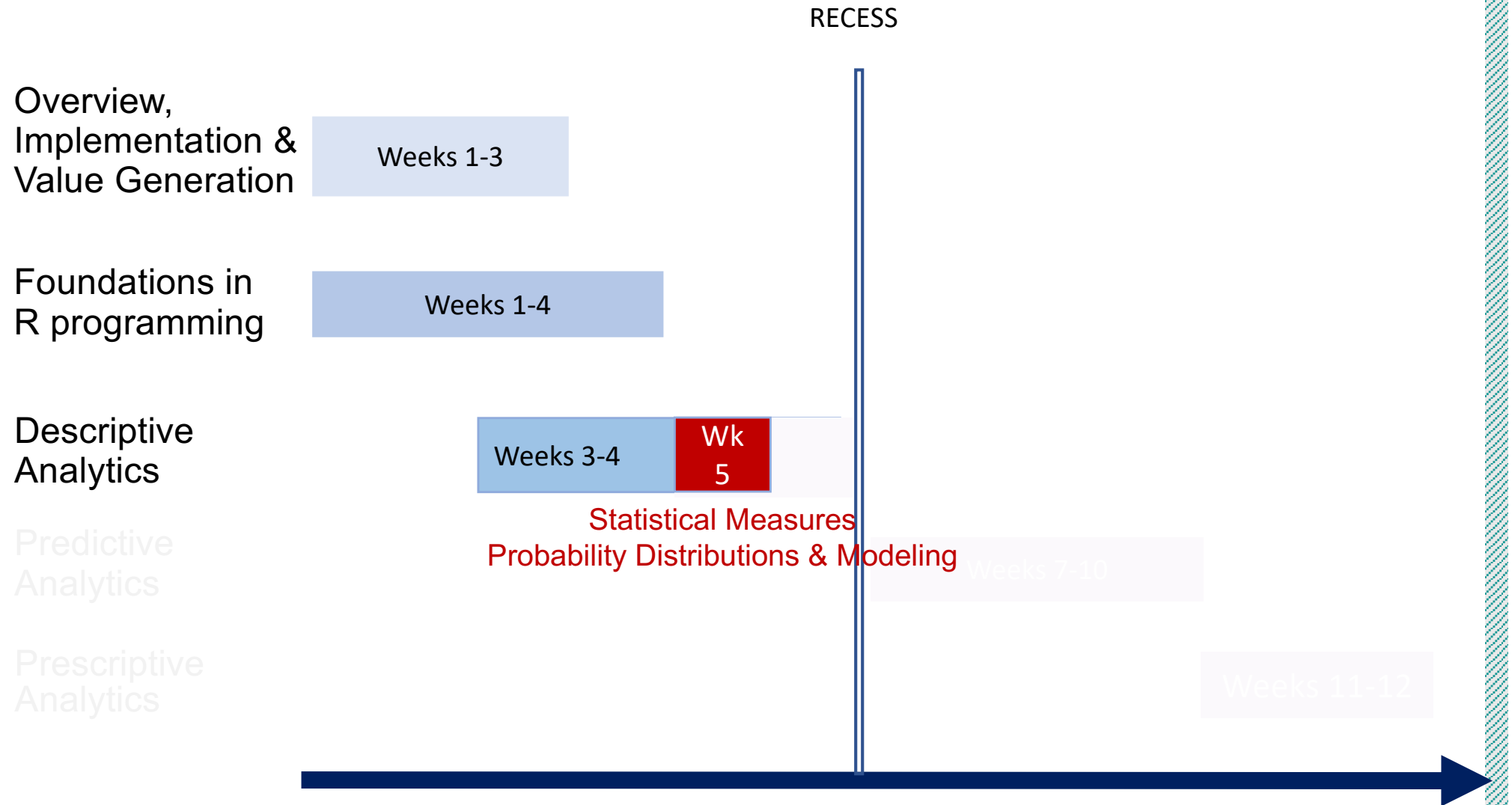**TBA2102 Introduction to Business Analytics**

