A/B Testing
Comparing *k* proportions

Comparing *k* populations

Two approaches

- Statistics approach
- Computer Science approach

Comparing two populations

Compare two headlines A and B

	A	B
Click	405	380
No click	495	570
	900	950

Does headline A have a higher rate over headline B

Comparing three populations

Compare headlines A,B and C

	A	B	C
Click	405	380	490
No click	495	570	510
Visits	900	950	1000

Which headline results in the largest click-rate?

Comparing three populations

Compare headlines A,B and C

	A	B	C
Click	405	380	490
No click	495	570	510
Visits	900	950	1000

This is a table of *observed* frequencies

Comparing three populations

Compare headlines A,B and C

	A	B	C
Click	405	380	490
No click	495	570	510
Visits	900	950	1000

Compare to a table of *expected* frequencies

Chi-square variables Theorem

If $(Z_1, Z_2, \ldots Z_n)$ are independent standard normal variables, then

$$\chi_1^2 = Z_1^2$$

$$\chi_k^2 = Z_1^2 + Z_2^2 \cdots + Z_k^2$$

Chi-square test of hypothesis

To test Ho: $p_1 = p_2 = \ldots p_k$

use the tables of observed frequencies and expected frequencies

$$\chi^2_{k-1} = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

Testing $p_1 - p_2$ from two populations

X_1 the number of successes in n_1 trials from population 1 $\quad X_1 \sim BINO(n_1, p_1)$

X_2 the number of successes in n_2 trials from population 2 $\quad X_2 \sim BINO(n_2, p_2)$

Testing $p_1 - p_2$ from two populations

X_1 the number of successes in n_1 trials from population 1 $\quad X_1 \sim BINO(n_1, p_1)$

X_2 the number of successes in n_2 trials from population 2 $\quad X_2 \sim BINO(n_2, p_2)$

$$\hat{p}_1 = \frac{X_1}{n_1} \qquad \hat{p}_1 \sim N\left[p_1, \frac{p_1(1-p_1)}{n_1}\right]$$

$$\hat{p}_2 = \frac{X_2}{n_2} \qquad \hat{p}_2 \sim N\left[p_2, \frac{p_2(1-p_2)}{n_2}\right] \qquad \hat{p}_1 - \hat{p}_2 \sim N\left[p_1 - p_2, \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}\right]$$

$$Z = \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\dfrac{p_1(1-p_1)}{n_1} + \dfrac{p_2(1-p_2)}{n_2}}}$$

Testing $p_1 - p_2$ from two populations

To test

$$H_0 : p_1 = p_2$$

$$H_a : p_1 > p_2$$

or

$$H_0 : p_1 - p_2 = 0$$

$$H_a : p_1 - p_2 > 0$$

use

Test Statistic (TS) $$Z = \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}}$$

Critical Value (CV) Z_α

Obs. Test Statistic (OTS) $$z_0 = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}}$$

p-value $P[Z > z_0]$

Testing $p_1 - p_2$ from two populations

To test

$$H_0 : p_1 = p_2$$

$$H_a : p_1 > p_2$$

or

$$H_0 : p_1 - p_2 = 0$$

$$H_a : p_1 - p_2 > 0$$

use

Test Statistic (TS) $\quad Z = \dfrac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\dfrac{p_1(1-p_1)}{n_1} + \dfrac{p_2(1-p_2)}{n_2}}}$

Critical Value (CV) $\quad Z_\alpha$

Obs. Test Statistic (OTS) $\quad z_0 = \dfrac{\hat{p}_1 - \hat{p}_2}{\sqrt{\dfrac{p_1(1-p_1)}{n_1} + \dfrac{p_2(1-p_2)}{n_2}}}$

p-value $\quad$ P[Z> z_0]

Testing $p_1 - p_2$ from two populations

.

To test

$$H_0 : p_1 = p_2$$

$$H_a : p_1 > p_2$$

or

$$H_0 : p_1 - p_2 = 0$$

$$H_a : p_1 - p_2 > 0$$

use

Test Statistic (TS) $$Z = \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}}$$

Critical Value (CV) Z_α

Obs. Test Statistic (OTS) $$z_0 = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\frac{\hat{p}_p(1-\hat{p}_p)}{n_1} + \frac{\hat{p}_p(1-\hat{p}_p)}{n_2}}}$$

p-value P[Z> z_0]

Testing $p_1 - p_2$ from two populations

use the *pooled* fraction of successes

$$
\begin{aligned}
\hat{p}_p &= \frac{x_1 + x_2}{n_1 + n_2} \\
&= \frac{x_1}{n_1 + n_2} + \frac{x_2}{n_1 + n_2} \\
&= \frac{n_1}{n_1 + n_2}\left(\frac{x_1}{n_1}\right) + \frac{n_2}{n_1 + n_2}\left(\frac{x_2}{n_2}\right) \\
&= \frac{n_1}{n_1 + n_2}\ \hat{p}_1 \ + \frac{n_2}{n_1 + n_2}\ \hat{p}_2
\end{aligned}
$$

Testing $p_1 - p_2$ Example

Compare two headlines A and B

	A	B
Click	405	380
No click	495	570
	900	950

Does headline A have a higher rate over headline B

Testing $p_1 - p_2$ Example

$$H_0 : p_A = p_B \quad n_A = 900 \quad \hat{p}_A = 0.45$$

$$H_a : p_A > p_B \quad n_B = 950 \quad \hat{p}_B = 0.40 \quad Z_\alpha = 1.645$$

pooled fraction of successes

	A	B
Click	405	380
No click	495	570
	900	950

$$\hat{p}_p = \frac{405 + 380}{900 + 950} = 0.42432$$

Testing $p_1 - p_2$ Example

the observed test statistic

$$
\begin{aligned}
z_0 &= \frac{\hat{p}_A - \hat{p}_B}{\sqrt{\frac{\hat{p}_p(1-\hat{p}_p)}{n_1} + \frac{\hat{p}_p(1-\hat{p}_p)}{n_2}}} \\
&= \frac{0.45 - 0.40}{\sqrt{0.24427\left(\frac{1}{900} + \frac{1}{950}\right)}} \\
&= 2.17486
\end{aligned}
$$

$$
\begin{aligned}
\text{p-value} &= P[Z > 2.17486] \\
&= 1 - \texttt{pnorm(2.17486)} \\
&= 0.01482
\end{aligned}
$$

Comparing *k* populations

Which one is preferable?

