Statistics for Business and decision Making

Dr. Ilaria Benedetti

Analysis of Variance

Socio-Economic Planning Sciences 89 (2023) 101701

Contents lists available at ScienceDirect

Socio-Economic Planning Sciences

journal homepage: www.elsevier.com/locate/seps

Digitalization in Europe: A potential driver of energy efficiency for the twin transition policy strategy

Ilaria Benedetti, Giulio Guarini *, Tiziana Laureti

University of Tuscia, Department of Economics, Engineering, Society and Business Organization, Viterbo, Italy

A R T I C L E I N F O

JEL classification:
L86
O33
O44
Q43

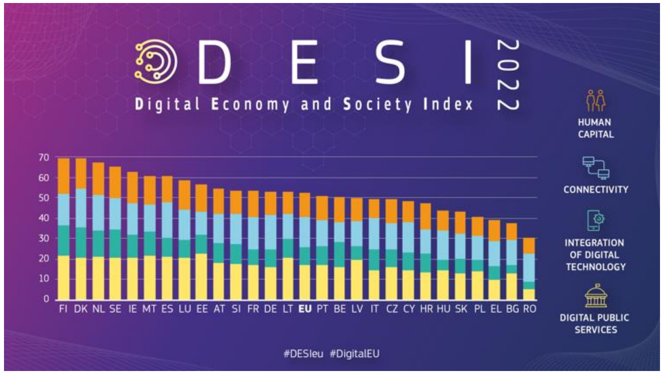
Keywords:
Twin transition
Digital transformation
Energy efficiency
DESI

A B S T R A C T

This paper aims to study the impact of digitalization on energy efficiency in Europe, where institutions are committed to promote both digital transformation and energy transition by implementing the European Twin Transition, ETT [1]. The paper starts by empirically analysing the state of play of the Twin Transition across European Member States (MS) by mapping the Digital Economy and Society Index (DESI), its sub-indices and the energy productivity. We then discuss the impact of the abovementioned digital indicators on energy productivity growth in 26 European MS during the period 2016–2020 by applying a system GMM model. The results show a significant and positive impact of digitalization on energy efficiency and relevant complementarities across diverse digitalization dimensions. This study contributes in an original manner to the literature by applying for the first time DESI and its sub-indices in a European twin transition analysis; moreover, empirical insights have

$$dEPROD_{it} = \delta_0 + \delta_1 dEPROD_{it-1} + \delta_2 EPROD_{GAP_{it}} + \delta_3 dPROD_{it} + \delta_4 HUM\_CAP_{it} + \sum_{t=1}^{t} \mu_t \rho_t + \varphi_{it}$$

$$dEPROD_{it} = \theta_0 + \theta_1 dEPROD_{it-1} + \theta_2 EPROD_{GAP_{it}} + \theta_3 dPROD_{it} + \theta_4 INT\_DIG\_TECH_{it} + \sum_{t=1}^{t} \mu_t \rho_t + \varphi_{it}$$



https://digital-strategy.ec.europa.eu/en/policies/desi

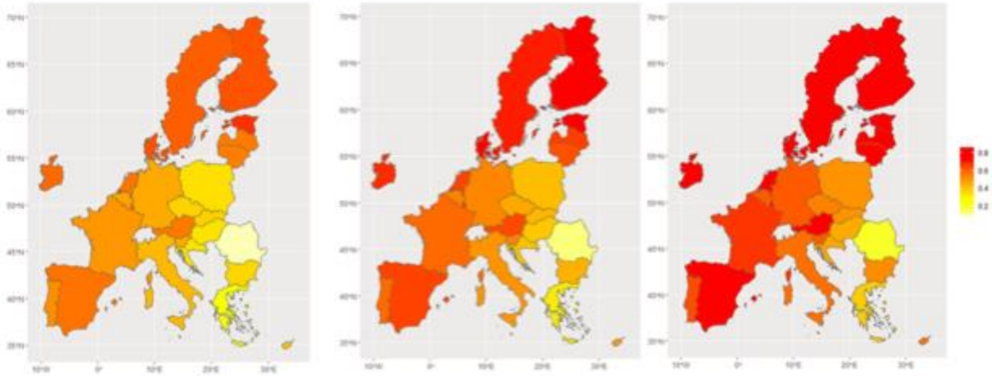


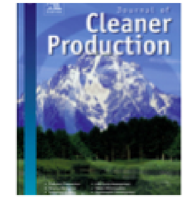
Fig. 3. DESI DIG_PUB_SERV, years 2016 (left), 2018 (middle) and 2020 (right).
Source: Author's elaboration from Eurostat data