Lecture 4
Statistical Measures
Probability Distribution & Data Modeling

Dr. Sharon Tan
9 Feb 2021

Course Map

RECESS

Overview, Implementation & Value Generation

Weeks 1-3

Foundations in R programming

Weeks 1-4

Descriptive Analytics

Weeks 3-4 | Wk 5

Statistical Measures
Probability Distributions & Modeling

Predictive Analytics

Weeks 7-10

Prescriptive Analytics

Weeks 11-12

Outline for today

Overview of Online Videos on Statistical Measures & Probability Distributions & Modeling
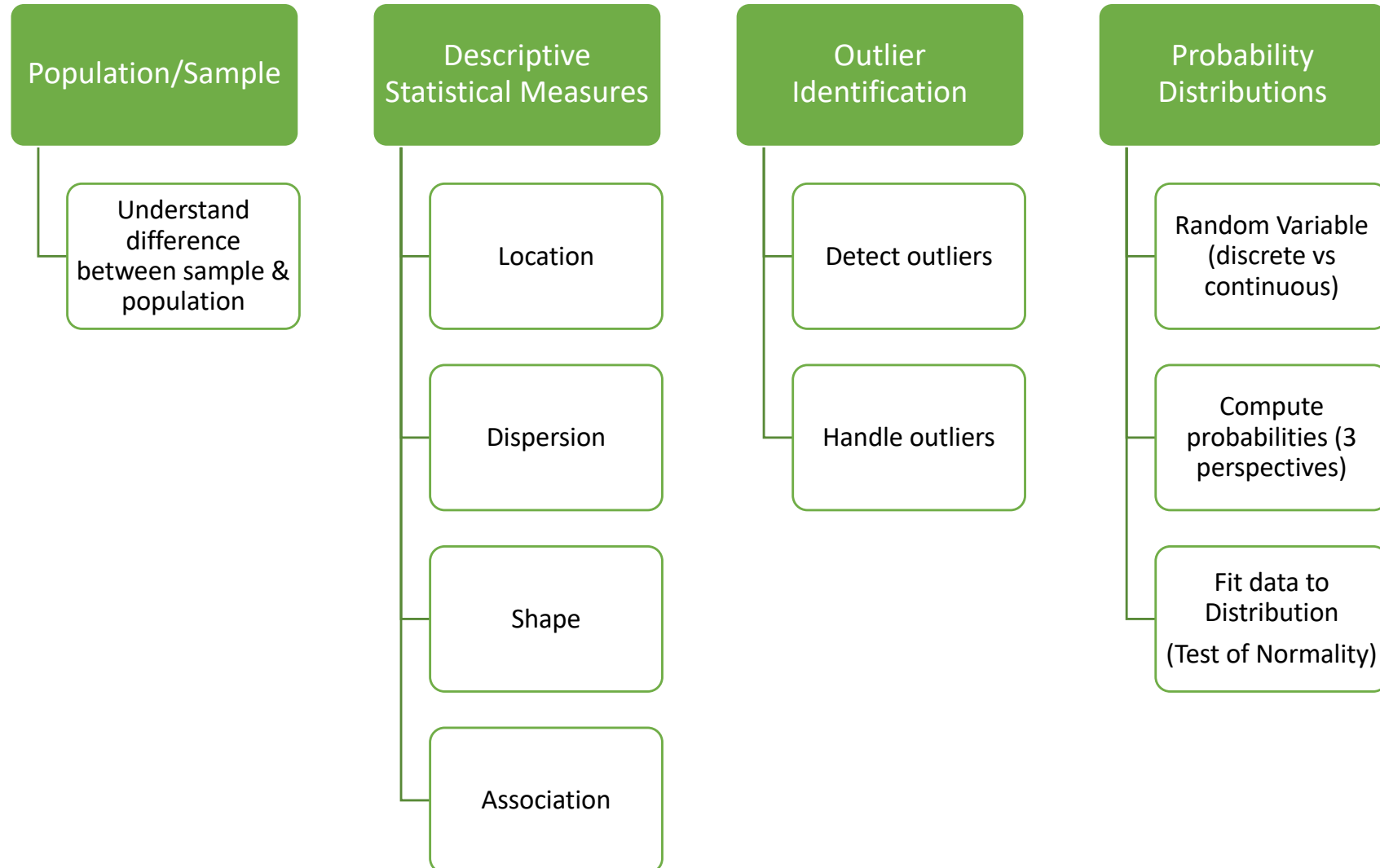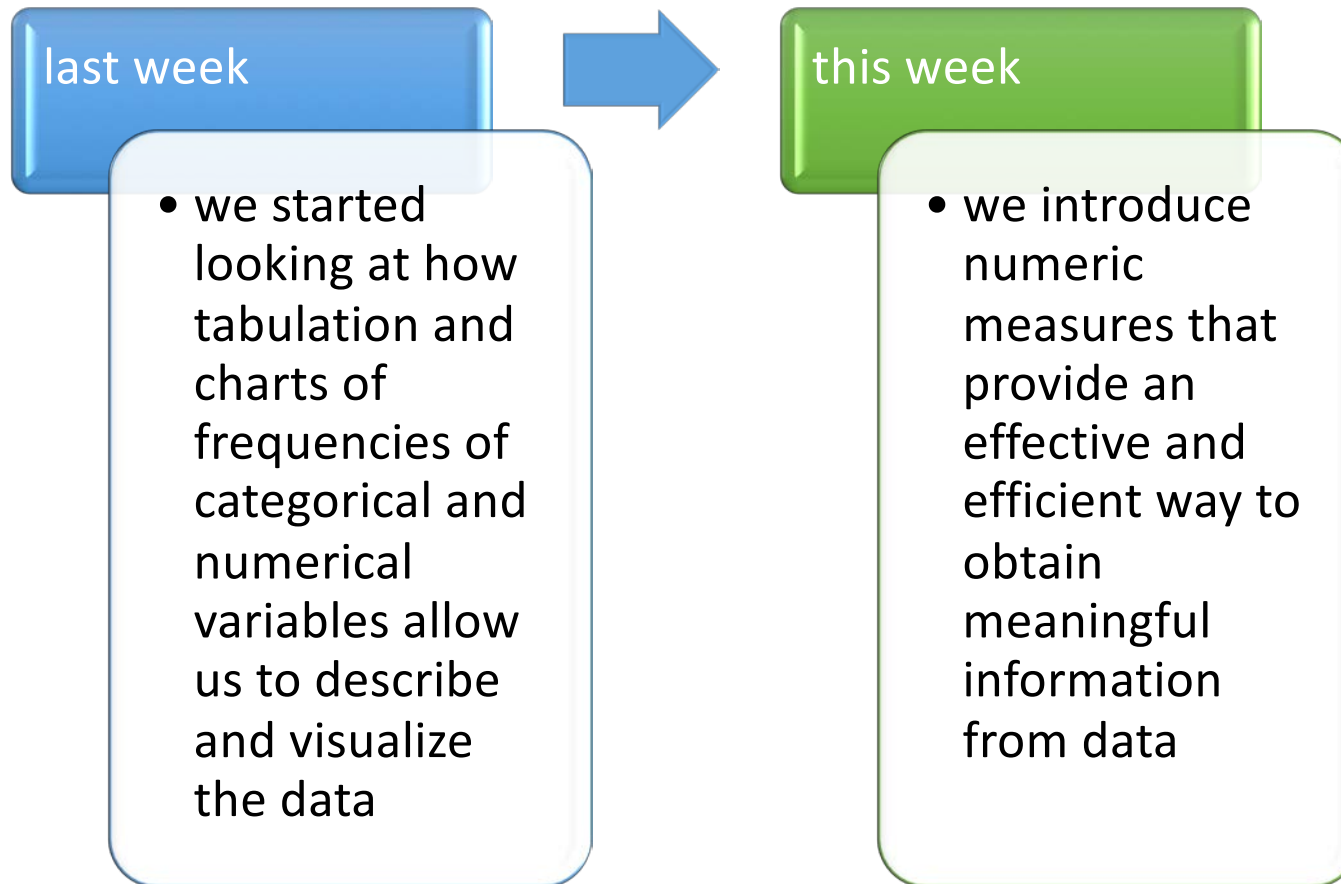
Applications of Statistical Measures

