

# **NISS Career Fair**

## **Statistics and Data Science Career Paths: Non-Healthcare Commercial Sectors**

**Victor S.Y. Lo**

**December 2019**

**“The best thing about being a statistician is that you get to play in everyone's backyard”. - From John Tukey, decades ago**



**“We no longer simply enjoy the privilege of playing in or cleaning up everyone's backyard. We are now being invited into everyone's study or living room, and trusted with the task of being their offspring's first quantitative nanny.”**

**- From Xiaoli Meng (2009), Harvard University**



Source:

Meng, X. (2009), “Desired and Feared—What Do We Do Now and Over the Next 50 Years?” *American Statistician*, v.63.

# Content

- ▶ **Statistics Career Opportunities**
- ▶ **Data Science and Analytics**
- ▶ **Operations Management**
- ▶ **Investment Management**
- ▶ **Risk Management**
- ▶ **2019 Career Rankings**

Disclaimer: This presentation does not represent any opinions from Fidelity Investments

# Statistics Career Opportunities

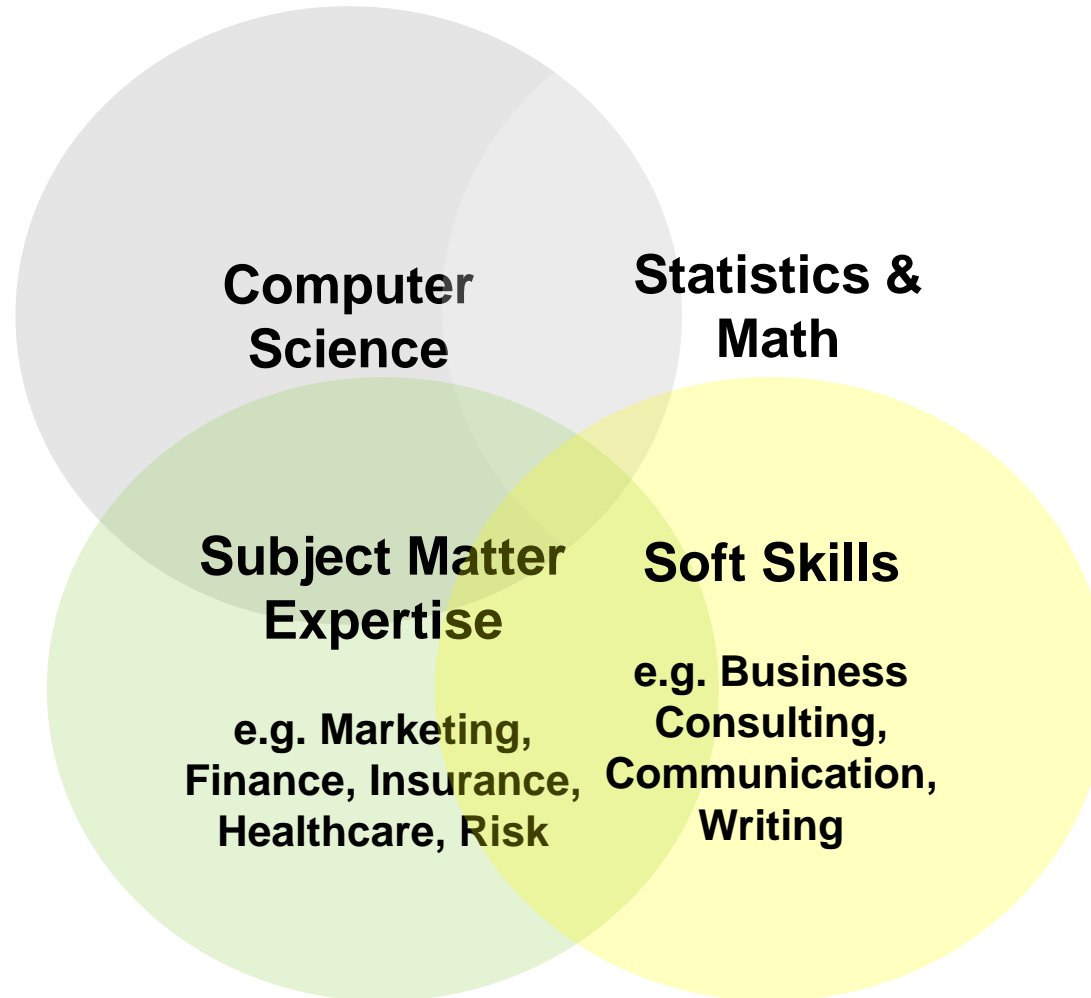
Career Category	Common Titles	Industry
<b>1) Healthcare</b>	Biostatistician, Statistician, Epidemiologist	Pharma, Hospitals, Medical Insurance, etc.
<b>2) Data Science, AI, and Analytics</b>	Data Scientist, Business Analyst, Machine Learning Engineer, Statistician, Marketing Scientist	<u>ALL</u> Industries
<b>3) Operations Management, Logistics, and Supply Chain Management</b>	Operations Research Analyst, Operations Analyst, Business Analyst, Data Scientist	Manufacturing, Transportation, Retail, Finance, Defense, and Others
<b>4) Investment Management</b>	Quantitative Analyst, Data Scientist	Finance
<b>5) Risk Management</b>	Quantitative Risk Analyst, Risk Manager	Banking, Consulting, Finance
<b>6) Actuarial Science</b>	Actuary, Actuarial Consultant	Insurance, Finance