ELSEVIER

Contents lists available at ScienceDirect

Journal of Cleaner Production

journal homepage: www.elsevier.com/locate/jclepro



Exploring pro-environmental food purchasing behaviour: An empirical analysis of Italian consumers

Tiziana Laureti [a, *], Ilaria Benedetti [b]

[a] Department of Economics, Engineering, Society and Business Organization, University of Tuscia, Via del Paradiso, 47, 01100 Viterbo, Italy
[b] Department of Economics and Legal Studies, University of Naples "Parthenope", Via Generale Parisi, 13, 80132 Napoli, Italy

A R T I C L E I N F O

Article history:
Received 30 June 2017
Received in revised form
7 October 2017
Accepted 11 November 2017
Available online 16 November 2017

Keywords:
Organic food
Sustainable consumption
Environmental attitudes
Environmental concerns
Theory of planned behaviour
Multilevel models

A B S T R A C T

Over the last decades, there has been growing consumers' interest in environmentally friendly products and green agriculture has experienced a steady increase in European countries. Although the purchasing behaviour of consumers is closely linked to their personal needs, environmental preservation has also become one of the main concerns since food production exerts significant pressures on the environment, especially through water, energy, pesticide and fertiliser use.

Environmentally-friendly purchasing decisions can reduce the environmental impact of food by substituting higher-impact products with "green" products which do not pollute the planet thus preserving the environment and public health, but bringing also significant benefits to the economy as a whole.

Therefore, the aim of this study is to analyse the factors influencing consumers' decision to buy green products by suggesting a two-level conceptual framework, based on the theory of planned behaviour and adding the context in which individuals reside.

We estimate two different type of ordered multilevel random effects models in order to examine whether and to what extent differences in individuals' behaviour concerning organic food consumption among Italian regions can be attributed to the socio-economic and environmental characteristics of the area in which the individuals reside. By using the 2014 Aspect of Daily Life Survey carried out by the Italian National Statistical Institute, we obtained interesting results concerning the role of individuals' environmental concern and attitudes. People concerned with animal welfare, soil pollution and deforestation have a higher probability of buying organic products on a daily basis. The results obtained from the random slope model specification, which allowed us to analyse if and to what extent the effect of awareness for soil pollution varies across regions, confirm that the across regions awareness for soil pollution significantly and positively influence individuals' consumption of organic food on a daily basis.

economies

Submit to this Journal

Review for this Journal

Propose a Special Issue

Article Menu

Academic Editor

Andreia Dionísio

Subscribe SciFeed

Recommended Articles

Related Info Link

More by Authors Links

Order Article Reprints

Open Access Article

Computation of High-Frequency Sub-National Spatial Consumer Price Indexes Using Web Scraping Techniques

by Ilaria Benedetti [1,*], Tiziana Laureti [1], Luigi Palumbo [1] and Brandon M. Rose [2]

1 Department of Economics, Engineering, Society and Business Organization, University of Tuscia, Via del Paradiso, 01100 Viterbo, Italy

2 Jataware LLC, Washington, DC 20015, USA

* Author to whom correspondence should be addressed.

Economies **2022**, *10*(4), 95; https://doi.org/10.3390/economies10040095

Received: 15 February 2022 / Revised: 6 April 2022 / Accepted: 10 April 2022 / Published: 14 April 2022

