

Storytelling with Data

Module 3: Effective business writing with audience analysis

Scott Spencer

Faculty and Lecturer
Columbia University

Agenda

Upcoming deliverable

Objectives of business writing

Readings

Example 250-word memo

Group work

Upcoming deliverable

Upcoming deliverable

For your chosen company and case study,
as an imagined member of the analytics team ...

250-word memo

Write a memo to **CAO** about an opportunity to leverage analytics. Consider background context, problem, data, solution, and impact.

750-word proposal

On approval of the memo, write a proposal to **CAO**, detailing the anticipated project.

Storyboard

Present project result in storyboard to **CMO**, using narrative forms, and with comparisons, metaphors and other storytelling concepts.

Infographic

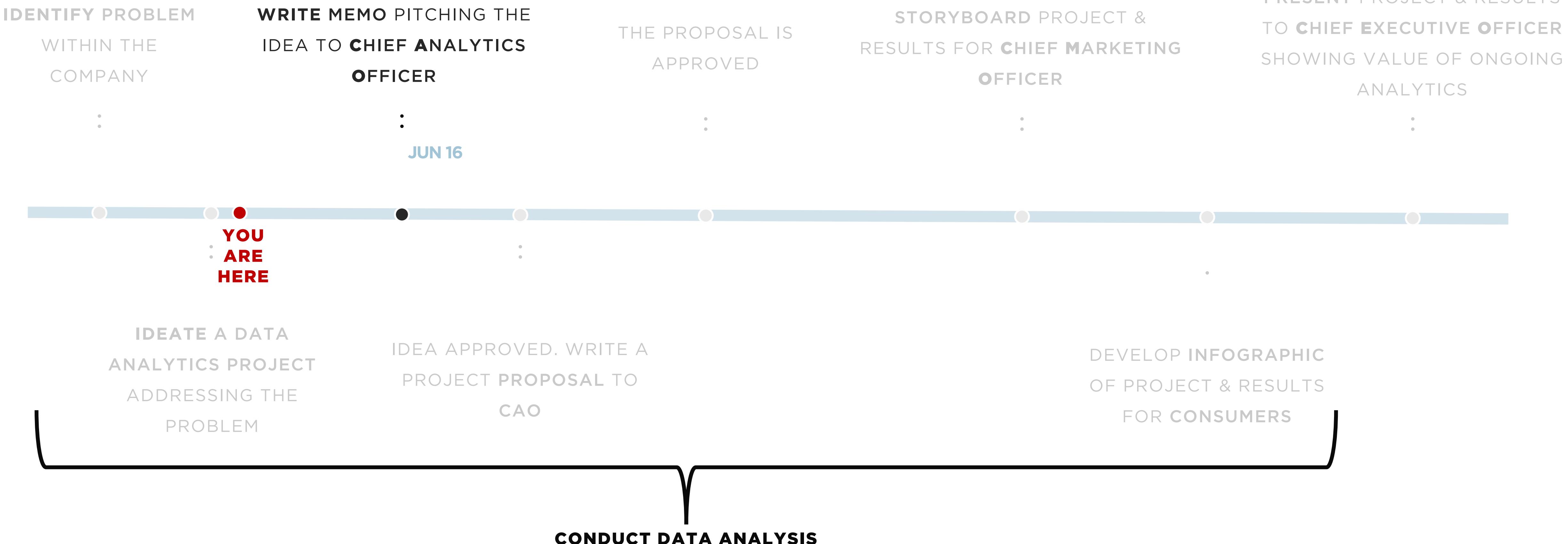
Recraft the results, telling the narrative through an infographic for the **public** or **potential consumers**, using data visualization with brand awareness.

Presentation

Construct and deliver a 4-5 minute persuasive presentation with up to 10 slides to the **CEO**, telling the story of the analytics project to convince them of further investment in analytics.

Upcoming deliverable

250-word memo — Write a memo to **CAO** about an opportunity to leverage analytics. Consider background context, problem, data, solution, and impact. Be sure you have data and a plan for analysis before pitching the idea in the memo.



Objectives of business writing

For **all** business
communications

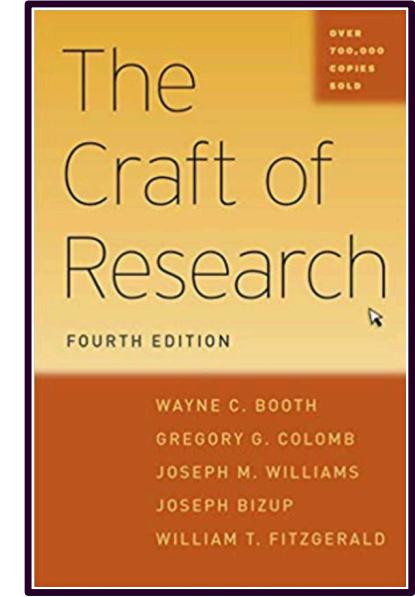
Step into your audiences' shoes to
get them to
pay attention to,
understand,
(be able to) act upon
a maximum of messages, given
constraints.

