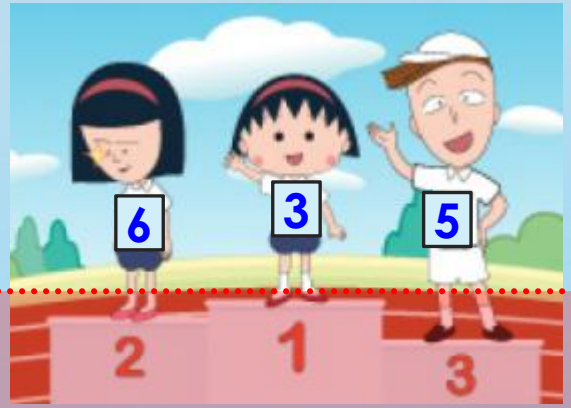
BACS2042 Research Methods

Data Analysis and Interpretation

Recap

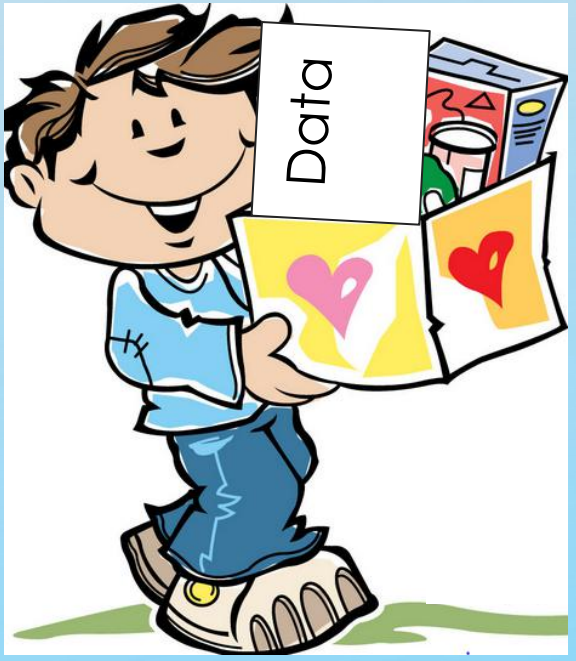
Nominal



Ordinal

Interval

Ratio

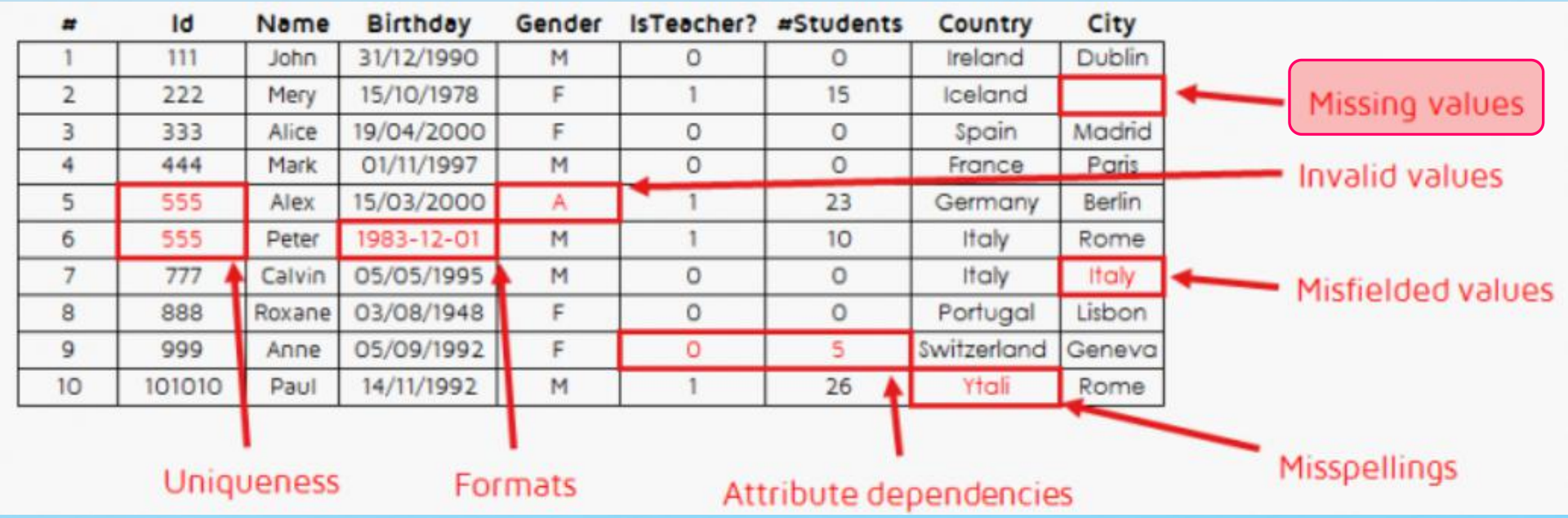


Data Collection

Data Preprocessing

Data Analysis and Interpretation

Data Preprocessing



#	Id	Name	Birthday	Gender	IsTeacher?	#Students	Country	City
1	111	John	31/12/1990	M	0	0	Ireland	Dublin
2	222	Mery	15/10/1978	F	1	15	Iceland	
3	333	Alice	19/04/2000	F	0	0	Spain	Madrid
4	444	Mark	01/11/1997	M	0	0	France	Paris
5	555	Alex	15/03/2000	A	1	23	Germany	Berlin
6	555	Peter	1983-12-01	M	1	10	Italy	Rome
7	777	Calvin	05/05/1995	M	0	0	Italy	Italy
8	888	Roxane	03/08/1948	F	0	0	Portugal	Lisbon
9	999	Anne	05/09/1992	F	0	5	Switzerland	Geneva
10	101010	Paul	14/11/1992	M	1	26	Ytali	Rome