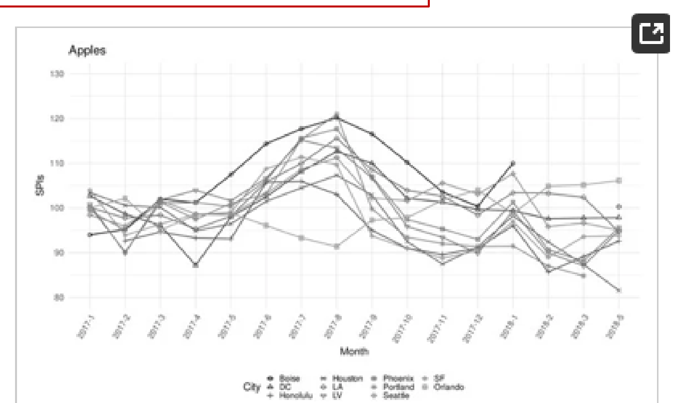
Download Browse Figures Versions Notes

Abstract

The development of Information and Communications Technology and digital economies has contributed to changes in the consumption of goods and services in various areas of life, affecting the growing expectations of users in relation to price statistics. Therefore, it is important to provide information on differences in consumer prices across space and over time in a timely manner. Web-scraped data, which is the process of collecting large amounts of data from the web, offer the potential to improve greatly the quality and efficiency of consumer price indices. In this paper, we explore the use of web-scraped data for compiling high-frequency price indexes for groups of products by using the time-interaction-region product model. We computed monthly average prices for five entry-level items according to the Consumer Price Index for All Urban Consumers (CPI-U) classification and tracked their evolution over time in 11 USA cities reported in our dataset. Even if our dataset covers a small percentage of the CPI-U index, results show how web scraping data may provide timely estimates of sub-national SPI evolution and unveil seasonal trends for specific categories.

Keywords: consumer spatial price indexes; data scraping; spatial index; time comparison; big data

$$lnP_{ijt} = \sum_{i=1}^{N} \beta_i D_{ijt} + \sum_{t=1}^{T} \sum_{j=1}^{M} \delta_{jt} C_{ij} T_{jt} + \upsilon_{ijt}$$



One-Way Analysis of Variance

- Evaluate the difference among the means of three or more groups

Examples: Average production for 1st, 2nd, and 3rd shifts
Expected mileage for five brands of tires

- Assumptions
  - Populations are normally distributed
  - Populations have equal variances
  - Samples are randomly and independently drawn

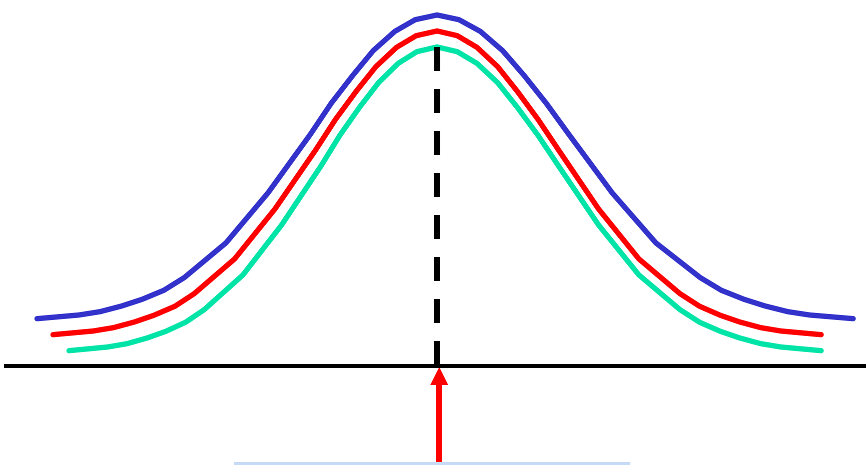
Hypotheses of One-Way ANOVA

- $H_0 : \mu_1 = \mu_2 = \mu_3 = \cdots = \mu_K$
  - All population means are equal
  - i.e., no variation in means between groups
- $H_1 : \mu_i \neq \mu_j$ for at least one i, j pair
  - At least one population mean is different
  - i.e., there is variation between groups
  - Does not mean that all population means are different (some pairs may be the same)

One-Way ANOVA

$$H_0: \mu_1 = \mu_2 = \mu_3 = \cdots = \mu_K$$

$$H_1: \text{Not all } \mu_i \text{ are the same}$$



$$\mu_1 = \mu_2 = \mu_3$$

All Means are the same:
The Null Hypothesis is True
(No variation between groups)