Who is your audience?

**What's your purpose
for communicating?**

tl;dr

Readings



Revising style: telling your story clearly

Booth and co-authors

The authors

All are university professors of English, and their book is first among Amazon's ranking of books in methodology and statistics.

A few writing principles

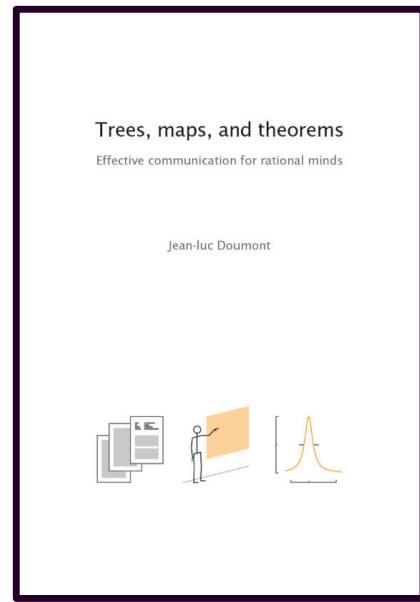
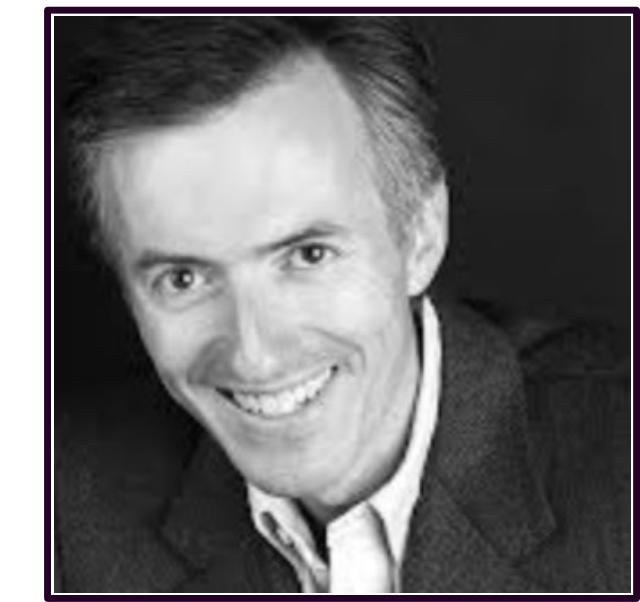
Express crucial actions in verbs.

Make your central characters the subjects of those verbs; keep those subjects short, concrete, and specific.

Old before new — “readers follow a story most easily if they can begin each sentence with a character or idea that is familiar to them, either because it was already mentioned or because it comes from the context.”

Complexity last, particularly important when:

introducing a new technical term,
presenting a long or complex unit of
information, introducing a concept to be
developed in what follows.



Trees, maps, theorems

Doumont

Author, audience, and purpose

An engineer from the Louvain School of Engineering and PhD in applied physics from Stanford University, Jean-luc Doumont wrote this book to help engineers, scientists, and managers with business communication.

Wear shoes of your audience

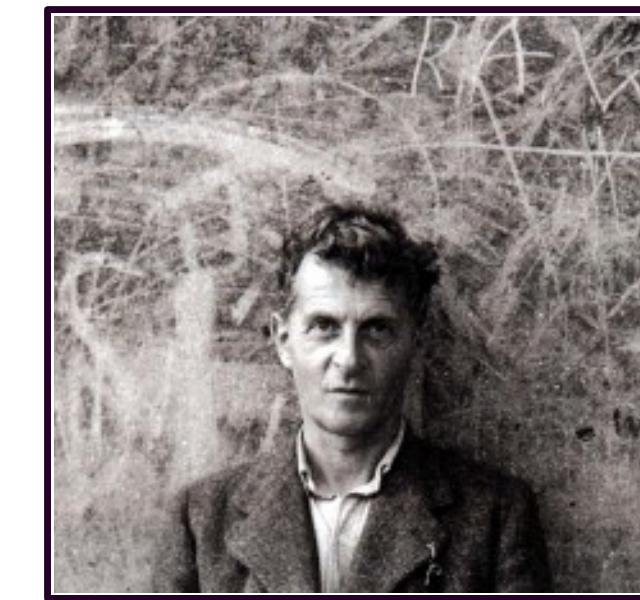
Put yourselves in the shoes of the audience, anticipating their situation, their needs, their expectations. Structure the story along their line of reasoning, recognizing the constraints they might bring: their familiarity with the topic, their mastery of the language, the time they can free for us.