Data Preprocessing

	id	school	gender	location	item1	item2	item3
1	1	1	2	1	5	3	4
2	2	2	1	1	5	3	4
3	3	3	1	2	4	5	4
4	4	4	1	2	2	4	3
5	5	3	2	2	1	5	5
6	6	4	1	2			5
7	7	2	2	1	4	5	4
8	8	1	2	1	4	4	4
9	9	2	1	1	5	5	5
10	10	3	2	2	1	4	5
11	11	4	2	2	1	5	5
12	12	3	1	2	2	4	5
13	13	3	2	2	.	4	4
14	14	3	1	2	1	5	5
15	15	2	2	1	4	3	5
16	16	3	2	2	5	5	3
17	17	4	1	2	5	4	4
18	18	4	2				
19	19	1	1				
20	20	2	2				
21	21	4	1				
22	22	3	1				
23	23	3	1				
24	24	1	2				
25	25	2	2				
26	26	3	1				

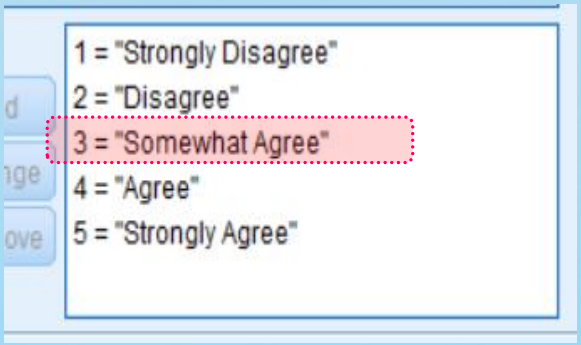
Missing value

location	Numeric	8	0	School location	{1, Urban}...	None	8
item1	Numeric	8	0	The lesson is interesting.	{1, Strongly ...	None	8
item2	Numeric	8	0	Teacher			
item3	Numeric	8	0	Time is			
item4	Numeric	8	0	Active p			
item5	Numeric	8	0	Underst			
pretest	Numeric	8	0	English			
posttest	Numeric	8	0	English			
problems	String	71	0	Problen			

Data Preprocessing: Missing Value

Missing Value

- Midpoint in the scale
- Mean Value
- Random number
- Remove the column



location	item1	item2	iter
1	5	3	
1	5	3	
2	4	5	
2	2	4	
2	1	5	
2	.	5	
1	4	5	
1	4	4	
1	5	5	
2	1	4	
2	1	5	
2	2	4	
2	.	4	
2	1	5	
1	4	3	
2	5	5	
2	5	4	
2	5	4	
1	4	4	
1	5	5	
2	4	4	
2	3	3	
2	2	5	
1	1	5	
1	3	4	
2	4	5	
1	5	3	

item1

location		item2
1		3
1		3
2		5
2		4
2		5
2		5
1		5
1		4
1		5
2		4
2		5
2		4
2		4
2		5
1		3
2		5
2		4
2		4
1		4
1		5
2		4
2		3
2		5
1		5
1		4
2		5

Data Analysis

Data Analysis

Objectives

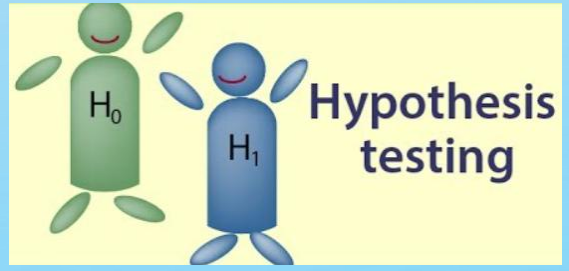
Getting a feel for the data

Testing the goodness of data

Testing the hypotheses

- Mean
- Std
- Range
- Variance
- Correlation

- Reliability of the data *(Cronbach's alpha)*



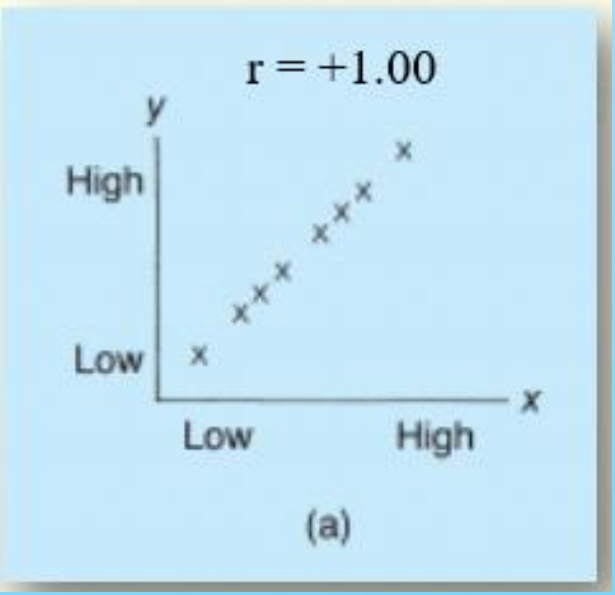
Data Analysis: Getting a feel for the data