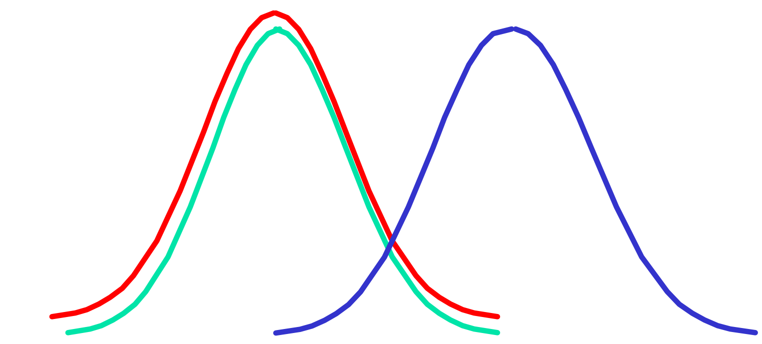
One-Way ANOVA

(continued)

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \cdots = \mu_K$$

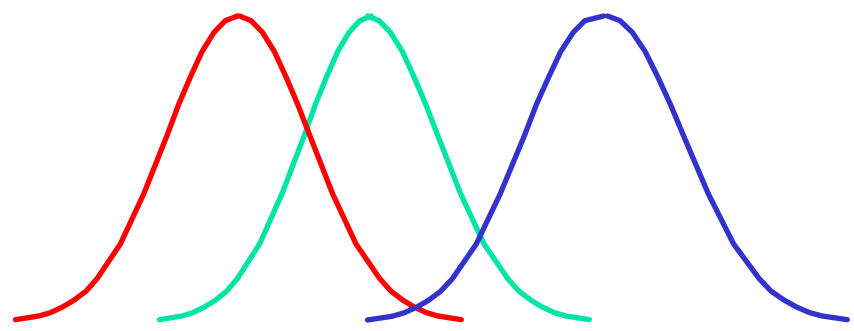
$$H_1 : \text{Not all } \mu_i \text{ are the same}$$

At least one mean is different:
The Null Hypothesis is NOT true
(Variation is present between groups)



$$\mu_1 = \mu_2 \neq \mu_3$$

or

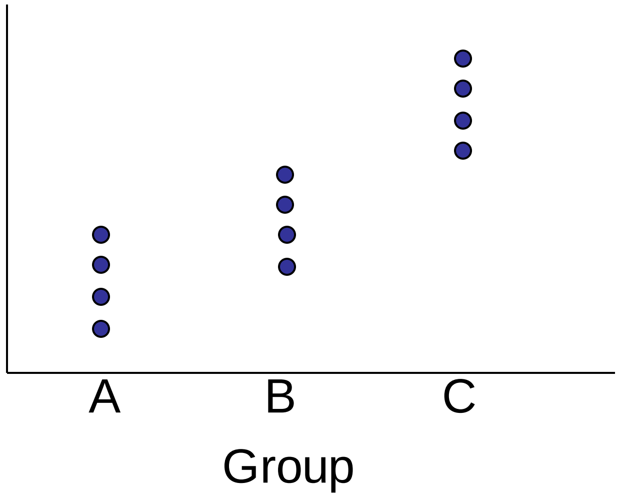


$$\mu_1 \neq \mu_2 \neq \mu_3$$

Variability

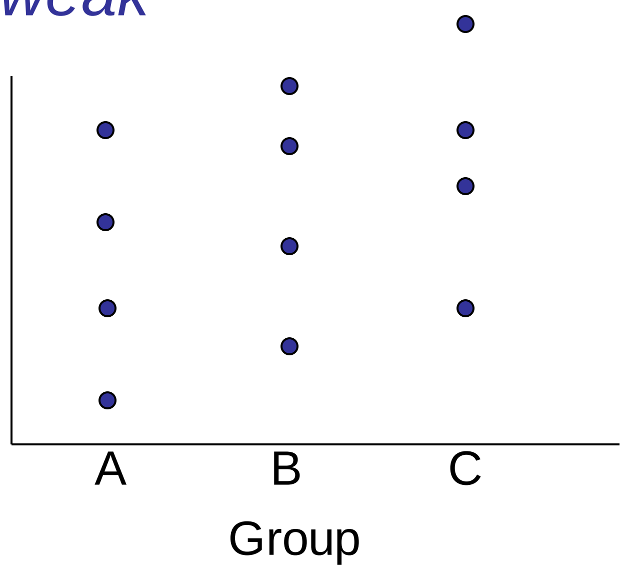
- The variability of the data is key factor to test the equality of means
- In each case below, the means may look different, but a large variation within groups in B makes the evidence that the means are different *weak*