Messages, not just information

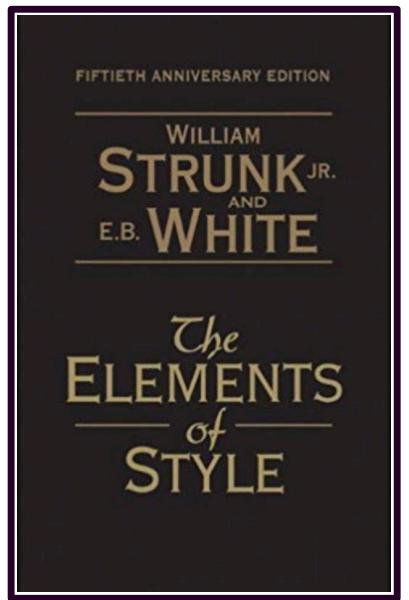
A message differs from raw information in that it interprets the information for a specific audience and for a specific purpose.

Organization, messages first

Get the audience interested. Once you have their attention, tell your main message. Last, support this message, tell how you got there.



&



The elements of style

Strunk & White

The authors

William Strunk Jr. was an English professor at Cornell University; E. B. White was his student. White wrote for *The New Yorker* — perhaps the best edited magazine — for sixty years, and won a Pulitzer for his writing.

Why read S&W, by Richard Ford

Mimic examples,
be concise,
don't overstate

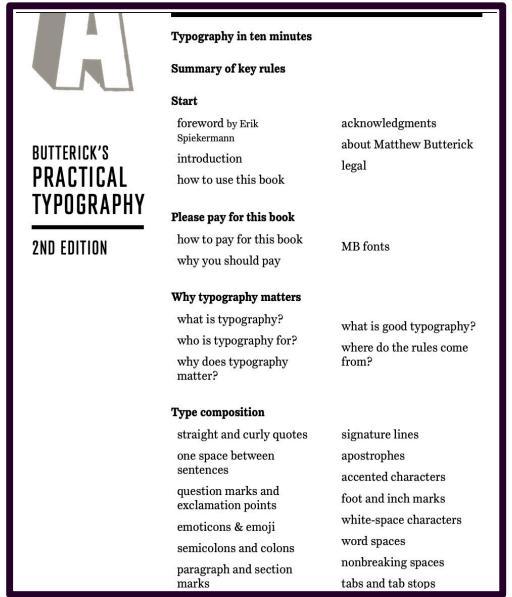
S&W doesn't really teach you how to write, it just tantalizingly reminds you that there's an orderly way to go about it, that clarity's ever your ideal, but — really — it's all going to be up to you.

Leading by example, this tiny book provides dos and don'ts with examples of each. Re-read.

Heed their warnings:

Vigorous writing is concise. A sentence should contain no unnecessary words, a paragraph no unnecessary sentences, for the same reason that a drawing should have no unnecessary lines and a machine no unnecessary parts. This requires not that the writer make all his sentences short, or avoid all detail and treat subjects only in outline, but that every word tell.

A single overstatement, wherever or however it occurs, diminishes the whole, and a carefree superlative has the power to destroy, for readers, the object of your enthusiasm.



Butterick's practical typography

Matthew Butterick

The author

The author earned a visual-studies degree from Harvard, and a law degree from UCLA. He is a writer, typographer, programmer, and lawyer.

Conserve limited reader attention

Typography is the visual component of the written word.

"Typography is for the benefit of the reader."

"Most readers are looking for reasons to stop reading. . . . Readers have other demands on their time. . . . The goal of most professional writing is persuasion, and attention is a prerequisite for persuasion. Good typography can help your reader devote less attention to the mechanics of reading and more attention to your message."

Very basic guidelines, body text

Point size: 10-12 (print), 15-25 pixels (web)
Line spacing: 120-145% of the point size
Line length: 45-90 characters per line
Fonts: see his recommendations

Grids are helpful when they encourage consistency. They make it easier to relate elements on the page to existing ones.