Comparing *k* populations

X_1 the number of successes in n_1 trials from population 1 $\quad X_1 \sim BINO(n_1, p_1)$

X_2 the number of successes in n_2 trials from population 2 $\quad X_2 \sim BINO(n_2, p_2)$

$$\vdots \qquad\qquad\qquad\qquad \vdots$$

X_k the number of successes in n_k trials from population k $\quad X_k \sim BINO(n_k, p_k)$

If n. trials is large, these variables are close to a normal variable

Comparing *k* populations

$$\hat{p}_1 = \frac{X_1}{n_1} \qquad \hat{p}_1 \sim N\left[p_1, \frac{p_1(1-p_1)}{n_1}\right] \qquad Z_1 = \frac{\hat{p}_1 - p_1}{\sqrt{\frac{p_1(1-p_1)}{n_1}}}$$

$$\hat{p}_2 = \frac{X_2}{n_2} \qquad \hat{p}_2 \sim N\left[p_2, \frac{p_2(1-p_2)}{n_2}\right] \qquad Z_2 = \frac{\hat{p}_2 - p_2}{\sqrt{\frac{p_2(1-p_2)}{n_2}}}$$

$$\vdots \qquad \vdots \qquad \vdots$$

$$\hat{p}_k = \frac{X_k}{n_k} \qquad \hat{p}_k \sim N\left[p_k, \frac{p_k(1-p_k)}{n_k}\right] \qquad Z_k = \frac{\hat{p}_k - p_k}{\sqrt{\frac{p_k(1-p_k)}{n_k}}}$$

Comparing *k* populations

$$\hat{p}_1 = \frac{X_1}{n_1} \qquad \hat{p}_1 \sim N\left[p_1, \frac{p_1(1-p_1)}{n_1}\right] \qquad Z_1 = \frac{\hat{p}_1 - p_1}{\sqrt{\frac{p_1(1-p_1)}{n_1}}}$$

$$\hat{p}_2 = \frac{X_2}{n_2} \qquad \hat{p}_2 \sim N\left[p_2, \frac{p_2(1-p_2)}{n_2}\right] \qquad Z_2 = \frac{\hat{p}_2 - p_2}{\sqrt{\frac{p_2(1-p_2)}{n_2}}}$$

$$\vdots \qquad \vdots \qquad \vdots$$

$$\hat{p}_k = \frac{X_k}{n_k} \qquad \hat{p}_k \sim N\left[p_k, \frac{p_k(1-p_k)}{n_k}\right] \qquad Z_k = \frac{\hat{p}_k - p_k}{\sqrt{\frac{p_k(1-p_k)}{n_k}}}$$

$$\mathrm{H}_0 : p_1 = p_2 = \cdots = p_k \qquad \chi_k^2 = Z_1^2 + Z_2^2 \cdots + Z_k^2$$

Comparing *k* populations

$$\hat{p}_1 = \frac{X_1}{n_1} \qquad \hat{p}_1 \sim N\left[p_1, \frac{p_1(1-p_1)}{n_1}\right] \qquad Z_1 = \frac{\hat{p}_1 - p_1}{\sqrt{\frac{p_1(1-p_1)}{n_1}}}$$

$$\hat{p}_2 = \frac{X_2}{n_2} \qquad \hat{p}_2 \sim N\left[p_2, \frac{p_2(1-p_2)}{n_2}\right] \qquad Z_2 = \frac{\hat{p}_2 - p_2}{\sqrt{\frac{p_2(1-p_2)}{n_2}}}$$

$$\vdots \qquad \vdots \qquad \vdots$$

$$\hat{p}_k = \frac{X_k}{n_k} \qquad \hat{p}_k \sim N\left[p_k, \frac{p_k(1-p_k)}{n_k}\right] \qquad Z_k = \frac{\hat{p}_k - p_k}{\sqrt{\frac{p_k(1-p_k)}{n_k}}}$$

$$\mathrm{H}_0 : p_1 = p_2 = \cdots = p_k = p_0$$

$$\mathrm{H}_1 : \text{at least one is different}$$

$$\chi_k^2 = Z_1^2 + Z_2^2 \cdots + Z_k^2$$

Comparing *k* populations

$$\hat{p}_1 = \frac{X_1}{n_1} \qquad \hat{p}_1 \sim N\left[p_1, \frac{p_1(1-p_1)}{n_1}\right] \qquad Z_1 = \frac{\hat{p}_1 - p_1}{\sqrt{\frac{p_1(1-p_1)}{n_1}}}$$

$$\hat{p}_2 = \frac{X_2}{n_2} \qquad \hat{p}_2 \sim N\left[p_2, \frac{p_2(1-p_2)}{n_2}\right] \qquad Z_2 = \frac{\hat{p}_2 - p_2}{\sqrt{\frac{p_2(1-p_2)}{n_2}}}$$

$$\vdots \qquad \vdots \qquad \vdots$$

$$\hat{p}_k = \frac{X_k}{n_k} \qquad \hat{p}_k \sim N\left[p_k, \frac{p_k(1-p_k)}{n_k}\right] \qquad Z_k = \frac{\hat{p}_k - p_k}{\sqrt{\frac{p_k(1-p_k)}{n_k}}}$$

$$\mathrm{H}_0 : p_1 = p_2 = \cdots = p_k = p_0$$

$$\mathrm{H}_1 : \text{at least one is different}$$

$$\chi_k^2 = Z_1^2 + Z_2^2 \cdots + Z_k^2$$

Comparing *k* populations

$$\hat{p}_1 = \frac{X_1}{n_1} \qquad \hat{p}_1 \sim N\left[p_1, \frac{p_1(1-p_1)}{n_1}\right] \qquad Z_1 = \frac{\hat{p}_1 - p_1}{\sqrt{\frac{p_1(1-p_1)}{n_1}}}$$

$$\hat{p}_2 = \frac{X_2}{n_2} \qquad \hat{p}_2 \sim N\left[p_2, \frac{p_2(1-p_2)}{n_2}\right] \qquad Z_2 = \frac{\hat{p}_2 - p_2}{\sqrt{\frac{p_2(1-p_2)}{n_2}}}$$

$$\vdots \qquad \vdots \qquad \vdots$$

$$\hat{p}_k = \frac{X_k}{n_k} \qquad \hat{p}_k \sim N\left[p_k, \frac{p_k(1-p_k)}{n_k}\right] \qquad Z_k = \frac{\hat{p}_k - p_k}{\sqrt{\frac{p_k(1-p_k)}{n_k}}}$$

$$\mathrm{H}_0 : p_1 = p_2 = \cdots = p_k = p_0$$

$$\mathrm{H}_1 : \text{at least one is different}$$

$$\chi_k^2 = Z_1^2 + Z_2^2 \cdots + Z_k^2$$

Comparing *k* populations

$$\begin{aligned}
\chi_k^2 &= \sum_{i=1}^{k} \left(\frac{\hat{p}_i - p_0}{\sqrt{\frac{p_0(1-p_0)}{n_i}}} \right)^2 \\
&= \sum_{i=1}^{k} \frac{n_i \left(\hat{p}_i - p_0\right)^2}{p_0(1-p_0)} \\
&= \sum_{i=1}^{k} \frac{n_i \left(\hat{p}_i - p_0\right)^2}{p_0(1-p_0)} \frac{n_i}{n_i} \\
&= \sum_{i=1}^{k} \frac{(n_i\hat{p}_i - n_i\, p_0)^2}{n_i\, p_0(1-p_0)} \\
&= \sum_{i=1}^{k} \frac{(x_i - n_i\, p_0)^2}{n_i\, p_0(1-p_0)}
\end{aligned}$$