# Statistics Career Opportunities

Career Category	Common Titles	Industry	Common Techniques	Typical Qualification <sup>1</sup>
<b>1) Healthcare</b>	Biostatistician, Statistician, Epidemiologist	Pharma, Hospitals, Medical Insurance, etc.	Experimental Design, Statistical Analysis, Causal Inference, Machine Learning	MS, PhD in Biostat, Stat, Epid
<b>2) Data Science, AI, and Analytics</b>	Data Scientist, Business Analyst, Machine Learning Engineer, Statistician, Marketing Scientist	<u>ALL</u> Industries	Machine Learning, Statistical Modeling, NLP, Experimental Design, Causal Inference, Advanced Programming, Survey Research	MS, MBA, PhD in CS, Stat, DS, Analytics, Econ, OR
<b>3) Operations Management, Logistics, and Supply Chain Management</b>	Operations Research Analyst, Operations Analyst, Business Analyst, Data Scientist	Manufacturing, Transportation, Retail, Finance, Defense, and Others	Time Series Forecasting, Machine Learning, Statistical Quality Control, Optimization, Discrete Event Simulation, NLP, Experimental Design	BA/BS, MS, MBA, PhD in OM, OR, IE, Stat, CS
<b>4) Investment Management</b>	Quantitative Analyst, Data Scientist	Finance	Time Series Analysis, Statistical Modeling, Optimization, Financial Engineering, NLP	MS, MBA, PhD in Finance, Econ, Physics, Stat, CS, Eng, plus CFA
<b>5) Risk Management</b>	Quantitative Risk Analyst, Risk Manager	Banking, Consulting, Finance	Risk Analytics, Time Series, Statistical Modeling, Machine Learning, Cryptography	MS, PhD in Econ, Stat, Finance, Eng, plus FRM
<b>6) Actuarial Science</b>	Actuary, Actuarial Consultant	Insurance, Finance	Probability, Statistical Modeling, Time Series, Risk Analytics	BS/BA, MS in Math, AS, Stat, plus ASA/FSA

<sup>1</sup> CS = Computer Science, DS = Data Science, OR = Operations Research, OM = Operations Management, IE = Industrial Engineering, Eng = Engineering, AS = Actuarial Science, CFA = Chartered Financial Analyst, FRM = Financial Risk Manager, ASA = Associate of the Society of Actuaries, FSA = Fellow of the Society of Actuaries

