phrasee

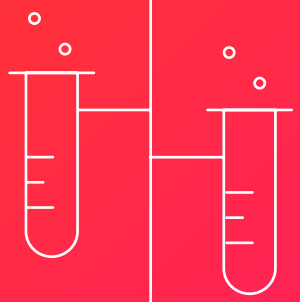
Artificial intelligence.
Human language.
Awesome.

The ultimate guide to subject line multivariate testing


Hello. We're Phrasee.
Phrasee empowers brands with AI-powered copywriting.

Multivariate testing (MVT) is a form of controlled experiment that many of the world's top marketing teams apply and is a direct contributor to their success.

MVT provides marketers with data-driven insight on what makes an audience tick. In the case of subject lines, multivariate testing can reveal the linguistic variables that incentivize your audience to spend more with you.



MVT is important, but many marketers don't test with much scientific or statistical rigor. Most email marketing experts will tell you that you should be multivariate testing, but what they don't tell you is how.

In this handy guide, you'll learn:

1. Why regularly multivariate testing your email subject lines is of such crucial importance.
2. How multivariate testing can be implemented into your brand's email marketing strategy quickly and easily.
3. Which factors you'll need to consider to ensure that you run the most effective email subject line multivariate tests, with the help of Phrasee's awesome end-to-end AI marketing platform.

Let's get started...

Why test email subject lines?

There are several factors that impact the success of an email marketing campaign and, while they are all relevant, a campaign's email subject line is one of the most important.

Here's why:

- Over a third (35%) of your recipients will decide whether or not to open your email based on the subject line alone. Research indicates that a good email subject line can generate up to 20% more opens than a bad one.
  [Source: Convince&Convert]

- The subject line will often be the only element of a campaign that many of your subscribers see. Our research shows that the average open rate across all industries is only 20.81% and, as we know, you need more opens to get more clicks, and more clicks to get more revenue… isn't that what this is really all about?
  [Source: Mailchimp]

Choosing a "good" email subject line based purely on gut instinct invites human bias into the equation. Multivariate testing ensures a good subject line

based purely on data-driven results.

How to run a multivariate test

Everything you need to know about the how and the why.

Most people will conduct a basic two-version split test, i.e. A/B testing. But you can test multiple variables with multivariate testing. If you test 10 different subject line variables, for example, you are testing hundreds of linguistic combinations and 45 pairs (A/B, A/B/C, A/B/C/D...etc.).

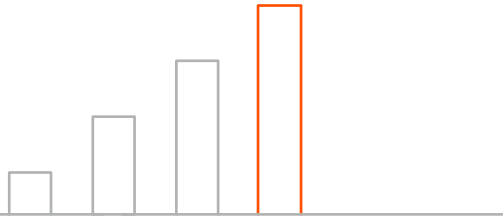
In theory, there is no limit to the number of splits you can do. If you had a limitless audience size, you would want to be testing billions of different variables. You would use this testing to isolate a local maximum on a campaign-to-campaign basis. But this is not realistic because you do not have a limitless list size.

The goal is to test as many variables as is technically feasible and as is statistically significant given your database size, while minimizing the opportunity cost.

Recommendation: Don't limit yourself to A/B testing.

Test many more combinations from which you can learn.

How to run a multivariate test



Create a test group.

With experimental methodology, you isolate part of your audience (the test group) to test a number of variables, such as subject lines, body copy, or whatever it is that you want to test and optimize. You separate part of the test group from the experiment, which will not be tested. This is called a 'control group' and can be used to rule out alternative explanations of the experimental results. In this way, you discover what variation works best and you use that for the remaining audience in your database. Each group must be statistically significant.
(➡️ More on statistical significance on p. 12)

Construct your test to ensure that your experimental methodology is sound and you're not just chasing random variance.

Set your goal.

What do you want your subject lines to achieve? In other words, is the goal to increase the number of conversions per person? Increase click-through rate? Increase open rate?

Create your campaign.

You define your independent variable (your subject line content) and change this variable to see what the effect is on the dependent variable (your email open rate). Define your desired outcome i.e. the data you want to be left with at the end of your experiment. In this case, it's about identifying the best subject line for your email marketing campaign.

Recommendation: Focus on open rate as a success metric. Numerous sources of research show that opens are highly correlated with clicks i.e. when you get more opens, you get more clicks. The subject line with the highest open rate can be deployed out to the rest of your audience (➡️ more on optimizing for opens on p. 9).

How to run a multivariate test

Determine your test sample size.

