence Intervals

- The entire probability represented by the curve must equal 1.
  - a probability of α must be shared equally among the two "tails" of the distribution.
- The t-multiplier, (α/2), is the t-value such that the probability "to the right of it" is α/2.



Confidence Intervals

- If our sample contained n = 15 measurements, yielding 14 degrees of freedom (from n-1 = 15-1 = 14)



t-distribution Table

$F(t)$	.75	.80	.85	.90	.95	.975	.98	.99	.995	.9975	.999	.9995
$\nu(\alpha)$	(.25)	(.20)	(.15)	(.10)	(.05)	(.025)	(.02)	(.01)	(.005)	(.0025)	(.001)	(.0005)
1	1.000	1.376	1.963	3.078	6.314	12.71	15.89	31.82	63.66	127.3	318.3	636.6
2	.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.09	22.33	31.60
3	.765	.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.21	12.94
4	.741	.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	.727	.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.894	6.859
6	.718	.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	.711	.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.405
8	.706	.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355	3.833	4.501	5.041
9	.703	.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250	3.690	4.297	4.781
10	.700	.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169	3.581	4.144	4.587
11	.697	.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106	3.497	4.025	4.437
12	.695	.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055	3.428	3.930	4.318
13	.694	.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012	3.372	3.852	4.221
14	.692	.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977	3.326	3.787	4.140
15	.691	.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947	3.286	3.733	4.073
16	.690	.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921	3.252	3.686	4.015
17	.689	.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898	3.222	3.646	3.965
18	.688	.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878	3.197	3.610	3.922
19	.688	.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861	3.174	3.579	3.883
20	.687	.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845	3.153	3.552	3.850
21	.686	.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831	3.135	3.527	3.819
22	.686	.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819	3.119	3.505	3.792
23	.685	.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807	3.104	3.485	3.767
24	.685	.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797	3.091	3.467	3.745
25	.684	.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787	3.078	3.450	3.725
26	.684	.856	1.058	1.315	1.706	2.056	2.162	2.479	2.779	3.067	3.435	3.707
27	.684	.855	1.057	1.314	1.703	2.052	2.158	2.473	2.771	3.057	3.421	3.690
28	.683	.855	1.056	1.313	1.701	2.048	2.154	2.467	2.763	3.047	3.408	3.674
29	.683	.854	1.055	1.311	1.699	2.045	2.150	2.462	2.756	3.038	3.396	3.659
30	.683	.854	1.055	1.310	1.697	2.042	2.147	2.457	2.750	3.030	3.385	3.646

Confidence Intervals

- A Z distribution is a standard normal distribution, and it can be used to construct confidence intervals in situations where the sample size is large.
- If the sample size is less than 30, a t-distribution must be used.

Z-scores for Common Confidence Intervals

Prob = r	z-score = z_r
0.90	1.645
0.95	1.96
0.98	2.326
0.99	2.576

Back to Case Study

Population	Impressions	Clicks	CTR
Control	39,010	468	1.20%
Variation 1	38,978	496	1.27%

$$\bar{x} \pm t_{\alpha/2,n-1}\left(\frac{s}{\sqrt{n}}\right)$$

Confidence level	95%
z score	1.96

Test Results

	Sample Size	Clicks	Response Rate	Std Dev	confidence limit (+)	confidence limit (-)
Control	39,010	468	1.20%	0.10887	1.308%	1.092%
Variation 1	38,978	496	1.27%	0.11209	1.384%	1.161%
				0.11049	*= pooled standard deviation*	
				0.00079	*= standard deviation of (test - control)*	