Integrating into Dashboards

Overview of Online Videos on Statistical Measures & Probability Distributions & Modeling

Population/Sample
- Understand difference between sample & population

Descriptive Statistical Measures
- Location
- Dispersion
- Shape
- Association

Outlier Identification
- Detect outliers
- Handle outliers

Probability Distributions
- Random Variable (discrete vs continuous)
- Compute probabilities (3 perspectives)
- Fit data to Distribution (Test of Normality)

Moving beyond Frequencies

last week

- we started looking at how tabulation and charts of frequencies of categorical and numerical variables allow us to describe and visualize the data

this week

- we introduce numeric measures that provide an effective and efficient way to obtain meaningful information from data

Applications of Statistical Measures

GRADUATE EMPLOYMENT SURVEY

6,609 fresh graduates and 870 follow-up graduates from NUS were surveyed in November 2019 and the overall response rates obtained were 77.9% and 74.7% respectively.

NUS: 2019 GES Employment Rates[1] and Salaries of Graduates by Bachelor Degree

Degree	Proportion of Graduates in the Labour Force who were		Basic Monthly Salary[4]		Gross Monthly Salary[5]			
	Employed[2]	In Full-Time Permanent Employment[3]	Mean	Median	Mean	Median	25th Percentile	75th Percentile
Faculty of Arts & Social Sciences								
Bachelor of Arts	84.9%	64.5%	$3,225	$3,000	$3,456	$3,200	$2,800	$3,700
Bachelor of Arts (Hons)	92.7%	68.2%	$3,421	$3,500	$3,537	$3,500	$3,250	$3,800
Bachelor of Social Sciences	88.0%	75.9%	$3,590	$3,500	$3,674	$3,500	$3,212	$3,850
Faculty of Dentistry								
Bachelor of Dental Surgery	100.0%	97.3%	$4,057	$4,200	$4,068	$4,200	$4,000	$4,500
Faculty of Engineering								
Bachelor of Engineering (Biomedical Engineering)	80.4%	68.6%	$3,480	$3,500	$3,535	$3,500	$3,200	$3,800
NUS Business School								
Bachelor of Business Administration	100.0%	82.6%	$4,799	$4,000	$4,925	$4,000	$3,400	$5,500
Bachelor of Business Administration (Hons)	95.5%	91.0%	$4,451	$4,000	$4,593	$4,150	$3,500	$5,000
Bachelor of Business Administration (Accountancy)*	93.8%	93.8%	$3,897	$3,000	$4,005	$3,000	$3,000	$3,800
Bachelor of Business Administration (Accountancy) (Hons)	97.8%	97.8%	$4,008	$3,500	$4,093	$3,600	$3,000	$4,500
School of Computing								
Bachelor of Computing (Computer Science)	92.8%	90.6%	$5,477	$5,000	$5,571	$5,000	$4,205	$5,585
Bachelor of Computing (Electronic Commerce)**	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
Bachelor of Computing (Information Security)**	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
Bachelor of Computing (Information Systems)	95.9%	95.9%	$4,210	$4,000	$4,330	$4,228	$4,000	$4,685
Bachelor of Science (Business Analytics)	100.0%	100.0%	$4,811	$5,000	$4,881	$5,000	$4,050	$5,200
Yale-NUS College								
Bachelor of Arts with Honours	96.0%	77.0%	$4,066	$3,600	$4,197	$3,684	$3,400	$4,567
Bachelor of Science with Honours*	94.7%	89.5%	$6,962	$5,000	$7,139	$5,000	$4,150	$9,967

aggregated data by university and year can be obtained here:
https://data.gov.sg/dataset/graduate-employment-survey-ntu-nus-sit-smu-suss-sutd?resource_id=9326ca53-9153-4a9c-b93f-8ae032637b70

GRADUATE EMPLOYMENT SURVEY

6,609 fresh graduates and 870 follow-up graduates from NUS were surveyed in November 2019 and the overall response rates obtained were 77.9% and 74.7% respectively.

NUS: 2019 GES Employment Rates[1] and Salaries of Graduates by Bachelor Degree

Degree	Proportion of Graduates in the Labour Force who were		Basic Monthly Salary[4]		Gross Monthly Salary[5]			
	Employed[2]	In Full-Time Permanent Employment[3]	Mean	Median	Mean	Median	25th Percentile	75th Percentile
Faculty of Arts & Social Sciences								
Bachelor of Arts	84.9%	64.5%	$3,225	$3,000	$3,456	$3,200	$2,800	$3,700
Bachelor of Arts (Hons)	92.7%	68.2%	$3,421	$3,500	$3,537	$3,500	$3,250	$3,800
Bachelor of Social Sciences	88.0%	75.9%	$3,590	$3,500	$3,674	$3,500	$3,212	$3,850
Faculty of Dentistry								
Bachelor of Dental Surgery	100.0%	97.3%	$4,057	$4,200	$4,068	$4,200	$4,000	$4,500
Faculty of Engineering								
Bachelor of Engineering (Biomedical Engineering)	80.4%	68.6%	$3,480	$3,500	$3,535	$3,500	$3,200	$3,800
NUS Business School								
Bachelor of Business Administration	100.0%	82.6%	$4,799	$4,000	$4,925	$4,000	$3,400	$5,500
Bachelor of Business Administration (Hons)	95.5%	91.0%	$4,451	$4,000	$4,593	$4,150	$3,500	$5,000
Bachelor of Business Administration (Accountancy)*	93.8%	93.8%	$3,897	$3,000	$4,005	$3,000	$3,000	$3,800
Bachelor of Business Administration (Accountancy) (Hons)	97.8%	97.8%	$4,008	$3,500	$4,093	$3,600	$3,000	$4,500
School of Computing								
Bachelor of Computing (Computer Science)	92.8%	90.6%	$5,477	$5,000	$5,571	$5,000	$4,205	$5,585
Bachelor of Computing (Electronic Commerce)**	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
Bachelor of Computing (Information Security)**	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
Bachelor of Computing (Information Systems)	95.9%	95.9%	$4,210	$4,000	$4,330	$4,228	$4,000	$4,685
Bachelor of Science (Business Analytics)	100.0%	100.0%	$4,811	$5,000	$4,881	$5,000	$4,050	$5,200
Yale-NUS College								
Bachelor of Arts with Honours	96.0%	77.0%	$4,066	$3,600	$4,197	$3,684	$3,400	$4,567
Bachelor of Science with Honours*	94.7%	89.5%	$6,962	$5,000	$7,139	$5,000	$4,150	$9,967

Population size = 6609

77.9% (or 5148) fresh grads responded
What does response rate tell us?
- any response bias?
- are the findings representative of the population?