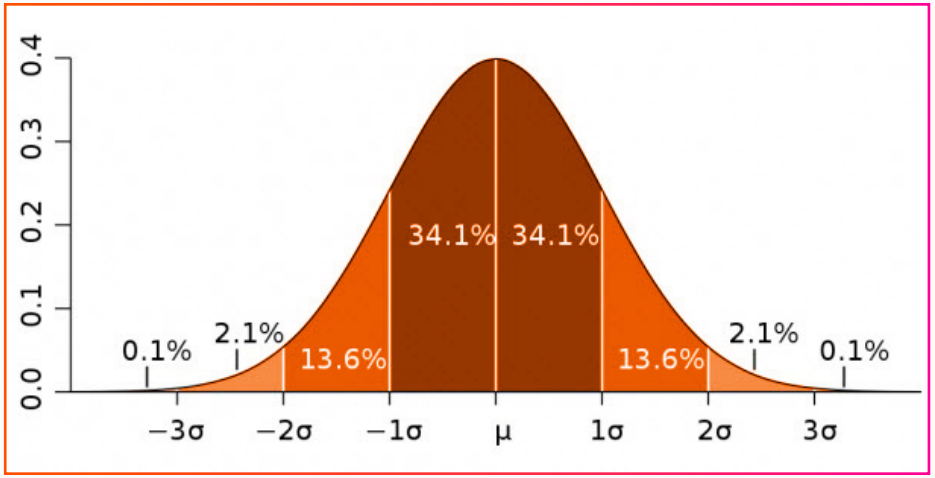
A lot of math goes into this. Email responses follow what's called a "binomial distribution". In other words, each result is binary – opened or not opened, clicked or not clicked, purchased or not purchased.

In the long run, a binomial distribution converges upon a "Gaussian distribution" i.e. "normal" or "bell curve".

Every subject line sits on a bell curve and you want to find the one that sits furthest right on the bell curve i.e. the largest number of standard deviations from the mean.

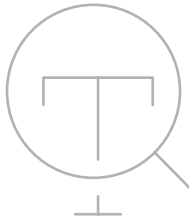
To work out the sample size, you need to determine the appropriate number of splits that you will use. There is a trade-off here. Lots of splits increase the possibility of finding a good subject line. In order to use a lot of splits, you need to use a larger proportion of your list for testing, which increases the opportunity cost of a poorly performing subject line.

Gaussian distribution bell curve



You also need to have the right size of splits. If the splits aren't big enough, the results won't be significant and will be due to chance. When your sample sizes are too small, you run the risk of making decisions based on random variance. As a general rule, your split test sizes should be in the tens of thousands (➡️ more on sample size on p. 11).

How to run a multivariate test



What's the point of testing if you don't apply those learnings?

Run your multivariate test.

Always work backwards from your optimum send time. Don't determine a winner too early. Wait at least four hours before sending out the winner i.e. the subject line with the highest open rate. This is because with broadcast email campaigns, you have a one-shot opportunity to engage with your audience. By allowing a sufficient amount of time for the test to run its course, you ensure that you have enough data to accurately determine the winner.

Learn from your results.

Sorting through your test data is important. Keep track of your results over time to find out why some things work and why some things don't work.

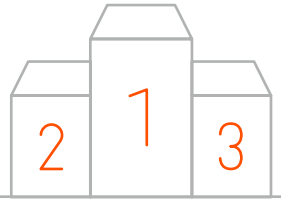
If you apply testing rigor, you will get better results. Building a model to learn from your results will allow you to run post-hoc analysis on which subject lines performed well, and which ones didn't. This can be challenging because in language there are multiple variants that can and cannot affect subject line performance. You can run into problems if you work with limited data, domain-specific data, or only data within your world. Your model is only going to be as good as the data in which you see it.

Thanks to Phrasee, you can make more data-informed decisions with your marketing. Phrasee's software looks at hundreds of thousands of different linguistic features, making controlled experiments easier and more insightful than ever.

For more insights on experimental methodology and how to run effective multivariate tests, watch Phrasee's free on-demand webinar:

➡️ phrasee.co/science_of_split_testing

Open rate reigns supreme



➡️ Find out what Phrasee's data scientists have to say about the performance metric that matters most in subject line split testing by reading this article and exploring our interactive charts.