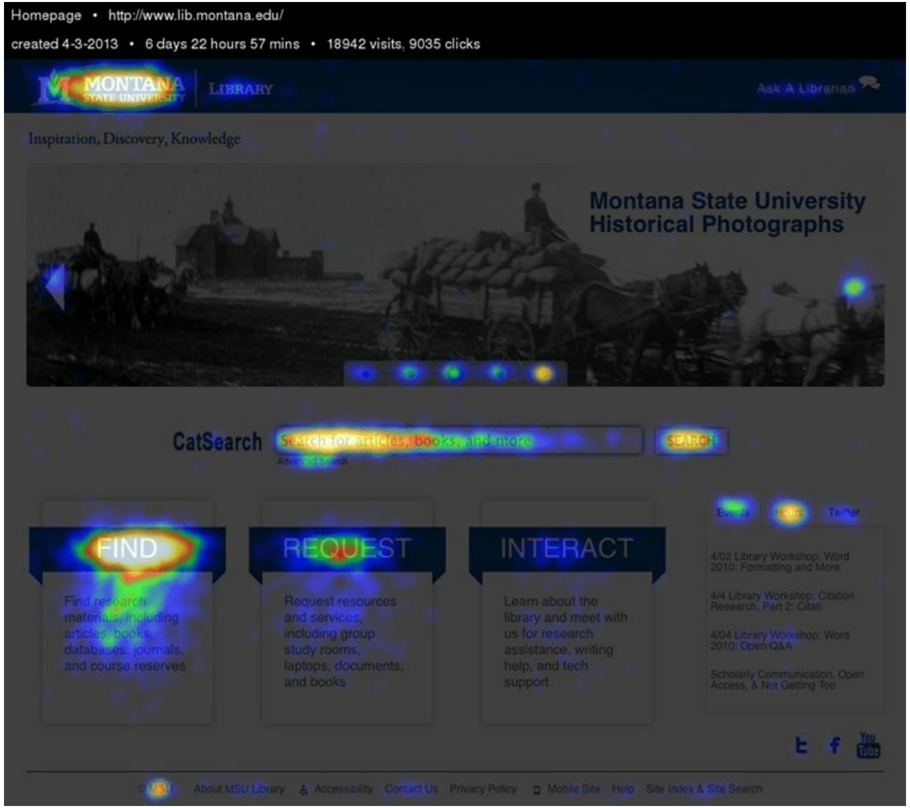
Paired t-test (test-control)

Confidence Limits for the difference in response rate (t-c)

Test - control	Confidence Interval (±)	Max difference	Min difference (-)	% Gain	Max	Min	*Significant?*
-0.073%	0.155%	0.082%	-0.228%	-5.7%	6.5%	-17.9%	No