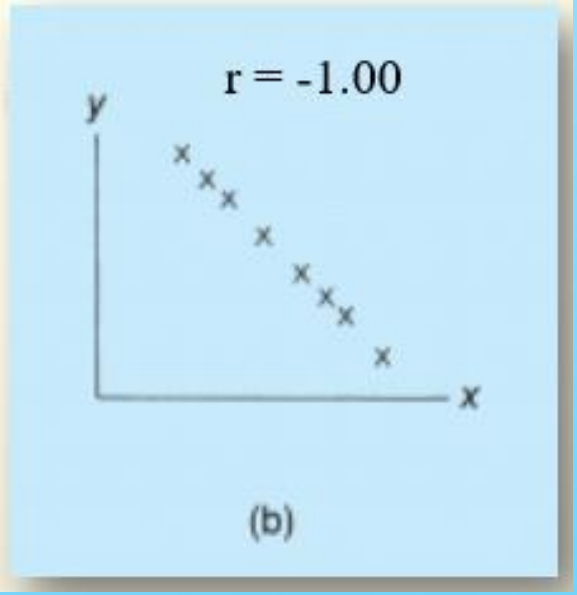
Correlation

Data Analysis: Getting a feel for the data

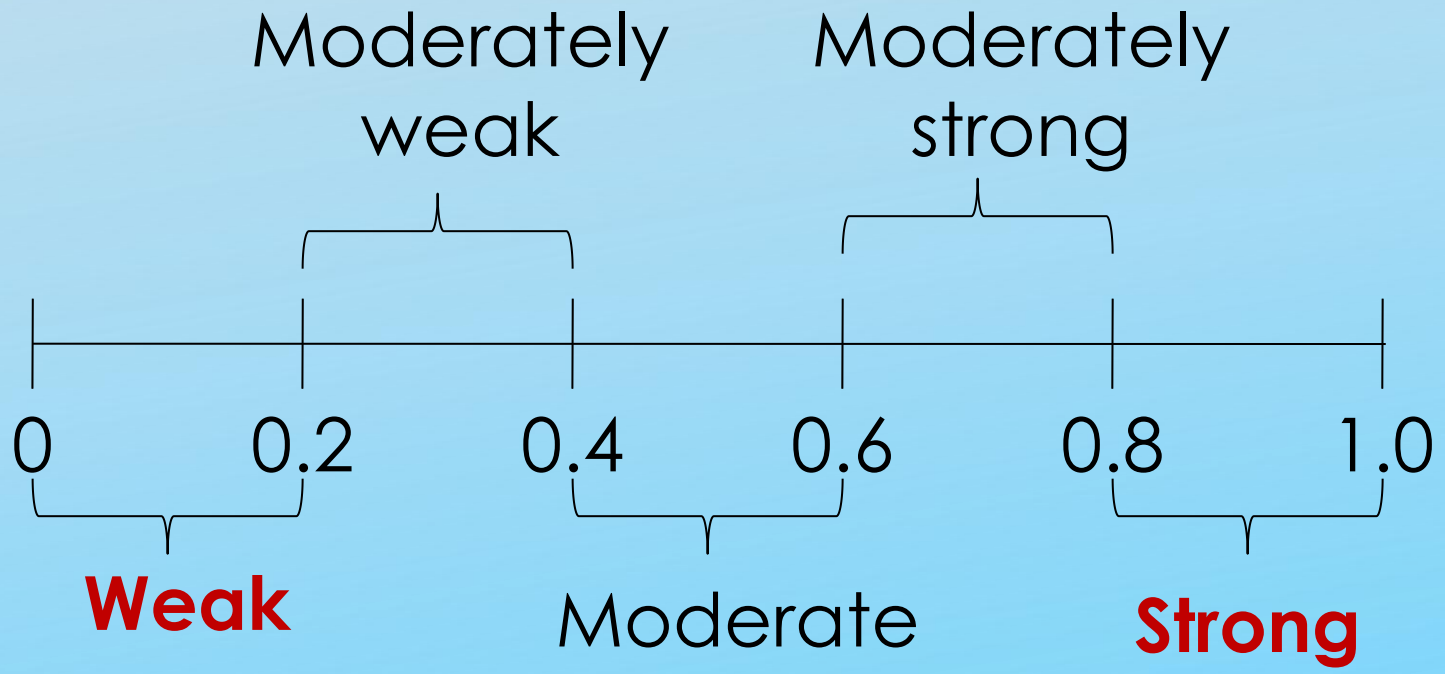
Correlation

- Correlation (*r*): the strength of the relationship between two variables
- a.k.a Pearson correlation coefficient
- +1.0 and -1.0