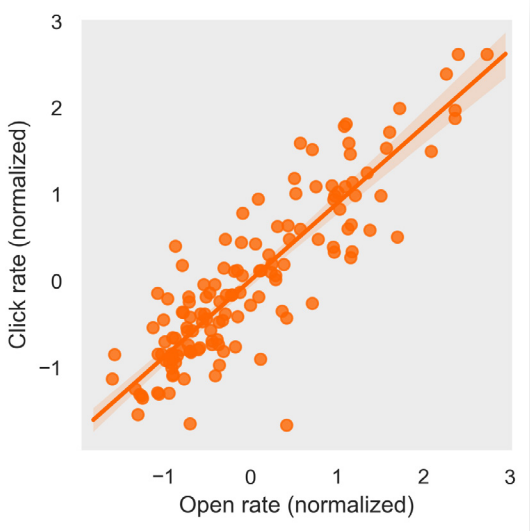
Always choose the subject line that performed best at generating opens. Open rate is the most robust email marketing performance metric.

Here's why...

Open rate is also correlated with other metrics (opens ➔ clicks ➔ orders). You may be tempted to choose the winning subject line based on secondary metrics such as click-throughs, conversions, or revenue. However, as you move further along the user journey, your sample sizes get smaller and your performance data becomes less reliable. For example, a single (random) large order can skew your email revenue results as the data pool is proportionally smaller. In comparison, the open rate is tested on the most data, making it the most dependable metric.

The open-click relationship is key. When a subject line is on-brand without being misleading, a high open rate will translate to more clicks.

Open / click correlation



1. Test correctly
2. Sit back
3. **Watch your open rates skyrocket**

HOW MANY SUBSCRIBERS SHOULD BE IN EACH TEST GROUP?

Does size matter?

Wonder no longer, dear reader

We asked our own Chief Scientist, Neil Yager PhD, how many subscribers should be in each segment and he replied with this ➔

Unfortunately, there is no single number that applies to every brand. Some brands have tens of millions of subscribers, while others may only have thousands. After multivariate testing millions of subject lines over the years, we at Phrasee have found that the larger the sample size, the more reliable the data.

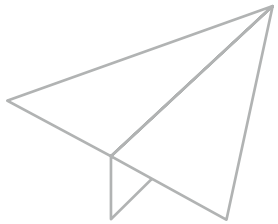
"Let's say someone gives you a coin and says: 'this coin comes up heads 60% of the time when flipped'. How many times would you need to flip that coin to determine if the person who handed it to you was telling the truth? 10 times? That's probably not enough. 1,000,000 times? That seems a bit excessive.

'Statistical significance' basically means that we can trust the results: they are as unlikely as possible to be due to chance."

- Neil Yager, Phd

Statistical significance

...and why it matters to you



Statistical significance is the likelihood that a relationship between two or more variables is caused by something other than chance.

At an individual level, the question of whether or not someone opens an email is almost entirely random.

Maybe they are having a busy week.

Maybe your email landed in their inbox right below an email they had been expecting.

Maybe they are really hungry and the microwave just pinged.

Many factors can come into play, none of which you, or even the most cutting-edge, world-leading AI technology, have any control over whatsoever.

That's why it's so important to make sure that when you run multivariate tests, you use a statistically significant sample size.

Subject line multivariate testing with a large enough group of recipients will get you results that are reliable enough to draw conclusions from. Test with small groups at your peril! All you'll get are a few random, largely meaningless data points. And what's the point of that?

Always make sure that your sample size is **statistically significant!**

Check out Phrasee's split test calculator to find out how many splits you should use and how big your sample size should be:

➡️ phrasee.co/resource/splits

Your own step-by-step checklist to subject line multivariate testing

1. Choose a campaign

   Set your goal (increase email open rate) and determine your independent variable (subject line content) and dependent variable (email open rate).

2. Set aside a random segment of your audience. This will be your "test" group.

3. Determine the number of splits you will do and the sample size for each split.

   Each segment must be a big enough size to ensure statistical significance. Try using ➡️ *Phrasee's Split Test Calculator tool to work out the number of splits and sample size for each one based on your database size and average open rate.*

4. Divide your test group into random equal segments.

5. Generate unique subject lines for each of the segments in your test group.

Your own step-by-step checklist to subject line multivariate testing

6 Send out emails with identifical body copy to each each of the segments in your test group, each with one of the subject lines you've generated.

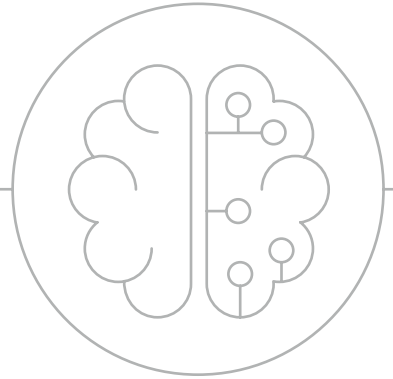
7 Wait a minimum of four hours.
Note: As a general rule of thumb, this is the minimum time required for the test to run its course.