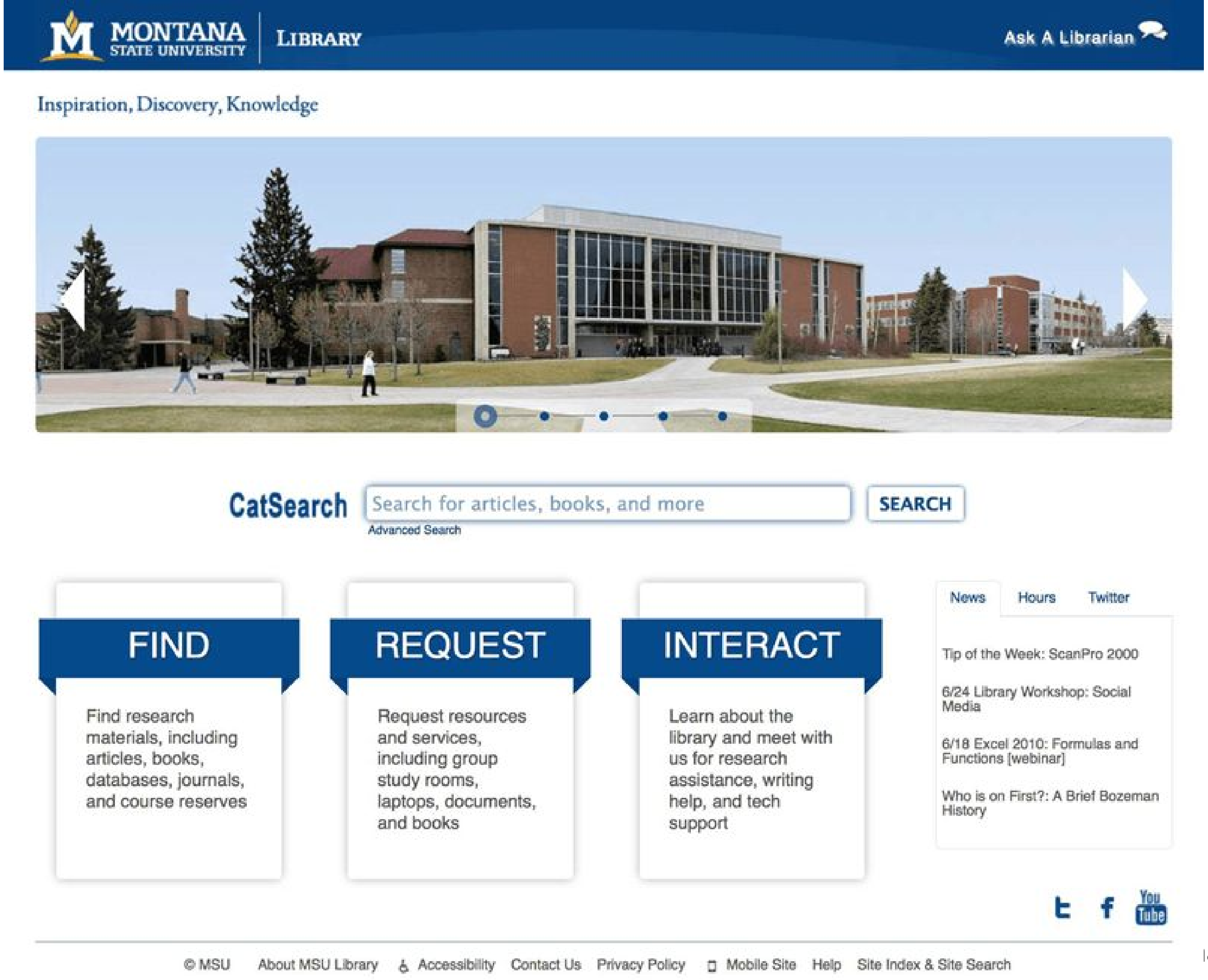
Montana University Case Study

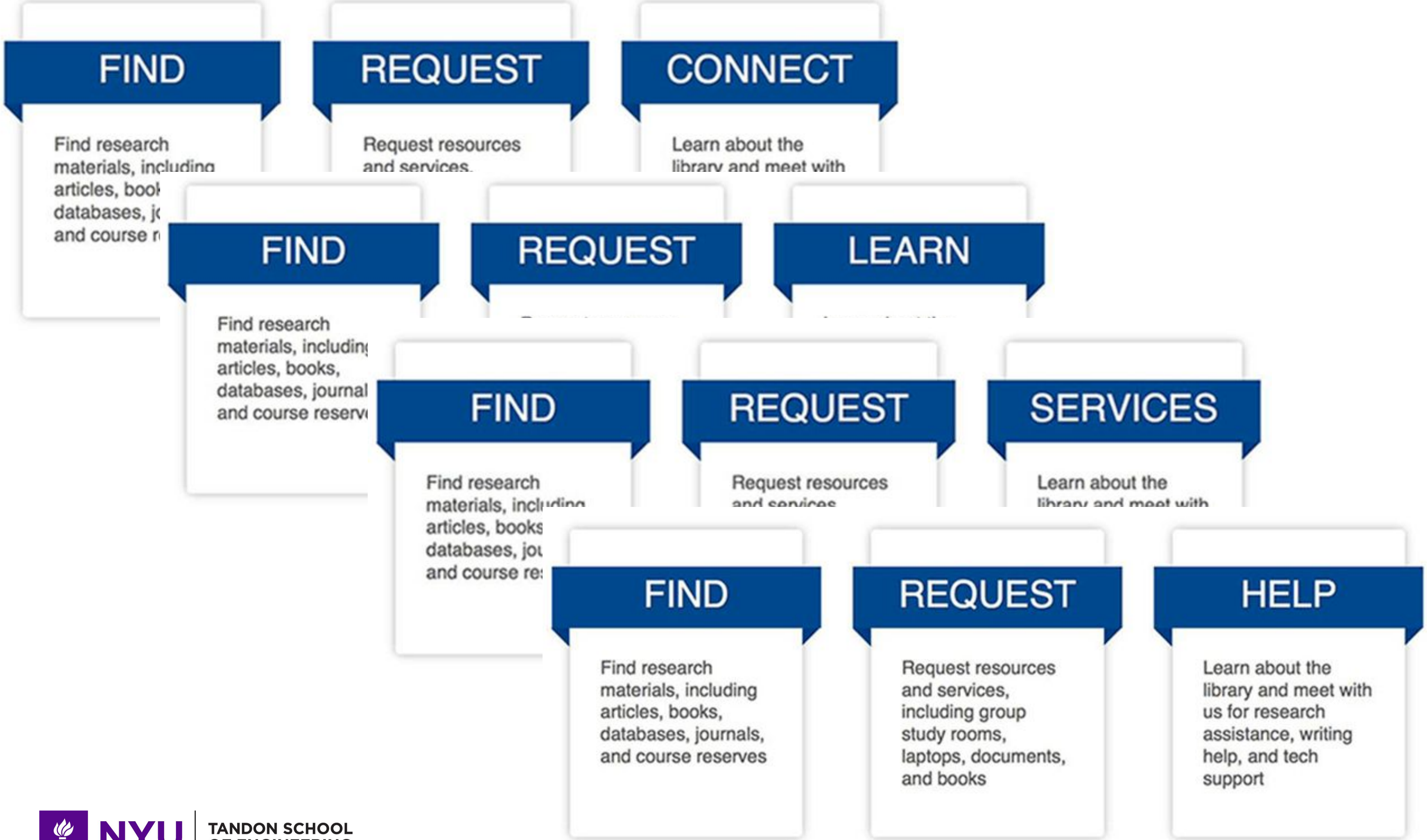
http://quod.lib.umich.edu/w/weave/12535642.0001.101?view=text;rgn=main