Correlation



Data Analysis: Getting a feel for the data

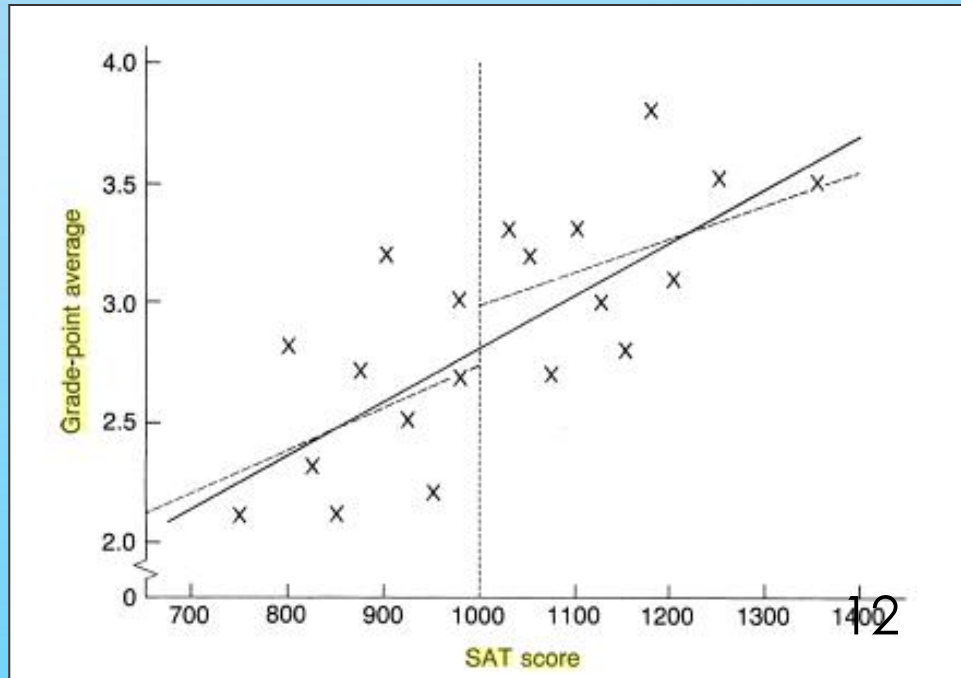
Regression

Data Analysis: Getting a feel for the data

Regression

- Definition: **predicting** the value of one variable from another based on their correlation
- Slope of the regression line (***m***) reflects relationship between two variables
- Equation:

$$y' = \boldsymbol{m}x + b$$



Data Analysis: Testing Goodness of Data

Learning
Self-motivation

		Strongly Disagree (1)	Disagree (2)	Undecided (3)	Agree (4)	Strongly Agree (5)
1	I like to read to learn	○	○	○	○	○
2	I seek out new information so I can learn new things	○	○	○	○	○
3	When I went to know more, I teach myself	○	○	○	○	○

Changes of words can impact our reactions

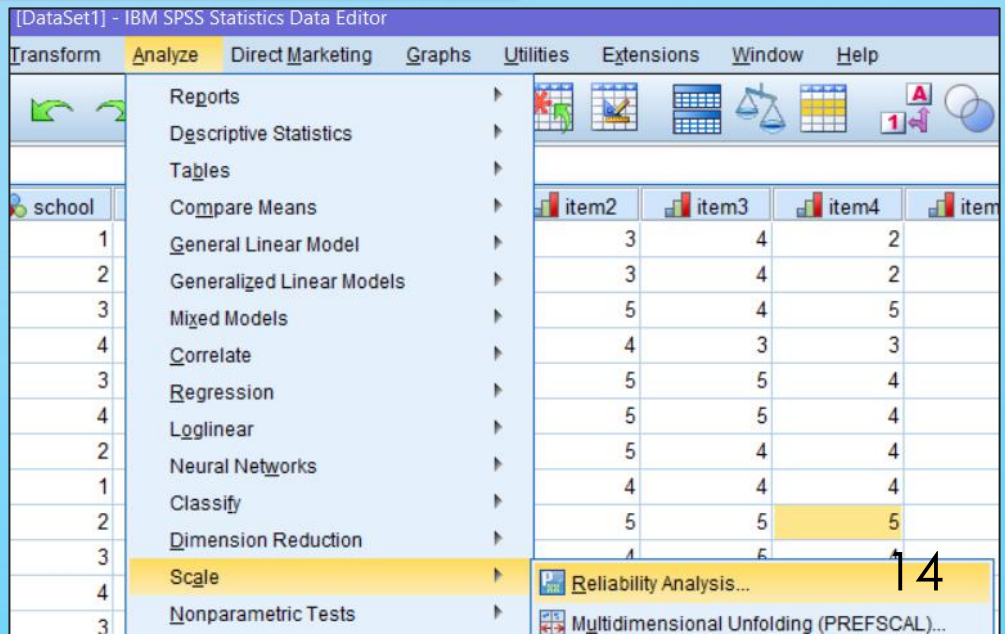
Data Analysis: Testing Goodness of Data

		Strongly Disagree (1)	Disagree (2)	Undecided (3)	Agree (4)	Strongly Agree (5)
1	I like to read to learn					
2	I seek out new information so I can learn new things					
3	When I went to know more, I teach myself					