Example 250-word memo



A perfect game

Perfect Game

Los Angeles Dodgers

Pitcher Sandy Koufax

Statcast data

(Attempting to) steal a base

Salary cap

Baseball

Statistics, probability, computing

Models

Mode, maximum likelihood

Probability distributions

Joint distributions

R language and packages

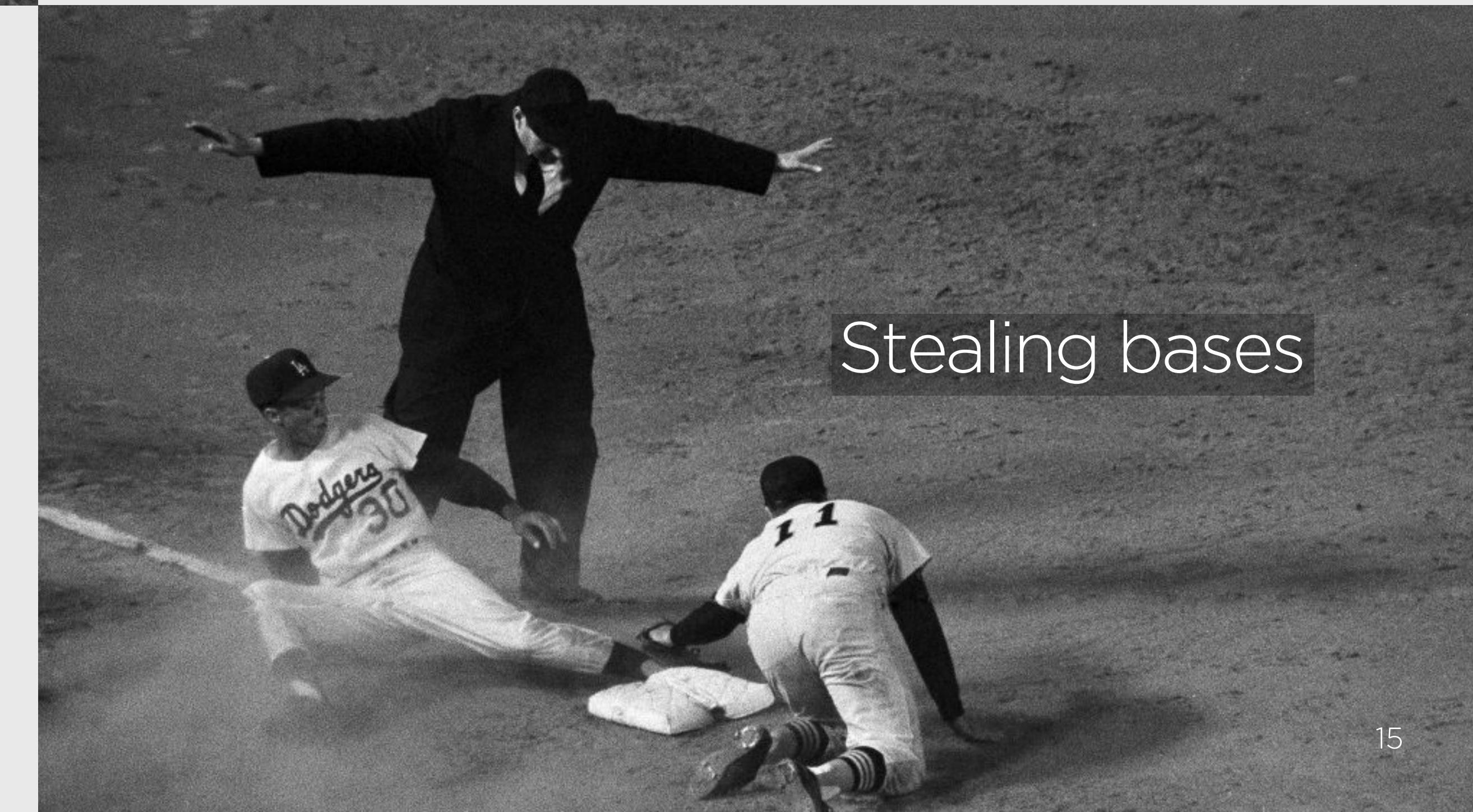
Inferences

Mean, expectations

Decision theory

Counterfactuals

Simulations



Stealing bases



To: **Scott Powers**
Senior Analyst, Baseball Analytics

2 February 2019

Our game decisions should optimize expectations. Proposal for testing: model decisions to steal

The most likely sequence of events on defense is a *perfect game* — occurring just 23 times in major-league baseball, once by our own Sandy Koufax. Decisions from what is most likely, however, leave wins unclaimed. To claim them, let's base decisions on expectations flowing from decision theory and probability models. A joint model of all events works best, but we can start small, with say, decisions to steal second base.