A



Small variation within groups

B



Large variation within groups

Partitioning the Variation

- Total variation can be split into two parts:

SST = SSW + SSB

SST = Total Sum of Squares

Total Variation = the aggregate dispersion of the individual data values across the various groups

SSW = Sum of Squares Within Groups

Within-Group Variation = dispersion that exists among the data values within a particular group

SSB = Sum of Squares Between Groups

Between-Group Variation = dispersion between the group sample means

Partition of Total Variation

Total Sum of Squares (SST)

= **Variation due to random sampling (SSW)** + **Variation due to differences between groups (SSB)**

Total Sum of Squares

$$SST = SSW + SSB$$

$$SST = \sum_{i=1}^{K} \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2$$

Where:

SST = Total sum of squares

K = number of groups (levels or treatments)

n_i = number of observations in group i

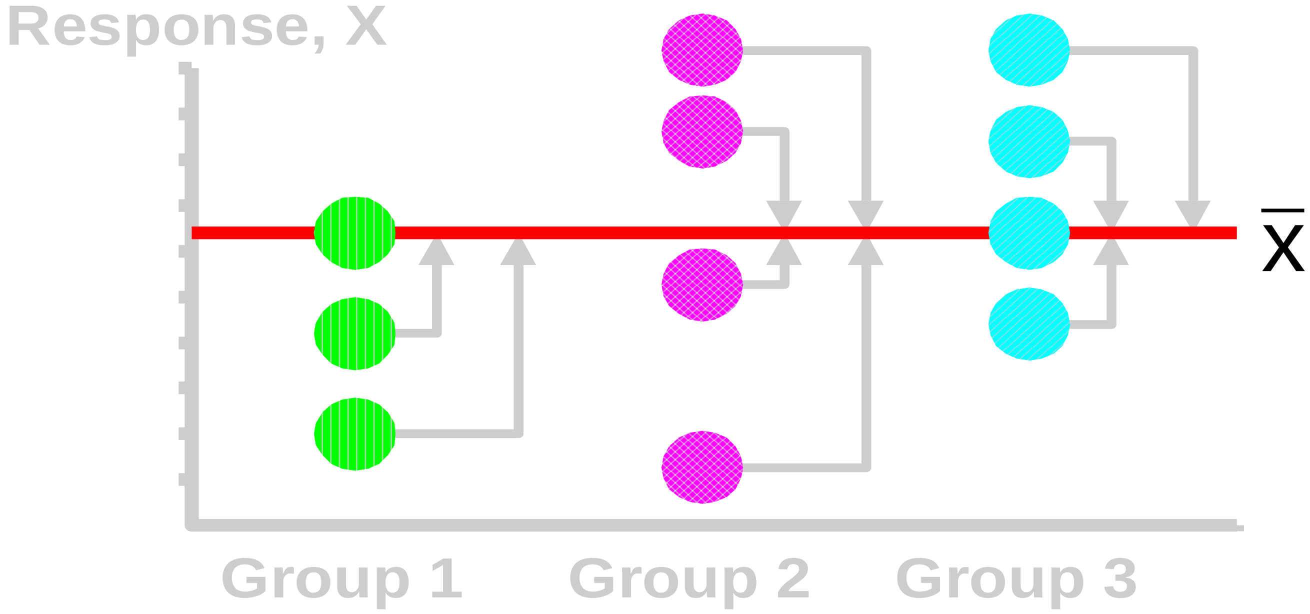
x_{ij} = j^{th} observation from group i

$\bar{x}$ = overall sample mean

Total Variation

(continued)

$$SST = (x_{11} - \bar{x})^2 + (X_{12} - \bar{x})^2 + ... + (x_{Kn_K} - \bar{x})^2$$



Within-Group Variation

$$SST = \boxed{SSW} + SSB$$

$$SSW = \sum_{i=1}^{K} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2$$

Where:

SSW = Sum of squares within groups

K = number of groups

n_i = sample size from group i

$\bar{x}_i$ = sample mean from group i

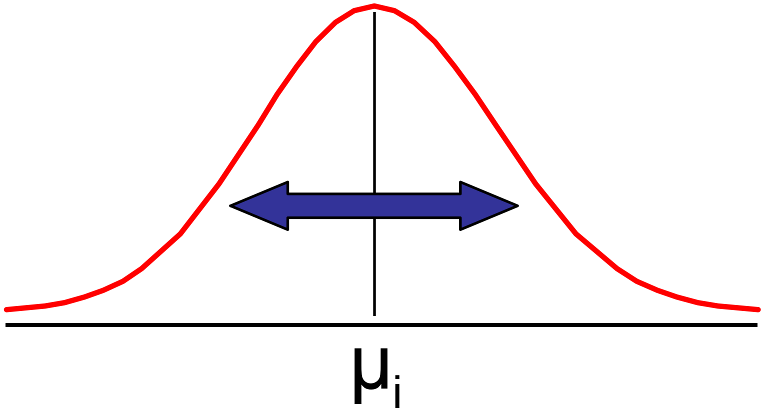
x_{ij} = j^{th} observation in group i

Within-Group Variation

(continued)

$$SSW = \sum_{i=1}^{K}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)^2$$

Summing the variation within each group and then adding over all groups



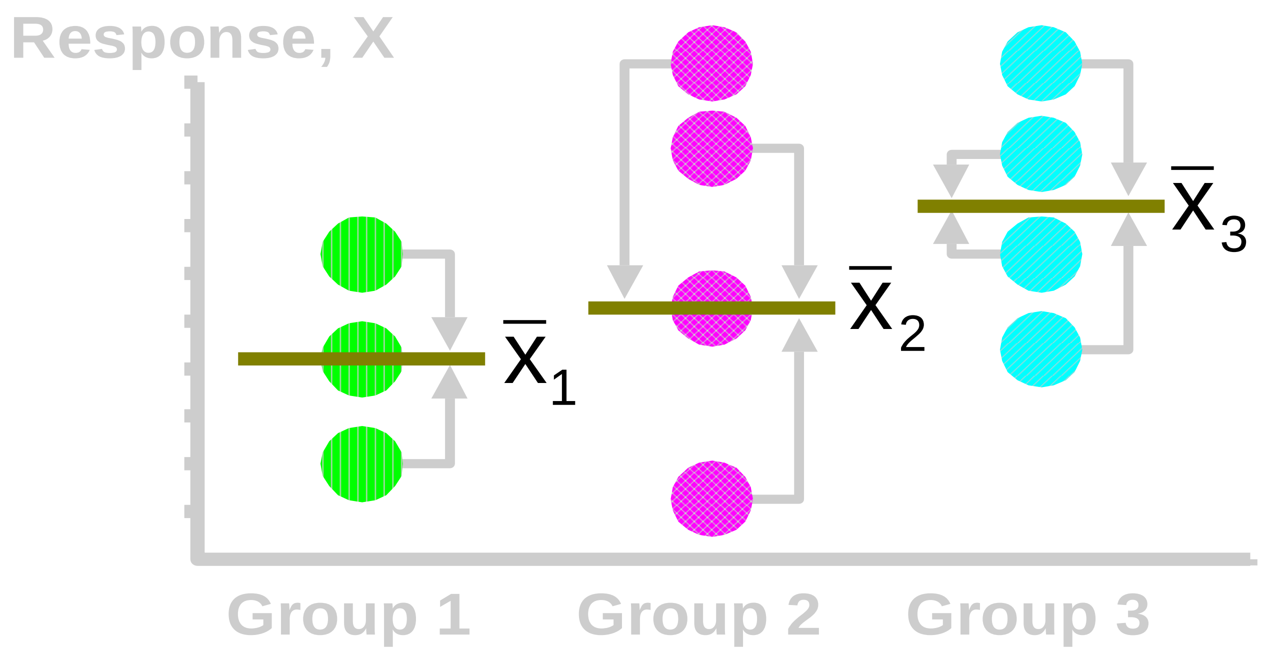
$$MSW = \frac{SSW}{n - K}$$

Mean Square Within = SSW/degrees of freedom

Within-Group Variation

(continued)

$$SSW = (x_{11} - \bar{x}_1)^2 + (x_{12} - \bar{x}_1)^2 + \ldots + (x_{Kn_K} - \bar{x}_K)^2$$