Cronbach's alpha

- Internal Consistency

Cronbach's alpha, >0.7



Data Analysis: Hypotheses Testing

Analysis of Variance (ANOVA)

ANOVA

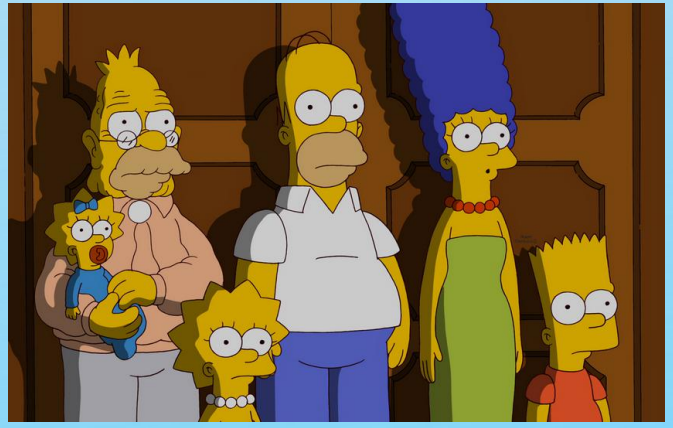
- Two estimates of the variability in the population:
  - Variability within the groups (within-conditions)
  - Variability between the groups (between-conditions)

Laziness is hereditary

ANOVA

Super Runner

Couch Potato

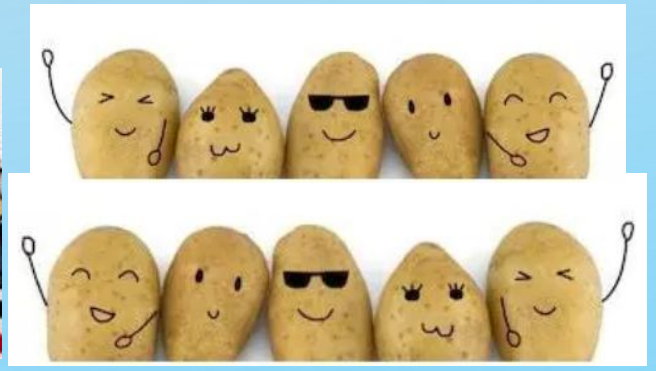


General

- Number of steps; different groups
- Is there any diff between the 3 groups?

ANOVA

Variability between the groups (between-conditions)



Variability within the groups (within-conditions)

ANOVA

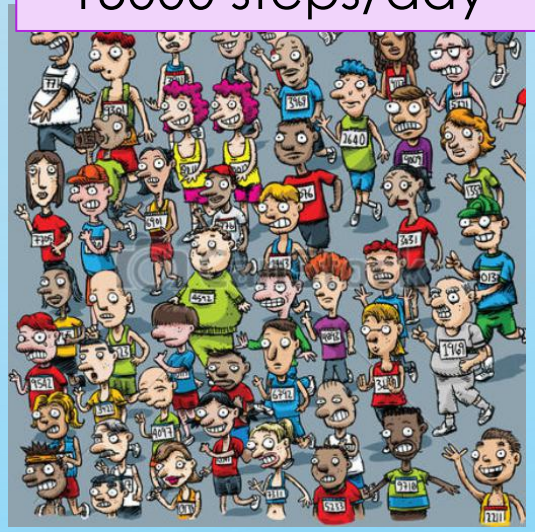
- Ratio **F**:

$$\frac{\text{Between-conditions variability}}{\text{within-conditions variability}}$$

$$F > 1.0$$

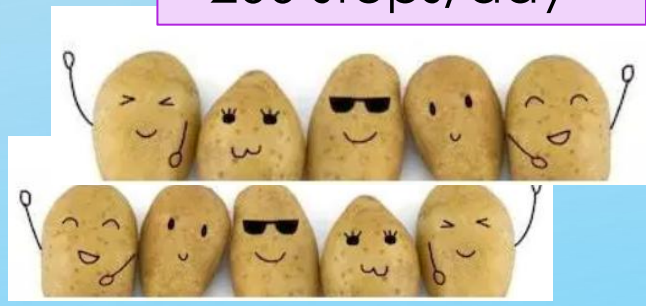
Variability between the groups (between-conditions)

~15000 steps/day



~5000 steps/day

~200 steps/day



Variability within the groups (within-conditions)

ANOVA Table Summary

- If there is an experimental effect, the variance attributable to the conditions (the mean square between conditions) will be larger than the variance attributable to the subjects (the mean square within conditions).
- If the F ratio is sufficiently greater than 1.0, it is considered significant.