Comparing *k* populations

$$\chi^2_k = \sum_{i=1}^{k} \left(\frac{\hat{p}_i - p_0}{\sqrt{\frac{p_0(1-p_0)}{n_i}}} \right)^2$$

$$= \sum_{i=1}^{k} \frac{n_i\,(\hat{p}_i - p_0)^2}{p_0(1-p_0)}$$

$$= \sum_{i=1}^{k} \frac{n_i\,(\hat{p}_i - p_0)^2}{p_0(1-p_0)} \frac{n_i}{n_i}$$

$$= \sum_{i=1}^{k} \frac{(n_i\hat{p}_i - n_i\,p_0)^2}{n_i\,p_0(1-p_0)}$$

$$= \sum_{i=1}^{k} \frac{(x_i - n_i\,p_0)^2}{n_i\,p_0(1-p_0)}$$

if p_0 is unknown,

use the *pooled* fraction of successes $\hat{p}_p$

$$\hat{p}_p = \frac{x_1 + x_2 + \cdots + x_k}{n_1 + n_2 + \cdots + n_k}$$

Comparing *k* populations

$$\chi_k^2 = \sum_{i=1}^{k} \left(\frac{\hat{p}_i - p_0}{\sqrt{\frac{p_0(1-p_0)}{n_i}}} \right)^2$$

$$= \sum_{i=1}^{k} \frac{n_i\,(\hat{p}_i - p_0)^2}{p_0(1-p_0)}$$

$$= \sum_{i=1}^{k} \frac{n_i\,(\hat{p}_i - p_0)^2}{p_0(1-p_0)} \frac{n_i}{n_i}$$

if p_0 is unknown,

$$= \sum_{i=1}^{k} \frac{(n_i\hat{p}_i - n_i\,p_0)^2}{n_i\,p_0(1-p_0)}$$

use the *pooled* fraction of successes $\hat{p}_p$

$$= \sum_{i=1}^{k} \frac{(x_i - n_i\,p_0)^2}{n_i\,p_0(1-p_0)} \qquad = \sum_{i=1}^{k} \frac{(x_i - n_i\,\hat{p}_p)^2}{n_i\,\hat{p}_p(1-\hat{p}_p)}$$

Comparing *k* populations

$$\chi_k^2 = \sum_{i=1}^{k} \frac{(x_i - n_i\,\hat{p}_p)^2}{n_i\,\hat{p}_p(1-\hat{p}_p)}$$

$$\chi_0^2 = \sum_{i=1}^{k} \sum_{j=1}^{2} \frac{(f_{ij} - e_{ij})^2}{e_{ij}}$$

f_{ij} observed frequency in row i and column j
e_{ij} expected frequency in row i and column j

Comparing *k* populations

$$\mathrm{H}_0 : p_1 = p_2 = \cdots = p_k = p_0$$

$$\mathrm{H}_1 : \text{at least one is different}$$

OTS $$\chi_0^2 = \sum_{i=1}^{k} \sum_{j=1}^{2} \frac{(f_{ij} - e_{ij})^2}{e_{ij}}$$

if $\chi_0^2 > \chi_{k-1,1-\alpha}^2$ reject H_0

Comparing *k* populations

$$H_0 : p_1 = p_2 = \cdots = p_k = p_0$$

$$H_1 : \text{at least one is different}$$

If Ho is rejected,

which one is preferable?

Comparing $k = 3$ populations

Compare headlines A,B and C

	A	B	C
Click	405	380	490
No click	495	570	510
Visits	900	950	1000

Which headline results in the largest click-rate?

Comparing $k = 3$ populations

```
test = binom.test(405,900)
test = binom.test(380,950)
test = binom.test(490,1000)
```

Comparing $k = 3$ populations

```
test = binom.test(405,900)
test = binom.test(380,950)
test = binom.test(490,1000)
```

	headlines	means	lls	uls
1	A	0.45	0.4171515	0.4831768
2	B	0.40	0.3686726	0.4319493
3	C	0.49	0.4585849	0.5214742

Use CIs to choose the one with the best proportion

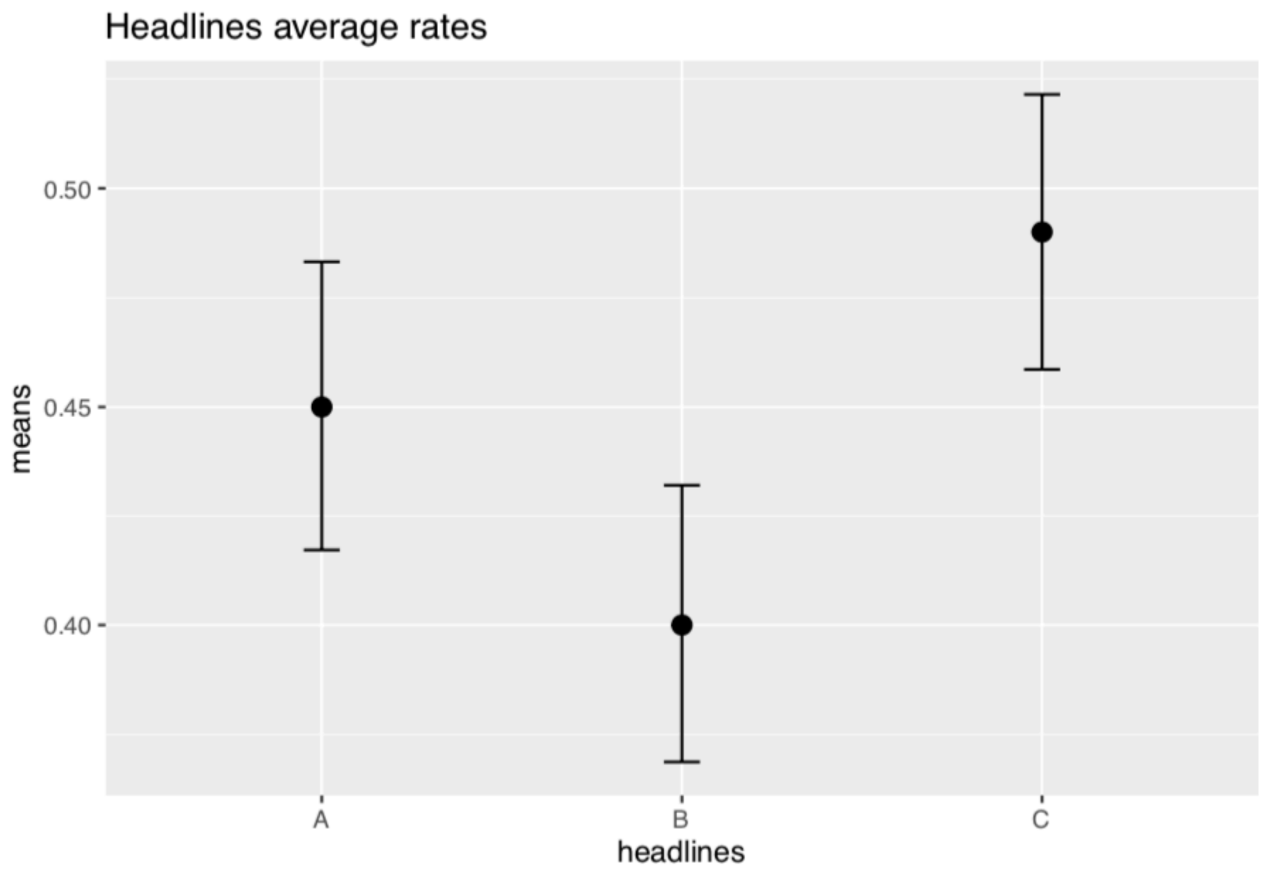
Comparing $k = 3$ populations

```
test = binom.test(405,900)
test = binom.test(380,950)
test = binom.test(490,1000)
```