After defining our objective (e.g. optimize expected runs) we will from Statcast data compute expectations: weight everything that could happen by its probability and accumulate these probability distributions. Joint distributions of all events, an eventual goal, will allow us to ask counterfactuals — “what if I do *this*” or “what if my opponent does *that*” — and simulate games to learn how decisions change win probability. It enables optimal strategy.

Rational and optimal, this approach is more efficient for gaining wins. For perspective, each added win from the free-agent market costs 10 million, give or take, and the league salary cap prevents unlimited spend on talent. There is no cap, however, on investing in rational decision processes.

Computational issues are being addressed in Stan, a tool that enables inferences through advanced simulations. This open-source software is free but teaching its applications will require time. To shorten our learning curve, we can start with Stan interfaces that use familiar syntax (like lme4) but return joint probability distributions: R packages *rethinking*, *brms*, or *rstanarm*. And we can test the concept with decisions to steal.

Sincerely,
Scott Spencer

Questions for Discussion

Was a few minutes enough time to understand?
Did you notice **structure** that guides the audience?

How do the audiences differ between this memo,
the *Jakarta* proposal, and *The Next Rembrandt*?

How do the structures and details of the
communications seem to depend on the audience?

Might the audience reply, tl;dr?



To: **Scott Powers**
Director, Quantitative Analytics

Audience background

Director of Quantitative Analytics
Ph.D. Statistics from Stanford University
Some publications use machine learning
Knows R programming
An employee, knows history of Dodgers

2 February 2019

Message first, context

Our game decisions should optimize expectations. Proposal for testing: model decisions to steal

Main body word count, 250

The most likely sequence of events on defense is a *perfect game* — occurring just 23 times in major-league baseball, once by our own Sandy Koufax. Decisions from what is most likely, however, leave wins unclaimed. To claim them, let's base decisions on expectations flowing from decision theory and probability models. A joint strategy for all events works best, but we can start small, with say, decisions to steal second base.

Readability Statistics	
Counts	After defining our objective (e.g. optimize expected runs) we will from Statcast data compute expectations:
Words	weight every thing that could happen by its probability and accumulate these probability distributions. Joint
Characters	distributions of all events, an eventual goal, will allow us to ask counterfactuals — “what if I do <i>this</i> ” or “what if
Paragraphs	my opponent <i>does that</i> ” — and simulate games to learn how decisions change win probability. It enables optimal
Sentences	strategy.
Averages	
Sentences per Paragraph	Rational and optimal, this approach is more efficient for gaining wins. For perspective, each added win from the
Words per Sentence	free-agent market costs 10 million, give or take, and the league salary cap prevents unlimited spend on talent.
Characters per Word	There is no cap, however, on investing in rational decision processes.
Readability	
Flesch Reading Ease	Computational issues are being addressed in Stan, a tool that enables inferences through advanced simulations.
Flesch-Kincaid Grade Level	This open-source software is free but teaching its applications will require time. To shorten our learning curve,
Passive Sentences	we can start with Stan interfaces that use familiar syntax (like lme4) but return joint probability distributions: R packages rethinking, brms, or rstanarm. And we can test the concept with decisions to steal.

Sincerely,
Scott Spencer



To: **Scott Powers**
Senior Analyst, Baseball Analytics

2 February 2019

Our game decisions should optimize expectations. Proposal for testing: model decisions to steal

Context, orient the audience

The most likely sequence of events on defense is a *perfect game* — occurring just 23 times in major-league baseball, once by our own Sandy Koufax. Decisions from what is most likely, however, leave wins unclaimed. To claim them, let's base decisions on expectations flowing from decision theory and probability models. A joint model of all events works best, but we can start small, with say, decisions to steal second base.

After defining our objective (e.g. optimize expected runs) we will from Statcast data compute expectations: weight everything that could happen by its probability and accumulate these probability distributions. Joint distributions of all events, an eventual goal, will allow us to ask counterfactuals — “what if I do *this*” or “what if my opponent does *that*” — and simulate games to learn how decisions change win probability. It enables optimal strategy.

Rational and optimal, this approach is more efficient for gaining wins. For perspective, each added win from the free-agent market costs 10 million, give or take, and the league salary cap prevents unlimited spend on talent. There is no cap, however, on investing in rational decision processes.

Computational issues are being addressed in Stan, a tool that enables inferences through advanced simulations. This open-source software is free but teaching its applications will require time. To shorten our learning curve, we can start with Stan interfaces that use familiar syntax (like lme4) but return joint probability distributions: R packages rethinking, brms, or rstanarm. And we can test the concept with decisions to steal.