GRADUATE EMPLOYMENT SURVEY

6,609 fresh graduates and 870 follow-up graduates from NUS were surveyed in November 2019 and the overall response rates obtained were 77.9% and 74.7% respectively.

NUS: 2019 GES Employment Rates[1] and Salaries of Graduates by Bachelor Degree

Degree	Proportion of Graduates in the Labour Force who were		Basic Monthly Salary[4]		Gross Monthly Salary[5]			
	Employed[2]	In Full-Time Permanent Employment[3]	Mean	Median	Mean	Median	25th Percentile	75th Percentile
Faculty of Arts & Social Sciences								
Bachelor of Arts	84.9%	64.5%	$3,225	$3,000	$3,456	$3,200	$2,800	$3,700
Bachelor of Arts (Hons)	92.7%	68.2%	$3,421	$3,500	$3,537	$3,500	$3,250	$3,800
Bachelor of Social Sciences	88.0%	75.9%	$3,590	$3,500	$3,674	$3,500	$3,212	$3,850
Faculty of Dentistry								
Bachelor of Dental Surgery	100.0%	97.3%	$4,057	$4,200	$4,068	$4,200	$4,000	$4,500
Faculty of Engineering								
Bachelor of Engineering (Biomedical Engineering)	80.4%	68.6%	$3,480	$3,500	$3,535	$3,500	$3,200	$3,800
NUS Business School								
Bachelor of Business Administration	100.0%	82.6%	$4,799	$4,000	$4,925	$4,000	$3,400	$5,500
Bachelor of Business Administration (Hons)	95.5%	91.0%	$4,451	$4,000	$4,593	$4,150	$3,500	$5,000
Bachelor of Business Administration (Accountancy)*	93.8%	93.8%	$3,897	$3,000	$4,005	$3,000	$3,000	$3,800
Bachelor of Business Administration (Accountancy) (Hons)	97.8%	97.8%	$4,008	$3,500	$4,093	$3,600	$3,000	$4,500
School of Computing								
Bachelor of Computing (Computer Science)	92.8%	90.6%	$5,477	$5,000	$5,571	$5,000	$4,205	$5,585
Bachelor of Computing (Electronic Commerce)**	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
Bachelor of Computing (Information Security)**	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
Bachelor of Computing (Information Systems)	95.9%	95.9%	$4,210	$4,000	$4,330	$4,228	$4,000	$4,685
Bachelor of Science (Business Analytics)	100.0%	100.0%	$4,811	$5,000	$4,881	$5,000	$4,050	$5,200
Yale-NUS College								
Bachelor of Arts with Honours	96.0%	77.0%	$4,066	$3,600	$4,197	$3,684	$3,400	$4,567
Bachelor of Science with Honours*	94.7%	89.5%	$6,962	$5,000	$7,139	$5,000	$4,150	$9,967

Salary data was based on full-time permanently employed

GRADUATE EMPLOYMENT SURVEY

6,609 fresh graduates and 870 follow-up graduates from NUS were surveyed in November 2019 and the overall response rates obtained were 77.9% and 74.7% respectively.