8 The winning subject line is the subject line that has the highest open rate - send it out to the rest of your audience (the 'final send').

9 Keep track of your multivariate testing results over time.
Insight into the macro trends within your test data can help your subject line copy perform better over time.

10 Keep testing!
Testing your subject lines rigorously on an ongoing basis is the only way to ensure that they stay optimized for your audience over time.

Important factoids



Looking at any multivariate test results in a vacuum will only get you so far. The real goal here is to learn over time. We all have theories about what does and doesn't work in an email subject line, but it's important to be data-driven and test a wide range of language. The opportunity cost is small, and the result just might surprise you!

Gain better insight into where your language optimization opportunities lie with Phrasee's subject line health check tool:
➡️ phrasee.co/language-calculator/

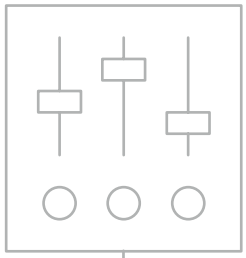
Isolating the true impact of the subject line language you're testing means holding all other variables (send times, demographics, email body content, discounts offered, etc.) constant. These variables can (and will) impact performance and skew your data if you don't.

Correlation does not mean causation. If subject lines containing the word "sparkle" generally produce a high open rate, this means that there's a correlation between using the word "sparkle" and high open rates. It doesn't mean that the relationship between open rates and the word "sparkle" is causal. Does the word "sparkle" cause people to open those emails? Probably not. A more likely explanation is that the word is used to describe certain products (e.g. diamonds) and that the reference to the product is driving opens.

Phrasee takes away gut instinct and generates awesome results

based purely on data-driven insights!

How could it possibly go wrong?



We have described a recipe for successful email subject line multivariate testing. But no process is perfect (at least with humans at the helm!). Here are a few potential problems to keep an eye out for:

Statistically insignificant test groups

As mentioned previously, if you don't run your email subject line tests on a large enough sample size, your performance results will be vulnerable to random variance. We might talk about this a lot, but it really does matter!

What might have happened:

Your sample size was too small!
Try using Phrasee's split test tool to calculate the number of splits you should use, and the sample size for each one:
➡️ phrasee.co/resource/splits

Ending your tests too early

Start your tests with plenty of time to spare, and don't declare a winning subject line until a significant amount of time has passed (we recommend at least four hours). This will allow the inbox-checking idiosyncrasies of your brand's audience to work themselves out and provide you with more robust results.

What might have happened:

You're an eager beaver, keen for those email marketing returns... but trust us. We've done the research, and you really do need to wait at least four hours.

Outliers (more like out-liars!)

The results of any email subject line multivariate tests should all be within a similar range. If you see results that look like this: 12%, 13%, 12%, 42%, 12%, it doesn't mean that you've just found the most effective email subject line in the history of the world, it means that something has gone wrong.

What might have happened:

Technical errors, bugs and bots can cause havoc with your results. No need to lose sleep - just ignore the outliers when analyzing your data.

The AI marketing difference

Subject line multivariate testing is an essential tool to getting more data-driven insight that can be used to positively impact your results. The caveat to this is: in order to deliver results that make a difference to your conversion rates, you need to invest your time into setup and measurement.

What if we told you we have a solution to this problem?

Take any group of email subject lines, follow the process outlined in this guide, and you'll ensure that your campaigns always use the best subject line of the bunch you started out with. But if you REALLY want to see an uptick in email marketing performance, what you need to do is improve the subject lines you start out with.

This is where true AI marketing really shines. Phrasee's predictive deep learning model tracks the results of every send, learning what works on your audience (and what doesn't) over time, and the natural language generation engine generates higher performing email subject line language based on the results.

Wish your brand had results like these?

Well, this one belongs to Domino's Pizza:

↑ 26%

email open uplift

Phrasee helped Gumtree achieve this:

↑ 35%

email open uplift

And this impressive stat belongs to Jacamo :

↑ 21.3%

email click uplift

...and your results can be awesome too!

Discover how Phrasee is empowering customers with AI-powered copywriting:
➡️ phrasee.co/case-studies

Want to find out how Phrasee can help your marketing?

Book a discovery session

phrasee

Artificial intelligence.
Human language.
Awesome.

+44 (0)20 8870 6968
awesome@phrasee.co
www.twitter.com/phrasee
www.linkedin.com/company/phrasee
www.phrasee.co

Phrasee UK
Tintagel House
92 Albert Embankment
London
SE1 7TY

Phrasee USA
100 Montgomery Street
Suite 1202
San Francisco
CA 94104