Between-Group Variation

$$SST = SSW + \boxed{SSB}$$

$$SSB = \sum_{i=1}^{K} n_i(\bar{x}_i - \bar{x})^2$$

Where:

SSB = Sum of squares between groups

K = number of groups

n_i = sample size from group i

$\bar{x}_i$ = sample mean from group i

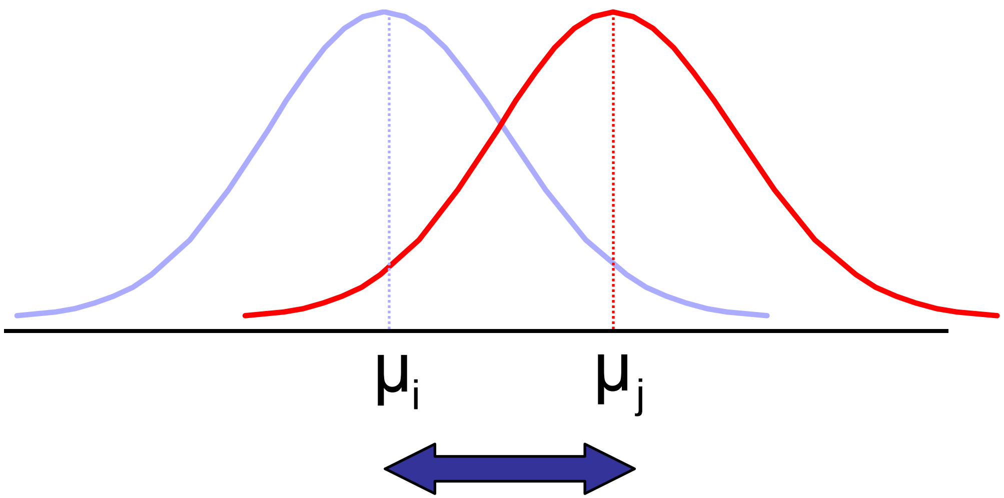
$\bar{x}$ = grand mean (mean of all data values)

Between-Group Variation

(continued)