ANOVA Summary

- Sum of squares (*SS*) – measure of the variability in the data
- Degrees of freedom (*df*) – quantity that depends on the number of groups, subjects
- Mean square (*MS*) – sum of squares divided by the number of degrees of freedom in the same row
- Ratio F (*F*)
- Probability (*p*) – significant of the effects. Normally if <0.05, the effects are significant.

Source	*SS*	*df*	*MS*	*F*	*p*
Between conditions	504	2	168	24	<.025
Within conditions	251	27	7		
Total	755	29			

Independent variable (IV)

Source	*SS*	*df*	*MS*	*F*	*p*
Between conditions	504	2	168	24	<.025
Within conditions	251	27	7		
Total	755	29			

ONE-WAY ANOVA because there is ONE IV

Sum of square: measure of the variability in the data

Source	*SS*	*df*	*MS*	*F*	*p*
Between conditions	504	2	168	24	<.025
Within conditions	251	27	7		
Total	755	29			

Degrees of freedom: quantity that depends on the number of groups, subjects

Source	*SS*	*df*	*MS*	*F*	*p*
Between conditions	504	2	168	24	<.025
Within conditions	251	27	7		
Total	755	29			

Source	*SS*	*df*	*MS*	*F*	*p*
Between conditions	504	2	168	24	<.025
Within conditions	251	27	7		
	755	29			

Four groups:
$df = K - 1$

Each group
10 subjects

30 subjects:
$df = N - 1$

Mean square:
SS/df

Source	*SS*	*df*	*MS*	*F*	*p*
Between conditions	504	2	168	24	<.025
Within conditions	251	27	7		
Total	755	29			

F = *MS* between condition / *MS* within conditions

Source	*SS*	*df*	*MS*	*F*	*p*
Between conditions	504	2	168	24	<.025
Within conditions	251	27	7		
Total	755	29			

Probability: significant of the effects.

Source	*SS*	*df*	*MS*	*F*	*p*
Between conditions	504	2	168	24	<.025
Within conditions	251	27	7		
Total	755	29			

Significant if <0.05

The probability that an experimental result happened by chance

Source	*SS*	*df*	*MS*	*F*	*p*
Between conditions	504	2	168	24	<.025
Within conditions	251	27	7		
Total	755	29			

Exercise

Exercise

Source	*SS*	*df*	*MS*	*F*	*p*
Between conditions	504	3	168	24	<.025
Within conditions	251	36	7		
Total	755	39			

Based on the ANOVA table given in the previous slide, answer the following questions:

a. How many subjects were in each group?

Exercise

Source	*SS*	*df*	*MS*	*F*	*p*
Between conditions	504	3	168	24	<.025
Within conditions	251	36	7		
Total	755	39			

Based on the ANOVA table given in the previous slide, answer the following questions:

b. How many subjects were in the experiment?

Exercise

Source	*SS*	*df*	*MS*	*F*	*p*
Between conditions	504	3	168	24	<.025
Within conditions	251	36	7		
Total	755	39			

Based on the ANOVA table given in the previous slide, answer the following questions:

c. Was the experiment effect significant?

Exercise

Source	*SS*	*df*	*MS*	*F*	*p*
Between conditions	504	3	168	24	<.025
Within conditions	251	36	7		
Total	755	39			

Based on the ANOVA table given in the previous slide, answer the following questions:

d. Show where the value of *F* came from.

Example: One-way ANOVA

Sources	*SS*	*df*	*MS*	*F*	*p*
Between subjects	167.33	9	18.59		
Between conditions	423.27	2	211.64	76.13	<.01
Residual (error)	50.07	18	2.78		
Total	640.67	29			

Repeated measure

df = n-1 = 10 groups

Example: One-way ANOVA

Sources	*SS*	*df*	*MS*	*F*	*p*
Between subjects	167.33	9	18.59		
Between conditions	423.27	2	211.64	76.13	<.01
Residual (error)	50.07	18	2.78		
Total	640.67	29			