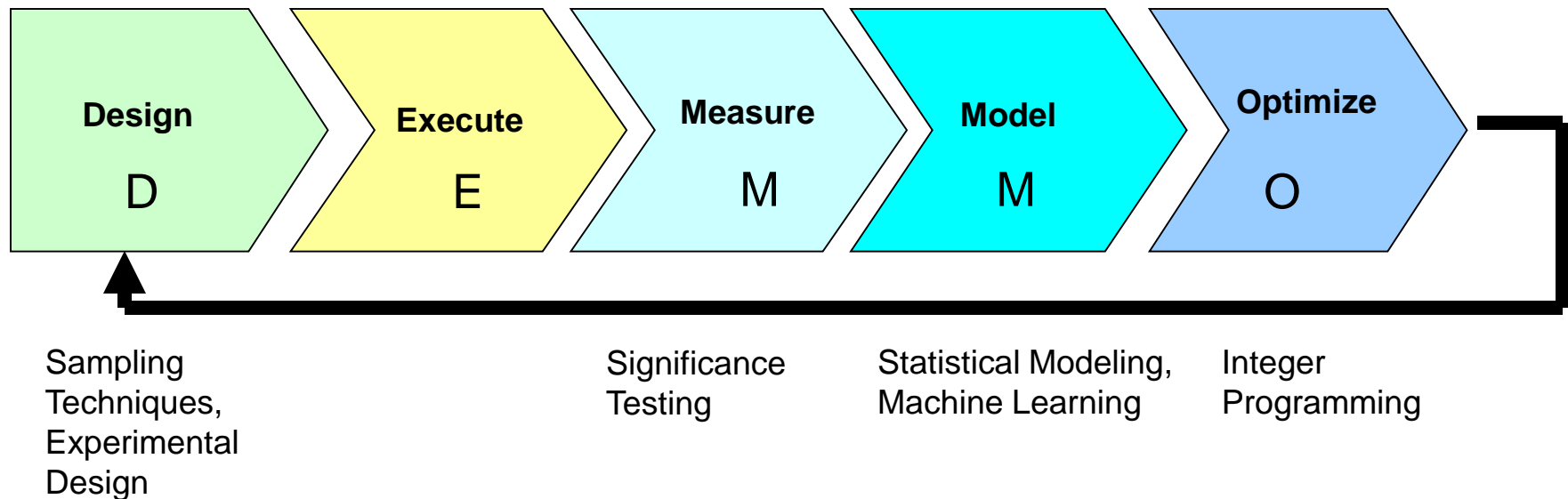
# Data Science Venn Diagram



See also: Lo (2019) <https://pubsonline.informs.org/doi/10.1287/LYTX.2019.04.02/full/>

# ▶ Data Science & Analytics: Customer Relationship Management (CRM)

- ▶ Track and optimize contacts with customers
- ▶ Use historical customer contact data and apply analytics to improve future customer interactions:



# Operations Management: Call Center Analytics



- Exponential Smoothing
- Box-Jenkins
- Deep Learning

Predictive Analytics



- Queuing Theory
- Discrete Event Simulation

Prescriptive Analytics



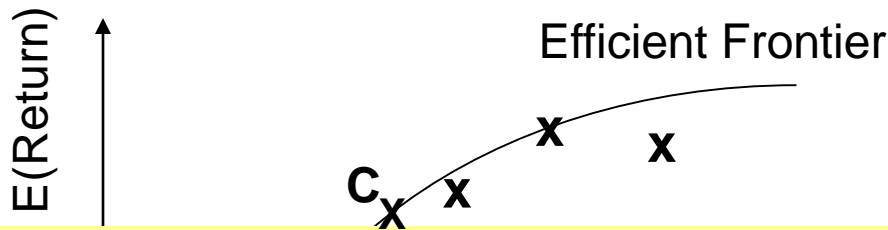
- Linear Programming
- Stochastic Programming
- Intra-day dynamic scheduling