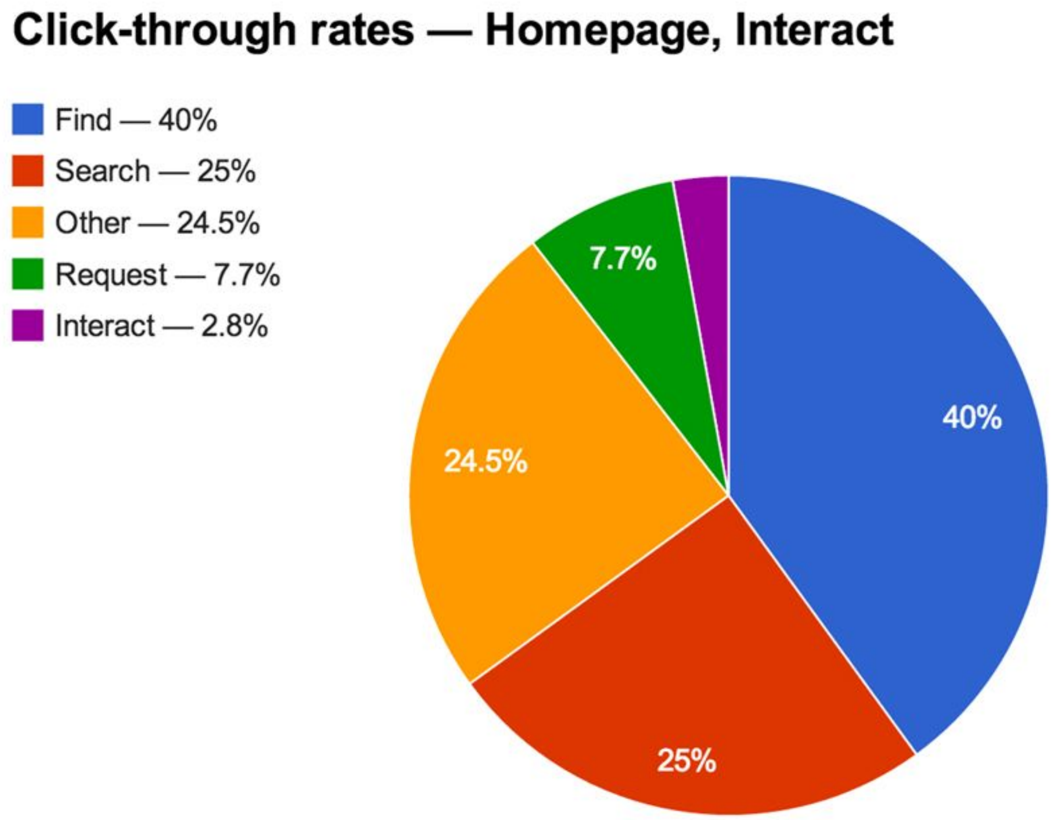


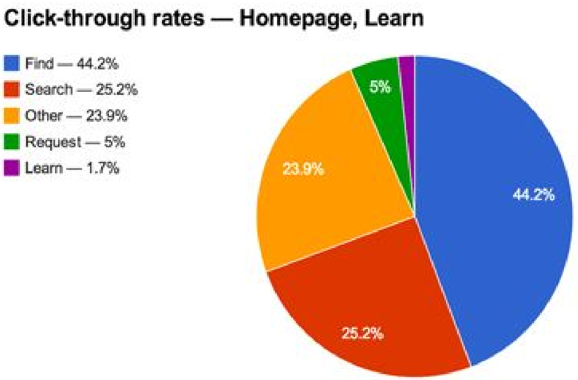
This observation prompted a question: "Why are Interact clicks so low?"