Sincerely,
Scott Spencer



To: **Scott Powers**
Senior Analyst, Baseball Analytics

2 February 2019

Our game decisions should optimize expectations. Proposal for testing: model decisions to steal

The most likely sequence of events on defense is a *perfect game* — occurring just 23 times in major-league baseball, once by our own Sandy Koufax. Decisions from what is most likely, however, leave wins unclaimed. To claim them, let's base decisions on expectations flowing from decision theory and probability models. A joint model of all events works best, but we can start small, with say, decisions to steal second base.

Goal, action problem

After defining our objective (e.g. optimize expected runs) we will from Statcast data compute expectations: weight everything that could happen by its probability and accumulate these probability distributions. Joint distributions of all events, an eventual goal, will allow us to ask counterfactuals — “what if I do *this*” or “what if my opponent does *that*” — and simulate games to learn how decisions change win probability. It enables optimal strategy.

Rational and optimal, this approach is more efficient for gaining wins. For perspective, each added win from the free-agent market costs 10 million, give or take, and the league salary cap prevents unlimited spend on talent. There is no cap, however, on investing in rational decision processes.

Computational issues are being addressed in Stan, a tool that enables inferences through advanced simulations. This open-source software is free but teaching its applications will require time. To shorten our learning curve, we can start with Stan interfaces that use familiar syntax (like lme4) but return joint probability distributions: R packages rethinking, brms, or rstanarm. And we can test the concept with decisions to steal.

Sincerely,
Scott Spencer



To: **Scott Powers**
Senior Analyst, Baseball Analytics

2 February 2019

Our game decisions should optimize expectations. Proposal for testing: model decisions to steal

Proposed solution

The most likely sequence of events on defense is a *perfect game* — occurring just 23 times in major-league baseball, once by our own Sandy Koufax. Decisions from what is most likely, however, leave wins unclaimed. To claim them, let's base decisions on expectations flowing from decision theory and probability models. A joint model of all events works best, but we can start small, with say, decisions to steal second base.

After defining our objective (e.g. optimize expected runs) we will from Statcast data compute expectations: weight everything that could happen by its probability and accumulate these probability distributions. Joint distributions of all events, an eventual goal, will allow us to ask counterfactuals — “what if I do *this*” or “what if my opponent does *that*” — and simulate games to learn how decisions change win probability. It enables optimal strategy.

Rational and optimal, this approach is more efficient for gaining wins. For perspective, each added win from the free-agent market costs 10 million, give or take, and the league salary cap prevents unlimited spend on talent. There is no cap, however, on investing in rational decision processes.

Computational issues are being addressed in Stan, a tool that enables inferences through advanced simulations. This open-source software is free but teaching its applications will require time. To shorten our learning curve, we can start with Stan interfaces that use familiar syntax (like lme4) but return joint probability distributions: R packages rethinking, brms, or rstanarm. And we can test the concept with decisions to steal.

Sincerely,
Scott Spencer



To: **Scott Powers**
Senior Analyst, Baseball Analytics

2 February 2019

Our game decisions should optimize expectations. Proposal for testing: model decisions to steal

The most likely sequence of events on defense is a *perfect game* — occurring just 23 times in major-league baseball, once by our own Sandy Koufax. Decisions from what is most likely, however, leave wins unclaimed. To claim them, let's base decisions on expectations flowing from decision theory and probability models. A joint model of all events works best, but we can start small, with say, decisions to steal second base.

After defining our objective (e.g. optimize expected runs) we will from Statcast data compute expectations: weight everything that could happen by its probability and accumulate these probability distributions. Joint distributions of all events, an eventual goal, will allow us to ask counterfactuals — “what if I do *this*” or “what if my opponent does *that*” — and simulate games to learn how decisions change win probability. It enables optimal strategy.

Objective, data, methods

Rational and optimal, this approach is more efficient for gaining wins. For perspective, each added win from the free-agent market costs 10 million, give or take, and the league salary cap prevents unlimited spend on talent. There is no cap, however, on investing in rational decision processes.

Computational issues are being addressed in Stan, a tool that enables inferences through advanced simulations. This open-source software is free but teaching its applications will require time. To shorten our learning curve, we can start with Stan interfaces that use familiar syntax (like lme4) but return joint probability distributions: R packages rethinking, brms, or rstanarm. And we can test the concept with decisions to steal.

Sincerely,
Scott Spencer



To: **Scott Powers**
Senior Analyst, Baseball Analytics

2 February 2019

Our game decisions should optimize expectations. Proposal for testing: model decisions to steal

The most likely sequence of events on defense is a *perfect game* — occurring just 23 times in major-league baseball, once by our own Sandy Koufax. Decisions from what is most likely, however, leave wins unclaimed. To claim them, let's base decisions on expectations flowing from decision theory and probability models. A joint model of all events works best, but we can start small, with say, decisions to steal second base.