	headlines	means	lls	uls
1	A	0.45	0.4171515	0.4831768
2	B	0.40	0.3686726	0.4319493
3	C	0.49	0.4585849	0.5214742

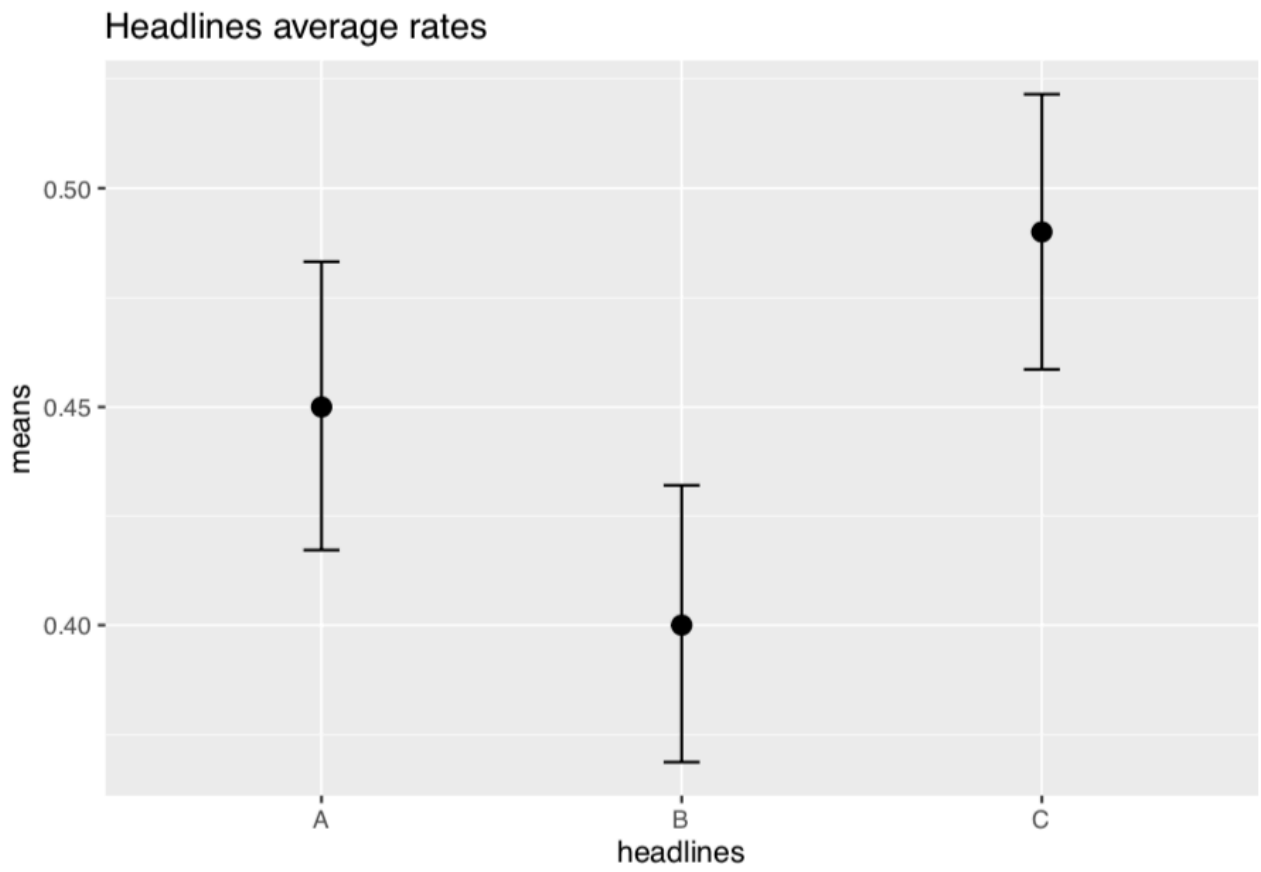
Use CIs to choose the one with the best proportion