NUS: 2019 GES Employment Rates[1] and Salaries of Graduates by Bachelor Degree

Degree	Proportion of Graduates in the Labour Force who were		Basic Monthly Salary[4]		Gross Monthly Salary[5]			
	Employed[2]	In Full-Time Permanent Employment[3]	Mean	Median	Mean	Median	25th Percentile	75th Percentile
Faculty of Arts & Social Sciences								
Bachelor of Arts	84.9%	64.5%	$3,225	$3,000	$3,456	$3,200	$2,800	$3,700
Bachelor of Arts (Hons)	92.7%	68.2%	$3,421	$3,500	$3,537	$3,500	$3,250	$3,800
Bachelor of Social Sciences	88.0%	75.9%	$3,590	$3,500	$3,674	$3,500	$3,212	$3,850
Faculty of Dentistry								
Bachelor of Dental Surgery	100.0%	97.3%	$4,057	$4,200	$4,068	$4,200	$4,000	$4,500
Faculty of Engineering								
Bachelor of Engineering (Biomedical Engineering)	80.4%	68.6%	$3,480	$3,500	$3,535	$3,500	$3,200	$3,800
NUS Business School								
Bachelor of Business Administration	100.0%	82.6%	$4,799	$4,000	$4,925	$4,000	$3,400	$5,500
Bachelor of Business Administration (Hons)	95.5%	91.0%	$4,451	$4,000	$4,593	$4,150	$3,500	$5,000
Bachelor of Business Administration (Accountancy)*	93.8%	93.8%	$3,897	$3,000	$4,005	$3,000	$3,000	$3,800
Bachelor of Business Administration (Accountancy) (Hons)	97.8%	97.8%	$4,008	$3,500	$4,093	$3,600	$3,000	$4,500
School of Computing								
Bachelor of Computing (Computer Science)	92.8%	90.6%	$5,477	$5,000	$5,571	$5,000	$4,205	$5,585
Bachelor of Computing (Electronic Commerce)**	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
Bachelor of Computing (Information Security)**	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
Bachelor of Computing (Information Systems)	95.9%	95.9%	$4,210	$4,000	$4,330	$4,228	$4,000	$4,685
Bachelor of Science (Business Analytics)	100.0%	100.0%	$4,811	$5,000	$4,881	$5,000	$4,050	$5,200
Yale-NUS College								
Bachelor of Arts with Honours	96.0%	77.0%	$4,066	$3,600	$4,197	$3,684	$3,400	$4,567
Bachelor of Science with Honours*	94.7%	89.5%	$6,962	$5,000	$7,139	$5,000	$4,150	$9,967

What aggregated salary data is provided here?

GRADUATE EMPLOYMENT SURVEY

6,609 fresh graduates and 870 follow-up graduates from NUS were surveyed in November 2019 and the overall response rates obtained were 77.9% and 74.7% respectively.

NUS: 2019 GES Employment Rates[1] and Salaries of Graduates by Bachelor Degree

Degree	Proportion of Graduates in the Labour Force who were		Basic Monthly Salary[4]		Gross Monthly Salary[5]			
	Employed[2]	In Full-Time Permanent Employment[3]	Mean	Median	Mean	Median	25th Percentile	75th Percentile
Faculty of Arts & Social Sciences								
Bachelor of Arts	84.9%	64.5%	$3,225	$3,000	$3,456	$3,200	$2,800	$3,700
Bachelor of Arts (Hons)	92.7%	68.2%	$3,421	$3,500	$3,537	$3,500	$3,250	$3,800
Bachelor of Social Sciences	88.0%	75.9%	$3,590	$3,500	$3,674	$3,500	$3,212	$3,850
Faculty of Dentistry								
Bachelor of Dental Surgery	100.0%	97.3%	$4,057	$4,200	$4,068	$4,200	$4,000	$4,500
Faculty of Engineering								
Bachelor of Engineering (Biomedical Engineering)	80.4%	68.6%	$3,480	$3,500	$3,535	$3,500	$3,200	$3,800
NUS Business School								
Bachelor of Business Administration	100.0%	82.6%	$4,799	$4,000	$4,925	$4,000	$3,400	$5,500
Bachelor of Business Administration (Hons)	95.5%	91.0%	$4,451	$4,000	$4,593	$4,150	$3,500	$5,000
Bachelor of Business Administration (Accountancy)*	93.8%	93.8%	$3,897	$3,000	$4,005	$3,000	$3,000	$3,800
Bachelor of Business Administration (Accountancy) (Hons)	97.8%	97.8%	$4,008	$3,500	$4,093	$3,600	$3,000	$4,500
School of Computing								
Bachelor of Computing (Computer Science)	92.8%	90.6%	$5,477	$5,000	$5,571	$5,000	$4,205	$5,585
Bachelor of Computing (Electronic Commerce)**	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
Bachelor of Computing (Information Security)**	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
Bachelor of Computing (Information Systems)	95.9%	95.9%	$4,210	$4,000	$4,330	$4,228	$4,000	$4,685
Bachelor of Science (Business Analytics)	100.0%	100.0%	$4,811	$5,000	$4,881	$5,000	$4,050	$5,200
Yale-NUS College								
Bachelor of Arts with Honours	96.0%	77.0%	$4,066	$3,600	$4,197	$3,684	$3,400	$4,567
Bachelor of Science with Honours*	94.7%	89.5%	$6,962	$5,000	$7,139	$5,000	$4,150	$9,967

What can you tell about the data by comparing Mean & Median?
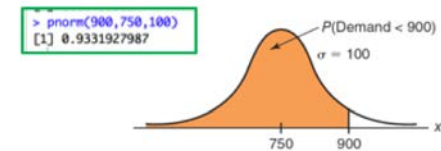
PSLE t-score

The Ministry of Education in Singapore has recently announced that it will change the Primary School Leaving Examination (PSLE) scoring system starting from 2021. Up till now, a student's PSLE performance is assessed by his/her aggregate T-score.