After defining our objective (e.g. optimize expected runs) we will from Statcast data compute expectations: weight everything that could happen by its probability and accumulate these probability distributions. Joint distributions of all events, an eventual goal, will allow us to ask counterfactuals — “what if I do *this*” or “what if my opponent does *that*” — and simulate games to learn how decisions change win probability. It enables optimal strategy.

Rational and optimal, this approach is more efficient for gaining wins. For perspective, each added win from the free-agent market costs 10 million, give or take, and the league salary cap prevents unlimited spend on talent. There is no cap, however, on investing in rational decision processes.

Computational issues are being addressed in Stan, a tool that enables inferences through advanced simulations. This open-source software is free but teaching its applications will require time. To shorten our learning curve, we can start with Stan interfaces that use familiar syntax (like lme4) but return joint probability distributions: R packages *rethinking*, *brms*, or *rstanarm*. And we can test the concept with decisions to steal.

**Eventual
benefit**

Sincerely,
Scott Spencer



To: **Scott Powers**
Senior Analyst, Baseball Analytics

2 February 2019

Our game decisions should optimize expectations. Proposal for testing: model decisions to steal

The most likely sequence of events on defense is a *perfect game* — occurring just 23 times in major-league baseball, once by our own Sandy Koufax. Decisions from what is most likely, however, leave wins unclaimed. To claim them, let's base decisions on expectations flowing from decision theory and probability models. A joint model of all events works best, but we can start small, with say, decisions to steal second base.

After defining our objective (e.g. optimize expected runs) we will from Statcast data compute expectations: weight everything that could happen by its probability and accumulate these probability distributions. Joint distributions of all events, an eventual goal, will allow us to ask counterfactuals — “what if I do *this*” or “what if my opponent does *that*” — and simulate games to learn how decisions change win probability. It enables optimal strategy.

**Benefit,
comparison,
financial**

Rational and optimal, this approach is more efficient for gaining wins. For perspective, each added win from the free-agent market costs 10 million, give or take, and the league salary cap prevents unlimited spend on talent. There is no cap, however, on investing in rational decision processes.

Computational issues are being addressed in Stan, a tool that enables inferences through advanced simulations. This open-source software is free but teaching its applications will require time. To shorten our learning curve, we can start with Stan interfaces that use familiar syntax (like lme4) but return joint probability distributions: R packages rethinking, brms, or rstanarm. And we can test the concept with decisions to steal.

Sincerely,
Scott Spencer



To: **Scott Powers**
Senior Analyst, Baseball Analytics

2 February 2019

Our game decisions should optimize expectations. Proposal for testing: model decisions to steal

The most likely sequence of events on defense is a *perfect game* — occurring just 23 times in major-league baseball, once by our own Sandy Koufax. Decisions from what is most likely, however, leave wins unclaimed. To claim them, let's base decisions on expectations flowing from decision theory and probability models. A joint model of all events works best, but we can start small, with say, decisions to steal second base.

After defining our objective (e.g. optimize expected runs) we will from Statcast data compute expectations: weight everything that could happen by its probability and accumulate these probability distributions. Joint distributions of all events, an eventual goal, will allow us to ask counterfactuals — “what if I do *this*” or “what if my opponent does *that*” — and simulate games to learn how decisions change win probability. It enables optimal strategy.

Rational and optimal, this approach is more efficient for gaining wins. For perspective, each added win from the free-agent market costs 10 million, give or take, and the league salary cap prevents unlimited spend on talent. There is no cap, however, on investing in rational decision processes.

Computational issues are being addressed in Stan, a tool that enables inferences through advanced simulations. This open-source software is free but teaching its applications will require time. To shorten our learning curve, we can start with Stan interfaces that use familiar syntax (like lme4) but return joint probability distributions: R packages `rethinking`, `brms`, or `rstanarm`. And we can test the concept with decisions to steal.

**Limitations,
short-term
solutions**

Sincerely,
Scott Spencer



To: **Scott Powers**
Senior Analyst, Baseball Analytics

2 February 2019

Our game decisions should optimize expectations. Proposal for testing: model decisions to steal

The most likely sequence of events on defense is a *perfect game* — occurring just 23 times in major-league baseball, once by our own Sandy Koufax. Decisions from what is most likely, however, leave wins unclaimed. To claim them, let's base decisions on expectations flowing from decision theory and probability models. A joint model of all events works best, but we can start small, with say, decisions to steal second base.