Comparing *k* populations



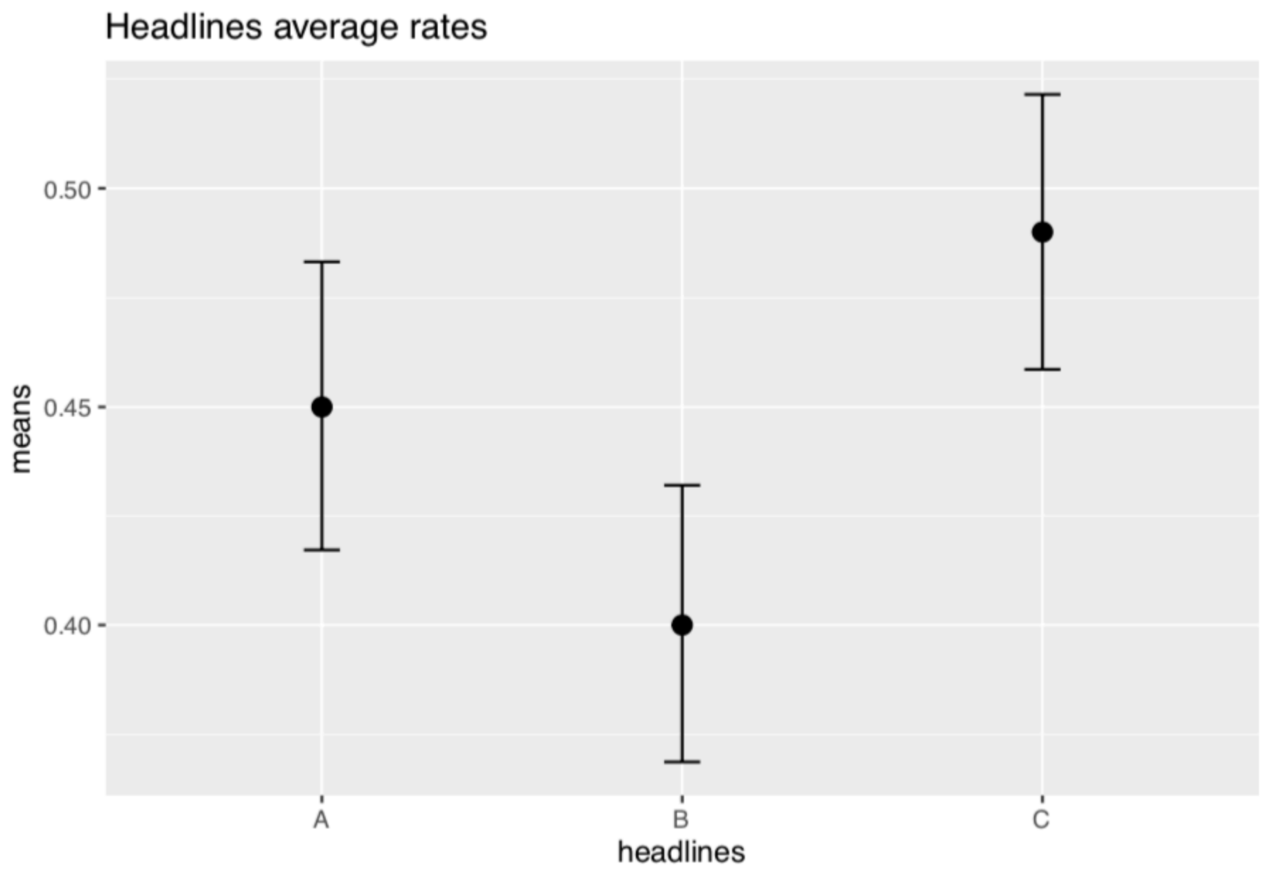
Use CIs to choose the one with the best proportion

