PASS Community News

February 2015

PASS
BUSINESS ANALYTICS
CONFERENCE
2015

APRIL 20 - 22, 2015
SANTA CLARA, CALIFORNIA

The Data Science Virtual Chapter is giving you the chance to save $150 off registration

Contact Jen.Stirrup@SQLPASS.org

Why Attend?

Learn from the world's top analytics experts in over 60 sessions

Network and connect with analytics experts from all over the world

Help your organization stay ahead of the curve by learning the latest

REGISTER TODAY

passbaconference.com

PASS

The discount code cannot be combined with other BA Conference
discount & cannot be used retroactively

Whatever your data passion – there's a Virtual Chapter for you!

PASS APPLICATION DEVELOPMENT VIRTUAL CHAPTER

PASS DATABASE ADMINISTRATION VIRTUAL CHAPTER

PASS GLOBAL PORTUGUESE VIRTUAL CHAPTER PT

PASS PERFORMANCE VIRTUAL CHAPTER

PASS BIG DATA VIRTUAL CHAPTER

PASS DBA FUNDAMENTALS VIRTUAL CHAPTER dba

PASS GLOBAL RUSSIAN VIRTUAL CHAPTER АЯ

PASS POWERSHELL VIRTUAL CHAPTER

PASS BOOK READERS VIRTUAL CHAPTER

PASS EXCEL BI VIRTUAL CHAPTER

PASS GLOBAL SPANISH VIRTUAL CHAPTER ñ

PASS PROFESSIONAL DEVELOPMENT VIRTUAL CHAPTER

PASS BUSINESS ANALYTICS VIRTUAL CHAPTER

PASS GLOBAL CHINESE VIRTUAL CHAPTER

PASS HEALTHCARE VIRTUAL CHAPTER

PASS SECURITY VIRTUAL CHAPTER

PASS BUSINESS INTELLIGENCE VIRTUAL CHAPTER

PASS GLOBAL FRANCOPHONE VIRTUAL CHAPTER

PASS HIGH AVAILABILITY & DISASTER RECOVERY VIRTUAL CHAPTER

PASS VIRTUALIZATION VIRTUAL CHAPTER

PASS CLOUD VIRTUAL CHAPTER

PASS GLOBAL HEBREW VIRTUAL CHAPTER עברית

PASS IN MEMORY VIRTUAL CHAPTER

PASS WOMEN IN TECHNOLOGY VIRTUAL CHAPTER W.I.T

PASS DATA ARCHITECTURE VIRTUAL CHAPTER

PASS GLOBAL ITALIAN VIRTUAL CHAPTER

PASS ORACLE/ SQL SERVER VIRTUAL CHAPTER

www.sqlpass.org/vc

*Data Science Virtual Chapter Coming Soon

Volunteering Opportunities

PASS would not exist without passionate, dedicated, and hardworking volunteers from around the globe.

Volunteer today!!

For local opportunities please visit
volunteer.sqlpass.org

For PASS HQ related activities, please update the "MyVolunteering" section of your *MyPASS* profile.

Stay Involved!

- Sign up for a *free* membership today at sqlpass.org

- Linked In: http://www.sqlpass.org/linkedin
- Facebook: http://www.sqlpass.org/facebook
- Twitter: @SQLPASS
- PASS: http://www.sqlpass.org

Introduction to Statistics and R for Rookies

Agenda:

The most important statistics for business

Using these statistics in R

You'll find the notes at:

- jenstirrup.com
- Datascience.sqlpass.org

Business and Statistics? Why?

Statistical analysis is used widely in businesses

Marketing – customer classification, spending patterns
Management consulting – efficient use of resources
Website design - testing

Numerical Measures – what is interesting?

Centre of the data
Spread of the data

Measures of Central Tendency

Mean – this is the average
Median – splits the data in two halves
Mode – the most popular value

Measures of Dispersion

Variance – average squared difference between the data points and the mean
Standard Deviation – square root of the variance, more intuitive
Percentiles – dataset is divided into 100 equal parts
Quartiles – dataset is divided into four equal parts
Interquartile range – middle 50% of data points

Measures of Association

Covariance – how variables vary together, rise together, fall together
Correlation – very similar, shown between -1 and 1

Measuring Uncertainty - Probability

Probability is based on SETS, which we use in SQL

We determine the probability of outcomes:

Addition Rule

Probability Distributions

Binomial distribution – one of two outcomes
Geometric Distribution – probability before success results
Poisson Distribution – probability that a number of events will occur within a time frame
Uniform Distribution – evenly distributed variables
Normal Distribution – bell shaped curve

What tools do we have in R?

80% of your time will be spent preparing and wrangling data

The remainder of your time will be spent complaining about it.

dplyr: the essential data manipulation toolset

In data wrangling, what are the main tasks?

– Filtering rows
– Selecting *columns* of data
– Adding new variables
– Sorting
– Aggregating