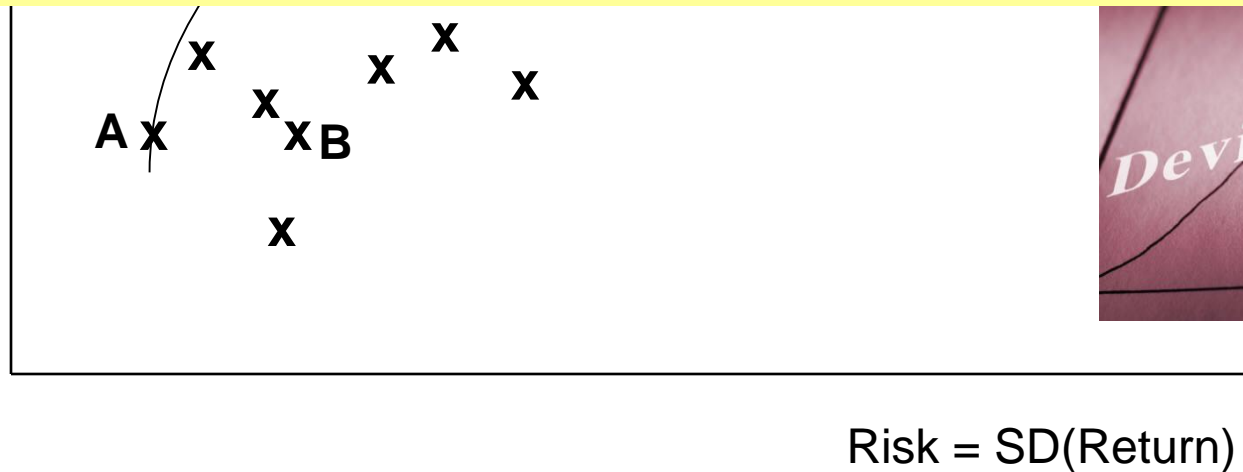
# Investment Management: Portfolio Optimization

## Modern Portfolio Theory – Harry Markowitz 1952

### Risk and Return Balance, where Risk = Volatility = SD



**Markowitz won the 1990 Nobel Prize in Economics on this!**



Find the investment portfolio that minimizes SD subject to mean constraint, based on Quadratic Programming, see Elton et al (2007) and Savage (2009) - a solid career by itself!

# Risk Management

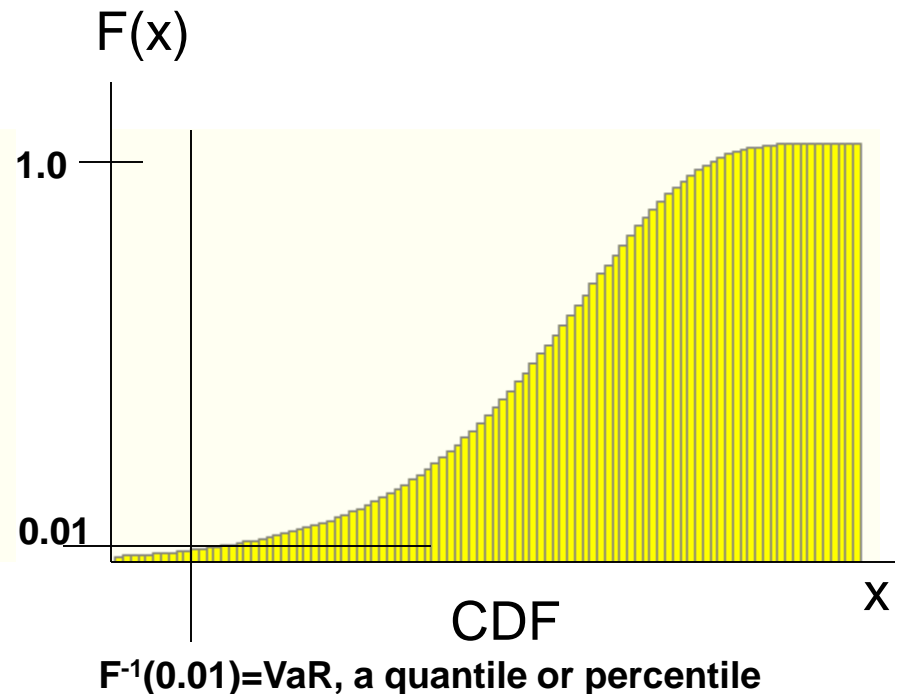
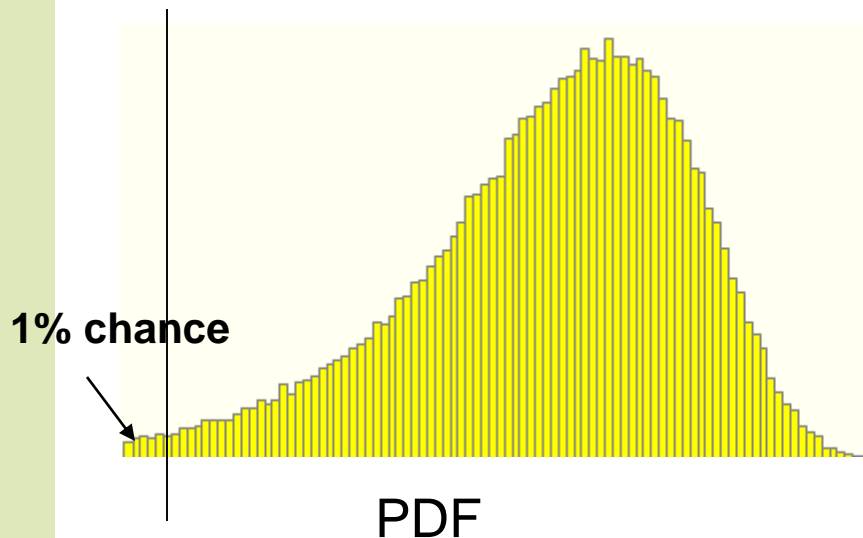
## Modern Risk Management – Value at Risk (VaR), Jorion (2007) and Bessis (2010):