Comparing *k* populations



Headline C better than headline B, not sure if than headline A

Comparing *k* populations



Keep headline C or collect more data to better compare with headline A

Comparing three populations

Compare headlines A,B and C

	A	B	C
Click	4050	3800	4900
Visits	9000	9500	10000

more data

Comparing three populations

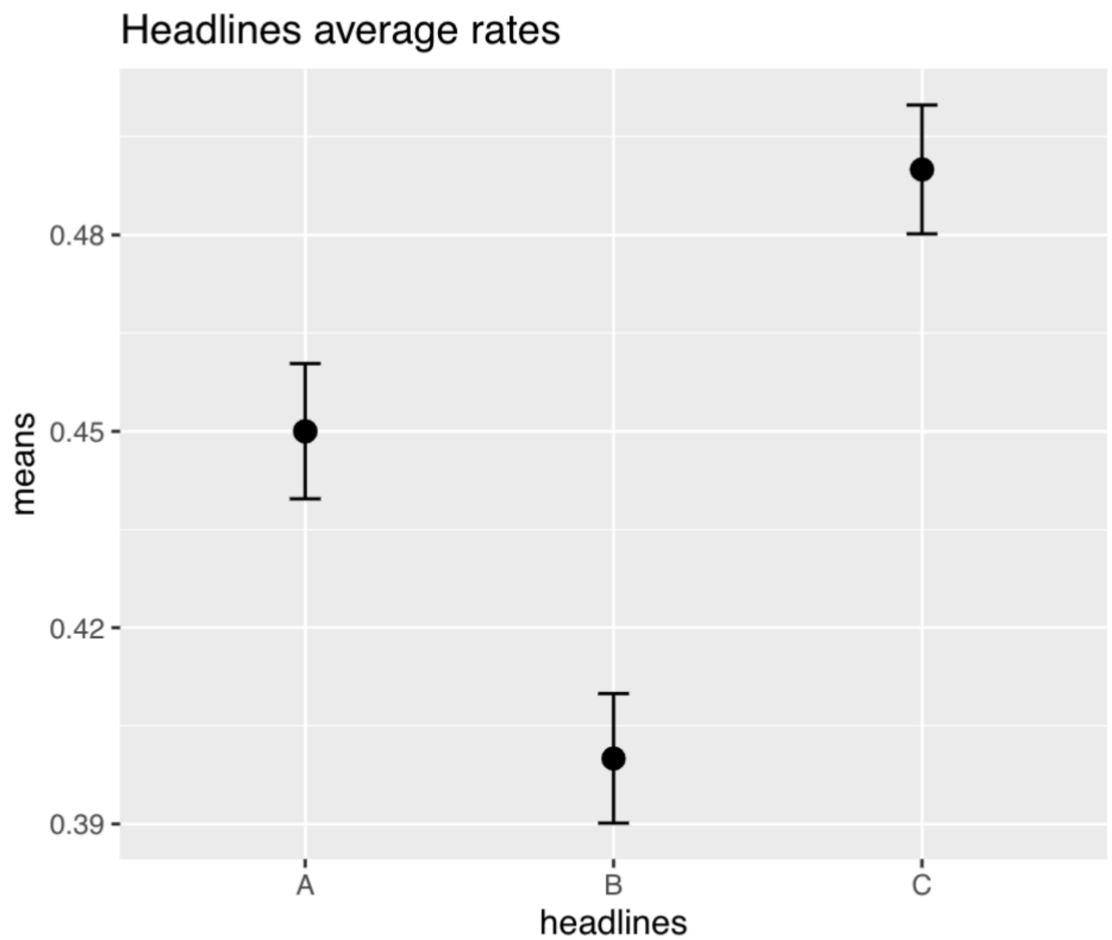
Compare headlines A,B and C

	A	B	C
Click	4050	3800	4900
Visits	9000	9500	10000

Same sample proportions, but from a

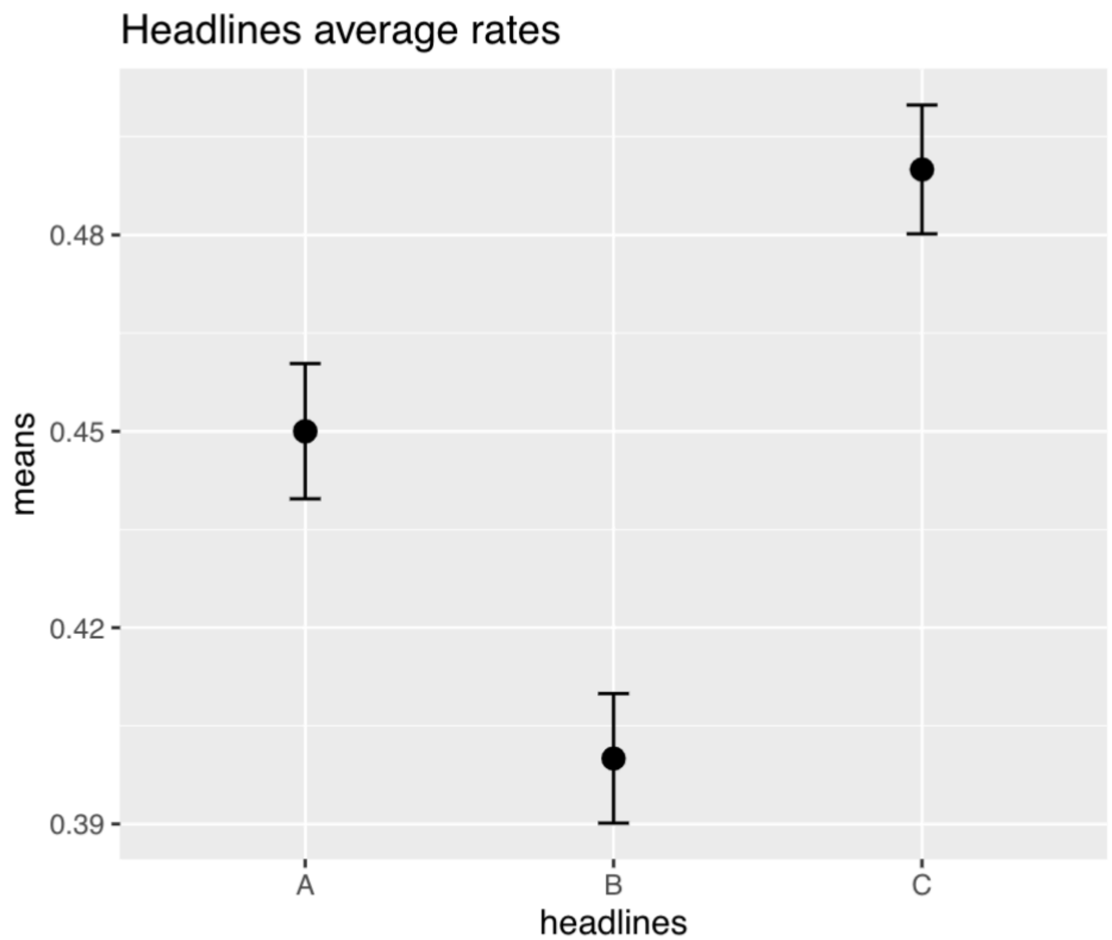
larger number of visits

Comparing *k* populations



With larger samples, CIs are smaller, and differences are more clear

Comparing *k* populations



Now it is clear that headline C is to be preferred

Comparing 2 populations

What to do if Ho is not rejected?

- Increase *n*
- Sequential approach

Comparing 2 populations

The

Computer Science

Approach

Bandit Algorithm for Website selection

Bandit = slot machine

-used for gambling-

Bandit Algorithm for Website selection

Bandit = slot machine

-used for gambling-

(designed to take the

money from gamblers)

Bandit Algorithm for Website selection

Bandit = slot machine

-used for gambling-

(designed to take the

money from gamblers)

Also called

one-armed bandits

Bandit Algorithm for Website selection

Imagine you want to gamble with 2 slot machines, each with different pay rates

Bandit Algorithm for Website selection

Imagine you want to gamble with 2 slot machines, each with different pay rates

- Try each several times to estimate the pay rate (exploration phase)

Bandit Algorithm for Website selection

Imagine you want to gamble with 2 slot machines, each with different pay rates

- Try each several times to estimate the pay rate (exploration phase)
- Select the best one to max profit (exploitation phase)

Armed-Bandit problem

- The problem involves an exploration / exploitation tradeoff
- How much money to spend exploring and how much is left for profiting

Bandit Algorithm for Website selection

Imagine you want to try two *designs*, each with different (but unknown) *user rates*

- Try each several times to estimate the user rate (exploration phase)
- Select the best one to max profit (exploitation phase)

Multi-armed bandit problem

Imagine you want to gamble with k slot machines, each with different pay rates

A room with k slot machines is equivalent to a single slot machine with k arms, each paying different rates

Multi-armed bandit problem

Imagine you want to gamble with k slot machines, each with different pay rates

A room with k slot machines is equivalent to a single slot machine with k arms, each paying different rates

When to select and settle with the one that you think is the *best*?

Multi-armed bandit problem

Objective

Find out the machine

that pays the best rate

and stay at that machine

Multi-armed bandit problem for Website selection

Objective

Find out the design that pays the best rate

Multi-armed bandit problem

The

epsilon – Greedy

algorithm

Multi-armed bandit problem

- greedy algorithm

  Always chooses the best option found after *m* attempts

  (keeps exploiting the best available option)

Multi-armed bandit problem

- greedy algorithm

  Always chooses the best option found after *m* attempts

  (keeps exploiting the best available option)

- almost greedy

  Almost always chooses the best option found after *m* attempts (sometimes it chooses to explore other options) allowing to update the *best* option