30 observations:
df = N-1

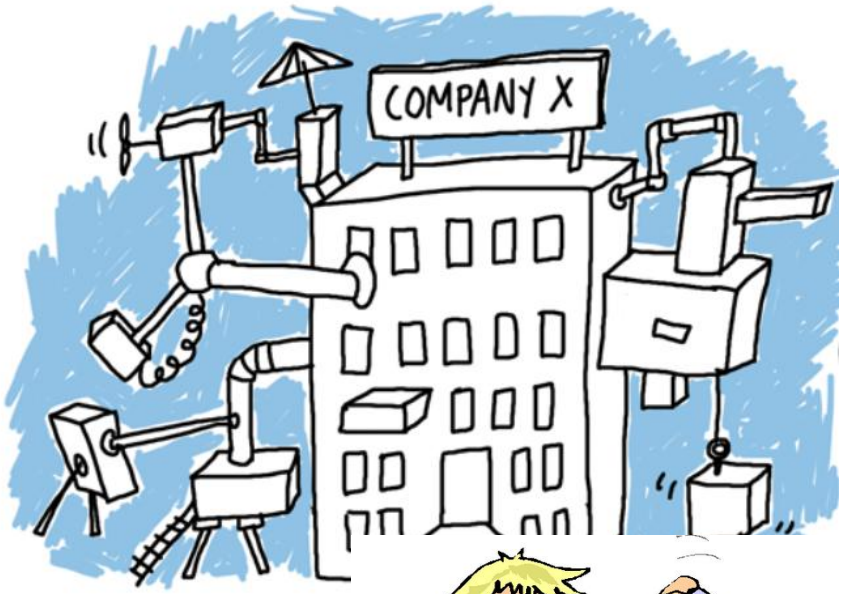
Example: One-way ANOVA

Sources	*SS*	*df*	*MS*	*F*	*p*
Between subjects	167.33	9	18.59		
Between conditions	423.27	2	211.64	76.13	<.01
Residual (error)	50.07	18	2.78		
Total	640.67	29			

F = MS between/MS within

Case Study

Data Analysis: Example



Factors influencing the intention to leave (ITL)

Data Analysis: Example

Dependent	Independent (Factors)
ITL	Job characteristic
	Perceived equality
	Burnout
	Job satisfaction

Data Analysis: Example

1. Men will perceive less equity than women (or women will perceive more equity than men).
2. The job satisfaction of individuals will vary depending on the shift they work.
3. Employees' intentions on leave (ITL) will vary according to their job title. In other words, there will be significant differences in the ITL of top managers, middle level managers, supervisors, and the clerical and blue-collar employees.
4. There will be a relationship between the shifts that people work (first, second, and third shift) and the part-time versus full-time status of employees. In other words, these two factors will not be independent.
5. The four independent variables of job characteristics, distributive justice, burnout, and job satisfaction will significantly explain the variance in intention to leave.

Data Analysis: Example

1. The establishment of Cronbach's alpha for the measures.
2. The frequency distribution of the variables.
3. Description statistics such as the mean and std.
4. The Pearson correlation matrix.
5. The results of hypotheses testing.

Data Analysis: Example

(1) Cronbach

Reliability Analysis

1. From the menus, choose:
   Analyze
   Scale
   Reliability Analysis…
2. Select the variables constituting the scale.
3. Choose Model *Alpha*.

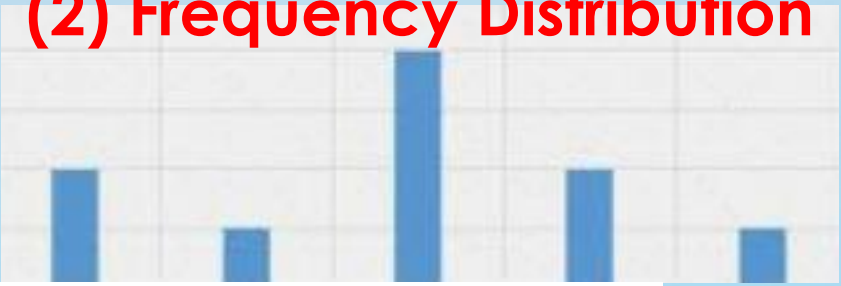
Reliability Output

Reliability Coefficients

Alpha = .8172 Standardized item alpha = .8168

Data Analysis: Example

(2) Frequency Distribution



Respondent's Department

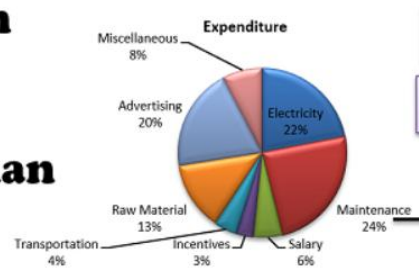
	Frequency	Percent	Valid Percent	Cumulative Percent
Marketing	13	7.5	7.5	7.5
Production	49	28.1	28.1	35.6
Sales	44	25.3	25.3	60.9
Finance	5	2.9	2.9	63.8
Servicing	34	19.5	19.5	83.3
Maintenance	5	2.9	2.9	86.2
Personnel	16	9.2	9.2	95.4
Public Relations	3	1.7	1.7	97.1
Accounting	5	2.9	2.9	100.0
Total	174	100.0	100.0	100.0

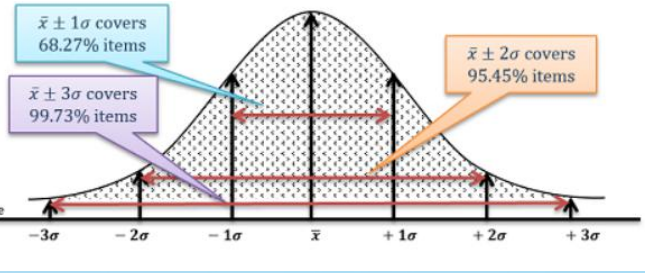
Data Analysis: Example

(3) Descriptive Analysis

Mean

Median





	N	Minimum	Maximum	Mean	Std Deviation	Variance
Dist Justice	173	1.00	5.00	2.379	.756	.570
Burnout	173	1.00	4.33	2.671	.521	.271
Job Sat	170	1.61	4.28	3.117	.507	.257
Job Char	167	2.31	4.69	3.474	.518	.268
ITL	174	1.00	4.00	2.212	.673	.453