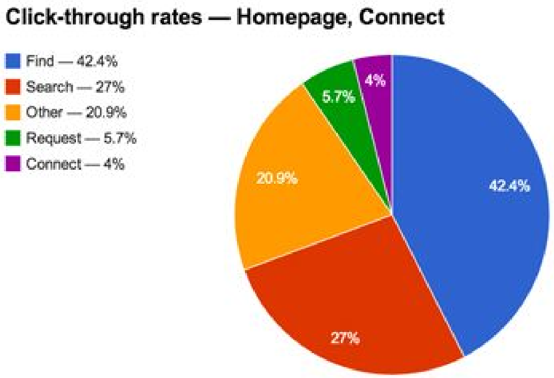
Four different category titles were then proposed as variations to be tested: Connect, Learn, Help, and Services.

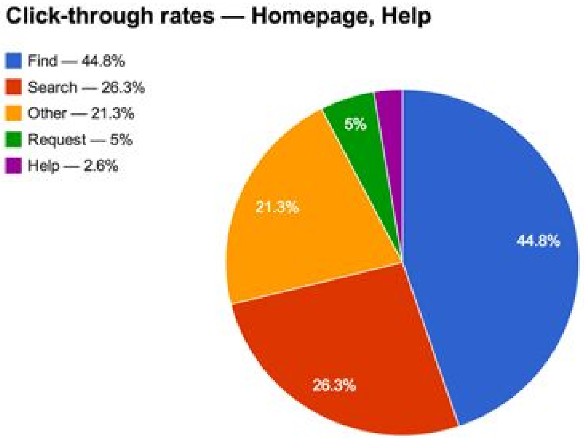

MONTANA STATE UNIVERSITY
LIBRARY
Ask A Librarian
Inspiration, Discovery, Knowledge
CatSearch
Search for articles, books, and more
SEARCH
Advanced Search
FIND
Find research materials, including articles, books, databases, journals, and course reserves
REQUEST
Request resources and services, including group study rooms, laptops, documents, and books
INTERACT
Learn about the library and meet with us for research assistance, writing help, and tech support
News
Hours
Twitter
Tip of the Week: ScanPro 2000
6/24 Library Workshop: Social Media
6/18 Excel 2010: Formulas and Functions [webinar]
Who is on First?: A Brief Bozeman History
© MSU
About MSU Library
Accessibility
Contact Us
Privacy Policy
Mobile Site
Help
Site Index & Site Search

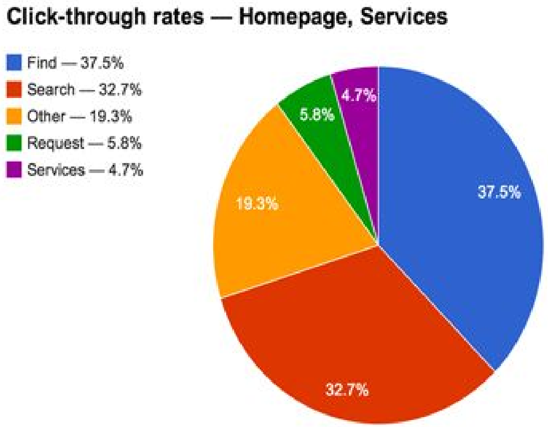

FIND
REQUEST
CONNECT
Find research
Request resources
Learn about the
FIND
REQUEST
LEARN
Find research
articles, books,
FIND
REQUEST
SERVICES
Find research
Request resources
Learn about the
FIND
REQUEST
HELP
Find research materials, including articles, books, databases, journals, and course reserves
Request resources and services, including group study rooms, laptops, documents, and books
Learn about the library and meet with us for research assistance, writing help, and tech support


Click-through rates — Homepage, Interact
Find — 40%
Search — 25%
Other — 24.5%
Request — 7.7%
Interact — 2.8%
7.7%
40%
24.5%
25%


Click-through rates — Homepage, Learn
Find — 44.2%
Search — 25.2%
Other — 23.9%
Request — 5%
Learn — 1.7%
44.2%
25.2%
23.9%
5%


Click-through rates — Homepage, Connect
Find — 42.4%
Search — 27%
Other — 20.9%
Request — 5.7%
Connect — 4%
42.4%
27%
20.9%
5.7%
4%


Click-through rates — Homepage, Help
Find — 44.8%
Search — 26.3%
Other — 21.3%
Request — 5%
Help — 2.6%
44.8%
26.3%
21.3%
5%


Click-through rates — Homepage, Services
Find — 37.5%
Search — 32.7%
Other — 19.3%
Request — 5.8%
Services — 4.7%
37.5%
32.7%
19.3%
5.8%
4.7%