Multi-armed bandit problem

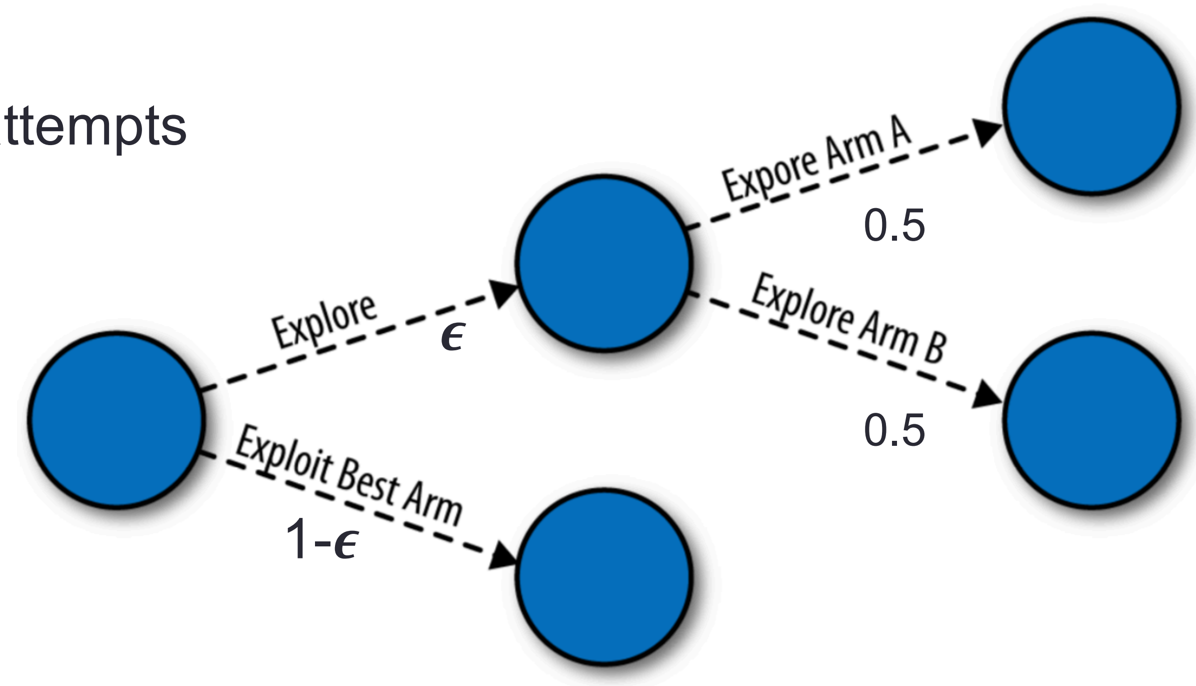
- epsilon - greedy algorithm

An almost greedy algorithm that every once in a while does not choose the best available option and prefers to explore other options

- epsilon (ϵ): probability that the algorithm explores new options and not the best available
- (ϵ = 0, for a greedy algorithm)

Multi-armed bandit problem

after *m* attempts



The epsilon-Greedy Algorithm

Epsilon-greedy algorithm (for A/B testing)

- Epsilon is fixed number $0 < \epsilon < 1$
- Generate a random value *x* between 0 and 1
- If $x < \epsilon$ show the next visitor one of the web designs randomly

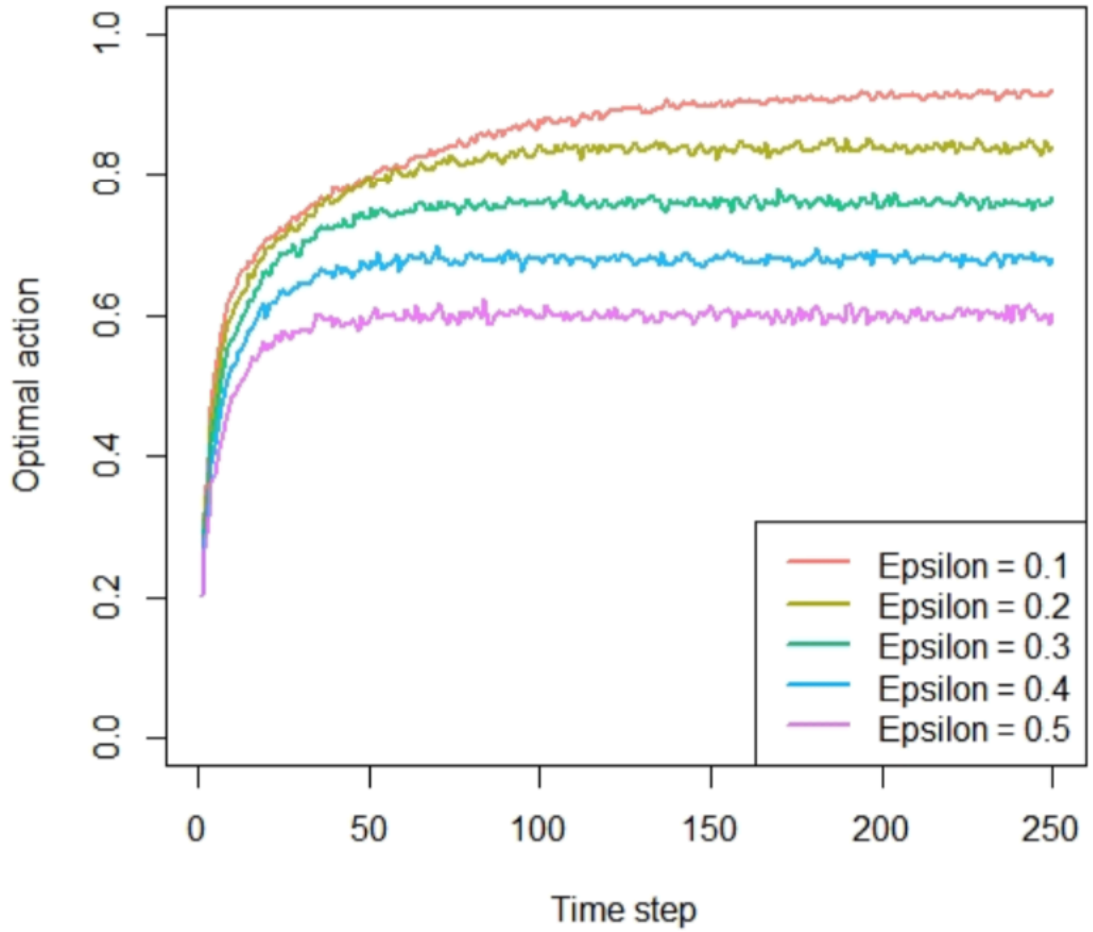
Otherwise, show web design with highest rate of purchases