$$P(\text{Return} \leq \text{VaR}) = 0.01, \text{ etc.}$$

I.e., the 1% worst case scenario

How to calculate – assume or fit a distribution for Return, and then calculate VaR

1% worst case = VaR



# 2019 Career Rankings

**Glassdoor - Top Job: Data Scientist**

[https://www.glassdoor.com/List/Best-Jobs-in-America-LST\\_KQ0,20.htm](https://www.glassdoor.com/List/Best-Jobs-in-America-LST_KQ0,20.htm)

**CareerCast - #1 Best Job: Data Scientist, #2: Statistician, #8: Mathematician, #9: Operations Research Analyst, #10 Actuary**

<https://www.careerCast.com/jobs-rated/2019-jobs-rated-report>

## Summary:

Data Scientist, Statistician, Actuary, Operations Research Analyst, Machine Learning Engineer, or Mathematician are among the best careers

**US News - #1 Best Business Job: Statistician, #2: Mathematician, #6: Actuary, #7: Operations Research Analyst**

<https://money.usnews.com/careers/best-jobs/rankings/best-business-jobs>

**LinkedIn - #1 Most Promising Job: Data Scientist, #15: Machine Learning Engineer**

<https://blog.linkedin.com/2019/january/10/linkedin-s-most-promising-jobs-of-2019>