Data Analysis: Example (Hypotheses Testing)

(4) Pearson Correlation

		Dist Justice	Burnout	Job Sat	Job Char	ITL
Pearson Correlation	Dist Justice	1.000	–.374**	.588**	.169*	–.357**
	Burnout	–.374**	1.000	–.474**	–.299**	.328**
	Job Sat	.588**	–.474**	1.000	.328**	–.535**
	Job Char	.169*	–.299**	.328**	1.000	–.274**
	ITL	–.357**	.328**	–.535**	–.274**	1.000

Data Analysis: Example (Hypotheses Testing)

H1

(5) Hypotheses Testing

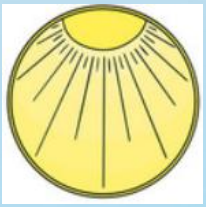
$H1_A$: Women will perceive more equity than men (or men will perceive less equity than women).

Statistically expressed: $H1_A$ is: $\mu_W > \mu_M$ **T-test**

		Levene's Test for Equality of Variance		t-test for Equality of Means						
									95% Confidence Interval of the Mean	
		F	Significance	*t*	*df*	Significance (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Dist Justice	Equal variance assumed	1.31	.352	.74	171	.461	.03	.10	.30	.91
	Equal variance not assumed		.67	29	.506	.03	.09	.29	.89	

Data Analysis: Example (Hypotheses Testing)

H2



$H2_0$: The job satisfaction of individuals will be the same irrespective of the shift they work (1, 2, or 3).

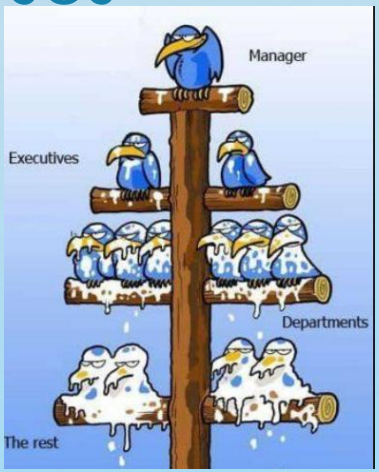
Statistically expressed, $H2_0$ is: $\mu_1 = \mu_2 = \mu_3$

		Sums of Squares	*df*	Mean Square	*F*	Sig.
Job Sat	Between Groups	1.659	2	.831	3.327	.038
	Within Groups	39.645	159	.249		
	Total	41.304	161			

0.038 < 0.05 **=> Reject NULL hypothesis**

Data Analysis: Example (Hypotheses Testing)

H3



$H3_0$: There will be no difference in the intention to leave of employees at the five different job levels.

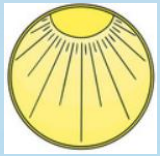
Statistically expressed, $H3_0$ is: $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$

		Sums of Squares	*df*	Mean Square	*F*	Sig.
ITL	Between Groups	2.312	4	.578	1.254	.288
	Within Groups	75.143	163	.461		
	Total	77.455	167			

0.288 > 0.05 **=> Accept NULL hypothesis**

Data Analysis: Example (Hypotheses Testing)

H4


$H4_o$: Shifts worked and employment status (part-time vs. full-time) will be independent (i.e., will not be related).

Employment Status * Shift Cross-tabulation

Employment Status	Shift			Total
	First	Second	Third	
Full Time	103	25	18	146
Part Time	16	8	4	28
Total	119	33	22	174

Chi-Square Tests

	Value	*df*	Asymp. Sig (2-sided)
Pearson Chi-square	2.312	2	.314
Likelihood ratio	2.163	2	.339
Linear-by-linear Association	1.103	1	.294
N of valid cases	174		

Test relationships between categorical var.

**** Chi-square***

Accept NULL hypothesis

Data Analysis: Example (Hypotheses Testing)

H5

$H5_0$: The four independent variables will *not* significantly explain the variance in intention to leave.

Multiple regression

Model Summary[3,4]

		Variables Entered	Variables Removed	*R*	*R* Square	Adjusted *R* Square	Std. Error of the Estimate
Model	1	Job Char Dist Just Burnout Job Sat.[1,2]		.548	.300	.282	.578

ANOVA[2]

			Sum of Squares	*df*	Mean Square	*F*	Significance
Model	1	Regression	22.366	4	5.591	16.717	.000[1]
		Residual	52.180	156	.335		
		Total	74.546	160			

[1] Indep.Vars: (constant) Job Char, Dist Justice, Burnout, Job Sat
[2] Dependent Variable: ITL

Reject NULL hypothesis

Coefficients[1]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	*t*	Sig
1	(Constant)	4.048	.603		6.713	.000
	Job Char	–.112	.095	–.084	–1.173	.243
	Dist Justice	–.115	.078	–.121	–1.461	.146
	Burnout	.143	.103	.109	1.393	.166
	Job Sat	–.498	.121	–.371	–4.121	.000

[1] Dependent Variable: INTENTION TO LEAVE