Example

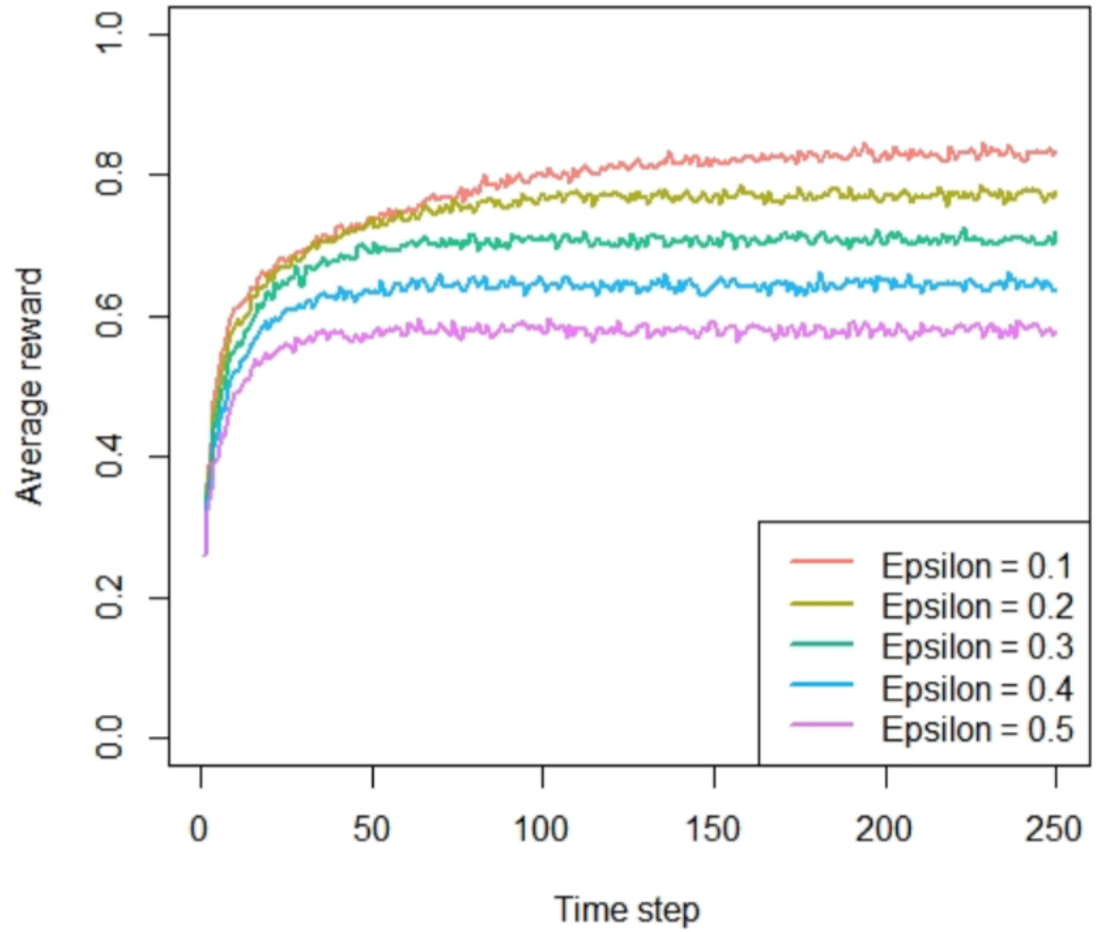
- Five designs (options)
  - 4 give reward 10% of the time, and
  - 1 give reward 90% of the time (this is best option)
- reward is $1
- Try policies with epsilon = 0.1,0.2,…,0.5

- Simulate N = 500 times each policy, and 250 visits, to find
  a) fraction of times the algorithm chooses best option
  b) average reward after each visit (game)
  c) cumulative reward after each visit (game)

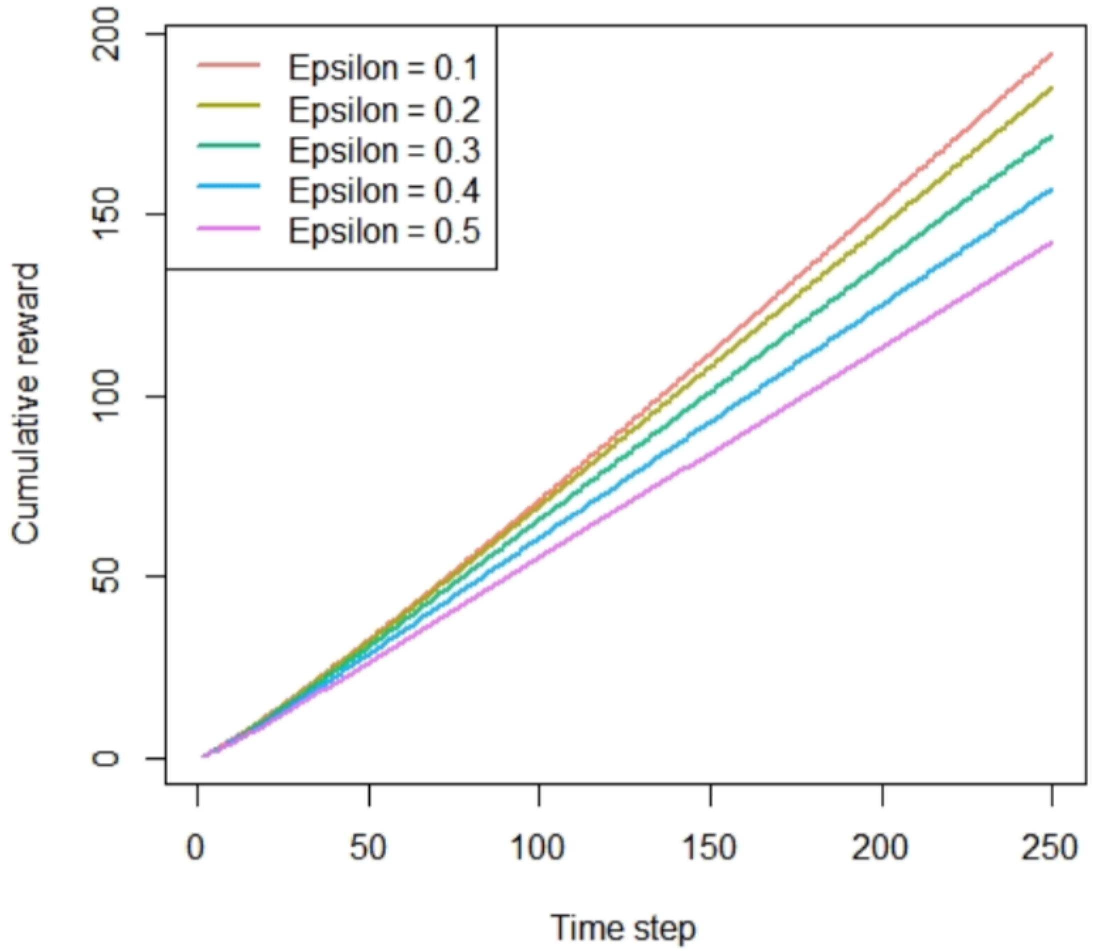
How often does the algorithm select the best?



How much reward does the it earn on average?



How much cumulative reward does the it earn on average?



Other bandit algorithms

- Softmax
- Upper Confidence Bound

Reference

Bandit Algorithms for Website Optimization, J. White