The aggregate T-score is a sum of the T-score for each of the 4 individuals subjects (English, Mother tongue, Science and Maths).

The T-score for each subject is computed by the formula: T-score = 50+ 10*(x-m)/s
where x is the candidate's mark for the subject;
m is the average mark (mean) scored by all the candidates
s is the spread of the marks around the average mark (standard deviation)

Below is an example of how a student's aggregate T-score will be computed:

Subjects	Student's raw score	Mean score (m)	Spread (s)	T-score
English	80	75	10	55
Mother Tongue	86	78	10	58
Science	83	73	20	55
Maths	90	81	15	56
			Aggregate T-score =	224

The aggregate T-score is thus a measure of how the student perform, relative to other students in the same cohort (or population). Assume students scores are normally distributed, answer the following questions.

Question: Timothy scored 90 marks for Maths. If the cohort mean and standard deviation for Maths is 78 and 7.5 respectively, what is Timothy's t-score for Math? What proportion of students would have a lower score than him?

PSLE t-score

The Ministry of Education in Singapore has recently announced that it will change the Primary School Leaving Examination (PSLE) scoring system starting from 2021. Up till now, a student's PSLE performance is assessed by his/her aggregate T-score.

The aggregate T-score is a sum of the T-score for each of the 4 individuals subjects (English, Mother tongue, Science and Maths).

The T-score for each subject is computed by the formula: T-score = 50+ 10*(x-m)/s
where x is the candidate's mark for the subject;
m is the average mark (mean) scored by all the candidates
s is the spread of the marks around the average mark (standard deviation)

Below is an example of how a student's aggregate T-score will be computed:

Subjects	Student's raw score	Mean score (m)	Spread (s)	T-score
English	80	75	10	55
Mother Tongue	86	78	10	58
Science	83	73	20	55
Maths	90	81	15	56
			Aggregate T-score =	224

The aggregate T-score is thus a measure of how the student perform, relative to other students in the same cohort (or population). Assume students scores are normally distributed, answer the following questions.

Question: What would the aggregate t-score be for a student who score the mean score for every subject?

PSLE t-score

The Ministry of Education in Singapore has recently announced that it will change the Primary School Leaving Examination (PSLE) scoring system starting from 2021. Up till now, a student's PSLE performance is assessed by his/her aggregate T-score.

The aggregate T-score is a sum of the T-score for each of the 4 individuals subjects (English, Mother tongue, Science and Maths).

The T-score for each subject is computed by the formula: T-score = 50+ 10*(x-m)/s
where x is the candidate's mark for the subject;
m is the average mark (mean) scored by all the candidates
s is the spread of the marks around the average mark (standard deviation)

Below is an example of how a student's aggregate T-score will be computed:

Subjects	Student's raw score	Mean score (m)	Spread (s)	T-score
English	80	75	10	55
Mother Tongue	86	78	10	58
Science	83	73	20	55
Maths	90	81	15	56
			Aggregate T-score =	224

The aggregate T-score is thus a measure of how the student perform, relative to other students in the same cohort (or population). Assume students scores are normally distributed, answer the following questions.

Question: What would be the minimum subject T-score for a student to be in the top 2.5% of the cohort for a subject? (hint: apply empirical rule)

Current Grading System till 2020

Grade	Mark Range
A*	91 – 100
A	75 – 90
B	60 – 74
C	50 – 59
D	35 – 49
E	20 -34
U	<20

Changes to PSLE in 2021

Standard level scoring bands

ACHIEVEMENT LEVEL*	RAW MARK RANGE	EXAMPLE
AL 1	≥ 90	ENGLISH + MATHEMATICS + SCIENCE + MOTHER TONGUE = **TOTAL AL SCORE**
AL 2	85 – 89	
AL 3	80 – 84	
AL 4	75 – 79	
AL 5	65 – 74	
AL 6	45 – 64	
AL 7	20 – 44	
AL 8	< 20	

* Students will have a score of AL1 - 8 for each subject

Foundation level scoring bands

AL EQUIVALENT*	RAW MARK RANGE	GRADE
AL 6	75 - 100	A
AL 7	30 – 74	B
AL 8	< 30	C

* Foundation level grades will be converted to their equivalent standard level subject scores for total AL score calculation

Secondary School Streaming Criteria	PSLE Score (until 2020)	PSLE Score (2021 onwards)
Express	≥ 200	4 - 20
Express / Normal (Academic) Option	188 - 199	21 - 22
Normal (Academic)	160 - 187	23 - 24
Normal (Academic) / Normal (Technical) Option	152 - 159	25
Normal (Technical)	< 152	26 - 30, with AL7 or better in both English **and** Mathematics

Secondary school placement criteria

TOTAL AL SCORE	PLACEMENT OUTCOME
4 - 20	EXPRESS
21 - 22	EXPRESS / N(A)
23 - 24	N(A)
25	N(A) / N(T)
26 - 30*	N(T)

* With AL 7 or better in both English and Mathematics

Question:
Do you think PSLE raw scores are normally distributed?