After defining our objective (e.g. optimize expected runs) we will from Statcast data compute expectations: weight everything that could happen by its probability and accumulate these probability distributions. Joint distributions of all events, an eventual goal, will allow us to ask counterfactuals — “what if I do *this*” or “what if my opponent does *that*” — and simulate games to learn how decisions change win probability. It enables optimal strategy.

Rational and optimal, this approach is more efficient for gaining wins. For perspective, each added win from the free-agent market costs 10 million, give or take, and the league salary cap prevents unlimited spend on talent. There is no cap, however, on investing in rational decision processes.

Computational issues are being addressed in Stan, a tool that enables inferences through advanced simulations. This open-source software is free but teaching its applications will require time. To shorten our learning curve, we can start with Stan interfaces that use familiar syntax (like lme4) but return joint probability distributions: R packages `rethinking`, `brms`, or `rstanarm`. And we can test the concept with decisions to steal.

**Reminder
of proposal**

Sincerely,
Scott Spencer



To: Scott Powers
Senior Analyst, Baseball Analytics

Audience background

Senior Analyst
Ph.D. Statistics from Stanford University
Some publications use machine learning
Knows R programming
An employee, knows history of Dodgers

2 February 2019

Message first, context

Context, orient the audience

The most likely sequence of events on defense is a *perfect game* — occurring just 23 times in major-league baseball, once by our own Sandy Koufax. Decisions from what is most likely, however, leave wins unclaimed. To claim them, let's base decisions on expectations flowing from decision theory and probability models. A joint model of all events works best, but we can start small, with say, decisions to steal second base.

Proposed solution

After defining our objective (e.g. optimize expected runs) we will from Statcast data compute expectations: weight everything that could happen by its probability and accumulate these probability distributions. Joint distributions of all events, an eventual goal, will allow us to ask counterfactuals — “what if I do *this*” or “what if my opponent does *that*” — and simulate games to learn how decisions change win probability. It enables optimal strategy.

Eventual benefit

Rational and optimal, this approach is more efficient for gaining wins. For perspective, each added win from the free-agent market costs 10 million, give or take, and the league salary cap prevents unlimited spend on talent. There is no cap, however, on investing in rational decision processes.

Benefit, comparison, financial

Computational issues are being addressed in Stan, a tool that enables inferences through advanced simulations. This open-source software is free but teaching its applications will require time. To shorten our learning curve, we can start with Stan interfaces that use familiar syntax (like lme4) but return joint probability distributions: R packages rethinking, brms, or rstanarm. And we can test the concept with decisions to steal.

Reminder of proposal

Sincerely,
Scott Spencer

Goal, action problem

Objective, data, methods

Limitations, short-term solutions

“When we read prose, we hear it... it’s variable sound. It’s sound with — pauses. With *emphasis*. With, well, you know, a certain rhythm.” — Richard Goodman

“If you **start your project early**, you’ll have time to **let your revised draft cool**. What seems good one day often looks different the next.” — Wayne Booth

Revise

“We write a first draft for ourselves; the drafts thereafter increasingly **for the reader**.” — Joseph Williams

Questions: tools for revising your memo

Sentence syntax, old before new? — Booth, chapter 17.

Needless words omitted? Every word tell? — S&W, composition 17.

Overstatements? — S&W, style 7.

Statements, in positive form? — S&W, composition 15.

Each paragraph, a separate unit of composition? — S&W, composition 16.

Definite, specific, concrete language? — S&W, composition 13.

Group work

Group help on case studies, memo

Get together in groups, work with your peers to get and give helpful feedback to refine your ideas for your choice of company, your data analytics project, potential sources of data, and writing it up.

Then, let's share some of these ideas together.

For Next Week, Module 4:

Agenda next week

The minimum

250-word memo due by 16th of June
Principles of Persuasion and Brief Proposals

Abelson, Robert P. Statistics as Principled Argument. Psychology Press, 1995. Print., Selected pages.

Read to understand his framework for using statistics as persuasive communication.

Also:

What's his ideal statistician?

What does he mean by *MAGIC*?

What are his thoughts about *comparisons*?

Conger J.A. (1998, May-June). *The necessary art of persuasion.* Harvard Business Review, 84-97.

What steps, in his view, must be considered?
What examples of successes and failures in these steps have you witnessed? Has he categorically omitted anything you consider important in his generalization of persuasion?

Checking in,

Reaching out?

Have you registered with Columbia's Writing Center? Scheduled a consultation with our Research Data Services?

**Coding?
Exploring data?**

What coding help do you need to gather, clean or visually explore your data? Be as specific as you can and we can setup a one-on-one.

See you
next week!