# References

- C. Alexander (2008), *Practical Financial Econometrics*. Wiley.
- I. Ayres (2007), *Super Crunchers*. Bantam Books.
- S. Baker (2006), “Math will Rock Your World,” *BusinessWeek*, Jan 23, 2006, available at [http://www.businessweek.com/magazine/content/06\\_04/b3968001.htm](http://www.businessweek.com/magazine/content/06_04/b3968001.htm)
- J. Bessis (2010), *Risk Management in Banking*, 3rd edition. Wiley.
- G. Box (2006), *Improving Almost Anything: Ideas and Essays*. Wiley.
- T.H. Davenport (2008), “The Rise of Analytical Performance Management,” *SAS Institute whitepaper*, at: [http://www.sas.com/resources/whitepaper/wp\\_5596.pdf](http://www.sas.com/resources/whitepaper/wp_5596.pdf)
- T.H. Davenport and J.G. Harris (2007), *Competing on Analytics*. Harvard Business School Press.
- T.H. Davenport, J.G. Harris, and R. Morison (2010), *Analytics at Work*. Harvard Business Press.
- M.G. Dekimpe and D.M. Hanssens (2002), “Time Series Models in Marketing: Past, Present, and Future,” *International Journal of Research in Marketing*, p.183-193, at: <http://www.anderson.ucla.edu/faculty/dominique.hanssens/content/timeseries.pdf>
- E.J. Elton, M.J. Gruber, S.J. Brown, and W.N. Goetzmann (2007), *Modern Portfolio Theory and Investment Analysis*, 7<sup>th</sup> edition. Wiley.
- I. Goodfellow, Y. Bengio, and A. Courville (2016), *Deep Learning*, MIT Press.
- G.W. Imbens and D.B. Rubin (2015), *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*, Cambridge.
- P. Jorion (2007), *Value At Risk: The New Benchmark For Managing Financial Risk*. McGrawHill.
- V.S.Y. Lo (2002) “The True Lift Model – A Novel Data Mining Approach to Response Modeling in Database Marketing.” *SIGKDD Explorations* 4, Issue 2, p.78-86, at: [https://www.researchgate.net/profile/Victor\\_Lo3/publication/220520042\\_The\\_True\\_Lift\\_Model\\_-\\_A\\_Novel\\_Data\\_Mining\\_Approach\\_to\\_Response\\_Modeling\\_in\\_Database\\_Marketing/links/00463530681bd8f8de000000/The-True-Lift-Model-A-Novel-Data-Mining-Approach-to-Response-Modeling-in-Database-Marketing.pdf](https://www.researchgate.net/profile/Victor_Lo3/publication/220520042_The_True_Lift_Model_-_A_Novel_Data_Mining_Approach_to_Response_Modeling_in_Database_Marketing/links/00463530681bd8f8de000000/The-True-Lift-Model-A-Novel-Data-Mining-Approach-to-Response-Modeling-in-Database-Marketing.pdf)
- V.S.Y. Lo (2019) “Searching for the Perfect Unicorn,” *Analytics Magazine*, INFORMS, available at: <https://pubsonline.informs.org/doi/10.1287/LYTX.2019.04.02/full/>
- V.S.Y. Lo and D. Pachamanova (2015), “A Practical Approach to Treatment Optimization While Accounting for Estimation Risk,” *Journal of Marketing Analytics*, v.3, Issue 2, p.79-95.
- J. Pearl (2018), *The Book of Why: The New Science of Cause and Effect*, Basic Books.
- S.L. Savage (2009), *The Flaw of Averages: Why We Underestimate Risk in the Face of Uncertainty*. Wiley.
- S.J. Taylor (2005), *Asset Price Dynamics, Volatility, and Prediction*. Princeton.

# APPENDIX

# ▶ Analytics (Data Analytics, Business Analytics)

**Prescriptive  
Analytics**

**What should we do? What is the Best Decision?**

- Support *decision making* and *proactive* actions

**Predictive  
Analytics**

**What will happen?**

- Predict *future* forward-looking behavior, events, probabilities, or trends

**Descriptive  
Analytics**

**What happened?**

- Reports and profiling
- Data visualization
- *Business Intelligence*

Source: [http://www.sas.com/news/sascom/2008q4/column\\_8levels.html](http://www.sas.com/news/sascom/2008q4/column_8levels.html), and <https://www.informs.org/Community/Analytics>

# ▶ A.I. and Machine Learning Relationship

## A.I.

- 1956 Dartmouth Workshop
- Rule-Based: Teach a Machine What to Do

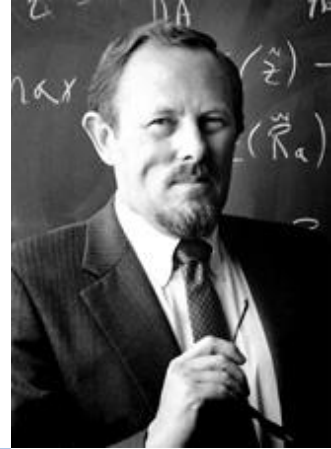
## **Machine Learning:** Let the machine learn

- Feed data and set a goal
- Highly related to Statistics

## Deep Learning

# ▶ Portfolio Management

## ▶ Capital Asset Pricing Model (CAPM) – William Sharpe



### Model:

Return of a stock - Risk-free Return  
=  $\beta \times (\text{Return of market index} - \text{Risk-free Return}) + \text{Error}$

- ▶ Easy to estimate statistically
- ▶ Simplified input requirement to Mean-Variance Optimization
- ▶ See Savage (2009), Elton et al. (2007), and Qian et al. (2007)