Integrating into Dashboards

Bank Credit Risk Analyses

Integrating into Dashboards

Bank Credit Risk Analyses

Let's add on to the dashboards we have created last week to help credit analysts understand their customers better.

- Customer Profile Dashboard
  - purpose: To understand the customer profile better. (Assume: customer characteristics of interests are gender, marital status, age, savings)
    - for categorical variables: pie chart (since there are few categories for each variable)
    - for continuous variables: histogram to show the distribution + descriptive statistics
    - association: scatterplot + correlation coefficient
- Loan Profile Dashboard
  - purpose: To understand the frequency of loans taken by the customers
    - loan purpose is categorical variable: bar chart (no change)
- Loan Customer Analyses Dashboard
  - purpose: To compare loan types across different customer profiles (e.g. gender)
    - grouped bar chart to compare frequencies across two categorical variables
  - purpose: To understanding Savings and Age across different loan types
    - Descriptive statistics grouped by loan types (and loan purpose & gender)
    - Barplot (or Grouped Barplot) of means
- Customer Savings Outlier Analyses Dashboard
  - purpose: To conduct outlier analyses on Savings
    - Check distribution, boxplot

Integrating into Dashboards

Restaurant Order Demand Analyses

|	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Day	Demand														
2	1	637														
3	2	718														
4	3	616														
5	4	859														
6	5	732														
7	6	617														
8	7	748														
9	8	773														
10	9	757														
11	10	669														
12	11	851														
13	12	738														
14	13	637														
15	14	478														
16	15	812														
17	16	695														
18	17	698														
19	18	794														
20	19	782														
21	20	759														
22	21	791														
23	22	778														
24	23	707														
25	24	501														
26	25	761														
27	26	694														
28	27	684														
29	28	552														
30	29	652														
31	30	741														
32	31	835														
33	32	689														
34	33	738														
35	34	694														
36	35	562														
37	36	658														
38	37	660														
39	38	694														

demand

Restaurant Order Demand Analyses

Let's build a dashboard to help interpret the demand data.

- Check for outliers
  - Check distribution, boxplot, rules of thumb
- Resolve outliers (if any)
  - ignore? remove? replace by mean?
- Build Demand Dashboard with "Clean" Dataset
  - Plots
  - Descriptive Statistics
  - Probability Computation
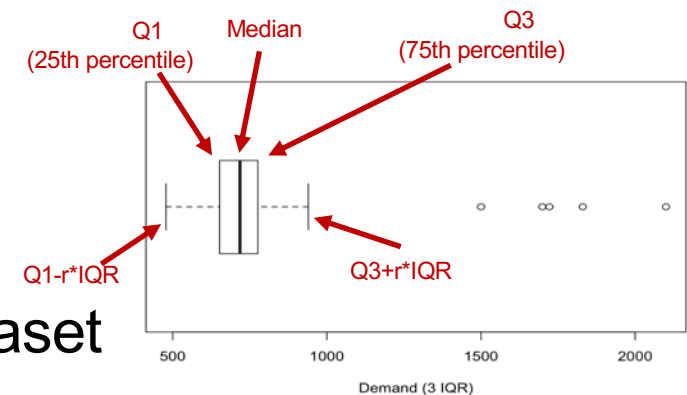
r: specified by "range" parameter in boxplot();
default = 1.5

Some rules of thumbs:

▸ z-scores > +3 or < -3

▸ Extreme outliers are > $3*IQR$ to the left of Q_1 or right of Q_3

▸ Mild outliers are between $1.5*IQR$ and $1.5*IQR$ to the left of Q_1 or right of Q_3

References for R:

- RStudio Cheatsheets
  - https://rstudio.com/resources/cheatsheets/
  - you may find some useful cheatsheets on packages you use (dplyr, base R, Rmarkdown)
- Rmarkdown Cheatsheet
  - https://github.com/rstudio/cheatsheets/raw/master/rmarkdown-2.0.pdf
- "psych" package documentation
  - https://cran.r-project.org/web/packages/psych/psych.pdf

THE END!

Thank You for Your Attention!