**Sharpe also won the 1990 Nobel Prize in Economics**



# ▶ Portfolio Management

- ▶ **Portfolio Optimization often requires a good estimate of future return and their variability (SD, variance)**
- ▶ **Robert Engle invented a way to model both mean and variance of return using historical data - “Autoregressive Conditional Heteroskedasticity: ARCH” :**



## Model:

**Return = Mean Return + residual**

**Variance Return =  $f$  (square of (previous return – mean return))**

- ▶ **Implying “volatility clustering,” See, e.g. Taylor (2005), Alexander (2008)**

**Engle won the 2003 Nobel Prize in Economics on this!**

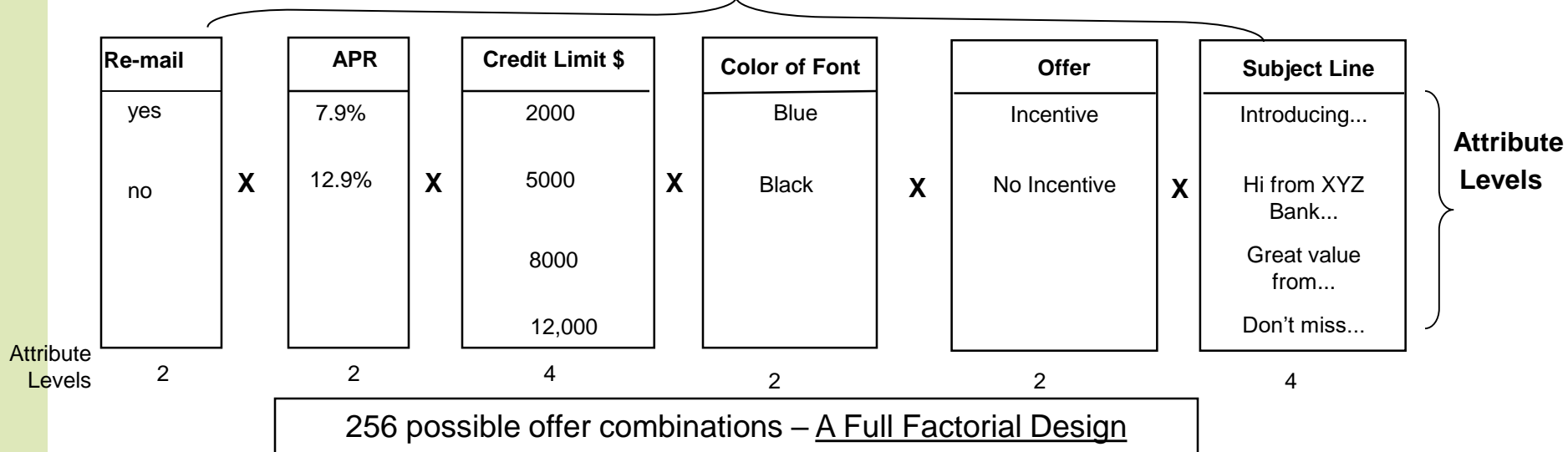
# ▶ Experimental Design Approach to Campaign Design

For Direct Marketing - testing particular treatment combinations of product and marketing features:

Treatment Combinations for a Banking Credit Card Email Campaign

*Hypothetical*

Attributes



From 256 to 32 cells

Fractional Factorial Design

Statistically Reduced to

cell	re-mail	APR	Credit limit	color	offer	subject
1	no	7.90%	2000	black	no incentive	Introducing...
2	no	7.90%	2000	blue	with incenti	Do not miss...
3	no	7.90%	5000	black	with incenti	Great value fr
4	no	7.90%	5000	blue	no incentive	Hi from XYZ Bank...
5	no	7.90%	8000	black	with incenti	Hi from XYZ Bank...
6	no	7.90%	8000	blue	no incentive	Great value fr
7	no	7.90%	12000	black	no incentive	Do not miss...
8	no	7.90%	12000	blue	with incenti	Introducing...
:	:	:	:	:	:	:
:	:	:	:	:	:	:

An 87.5% savings in # cells

# Application: CRM Optimization