$$SSB = \sum_{i=1}^{K} n_i(\bar{x}_i - \bar{x})^2$$

Variation Due to
Differences Between Groups



$$\boldsymbol{MSB = \frac{SSB}{K-1}}$$

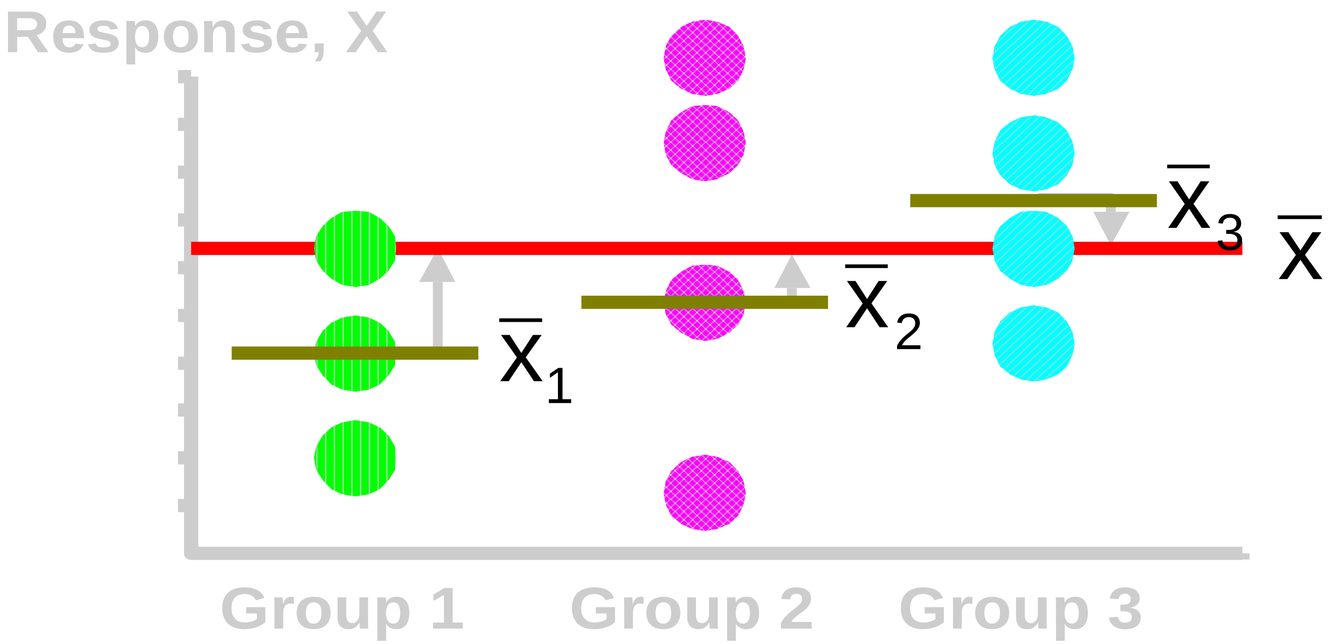
Mean Square Between Groups

= SSG/degrees of freedom

Between-Group Variation

(continued)

$$SSB = n_1(\bar{x}_1 - \bar{x})^2 + n_2(\bar{x}_2 - \bar{x})^2 + \ldots + n_K(\bar{x}_K - \bar{x})^2$$



Obtaining the Mean Squares

$$\text{MST} = \frac{\text{SST}}{n-1}$$

$$\text{MSW} = \frac{\text{SSW}}{n-K}$$

$$\text{MSB} = \frac{\text{SSB}}{K-1}$$

One-Way ANOVA Table

Source of Variation	SS	df	MS (Variance)	F ratio
Between Groups	SSB	K - 1	$MSB = \frac{SSB}{K - 1}$	$F = \frac{MSB}{MSW}$
Within Groups	SSW	n - K	$MSW = \frac{SSW}{n - K}$	
Total	SST = SSG+SSW	n - 1		

K = number of groups

n = sum of the sample sizes from all groups

df = degrees of freedom

One-Factor ANOVA F Test Statistic

$H_0: \mu_1 = \mu_2 = \ldots = \mu_K$

H_1: At least two population means are different

- Test statistic

$$F = \frac{MSB}{MSW}$$

MSB is mean squares between variances

MSW is mean squares within variances

- Degrees of freedom
  - $df_1 = K - 1$ (K = number of groups)
  - $df_2 = n - K$ (n = sum of sample sizes from all groups)

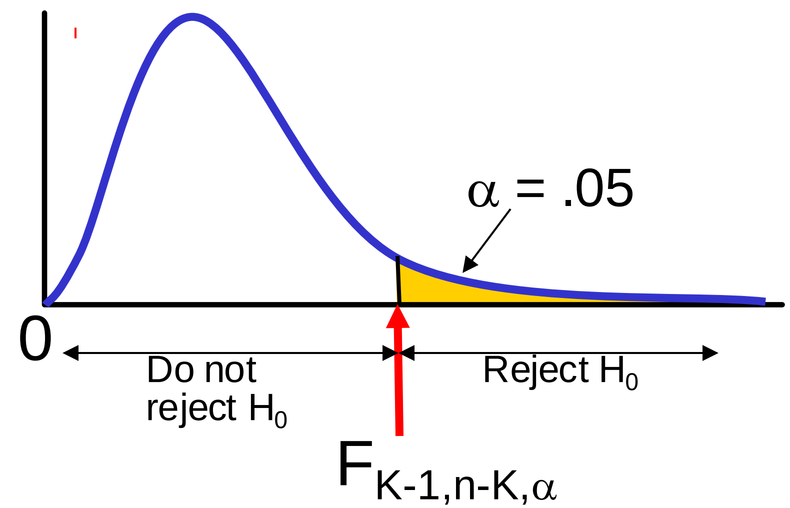
Interpreting the F Statistic

- The F statistic is the ratio of the between estimate of variance and the within estimate of variance
  - The ratio must always be positive
  - $df_1 = K - 1$ will typically be small
  - $df_2 = n - K$ will typically be large

Decision Rule:

- Reject H_0 if

  $F > F_{K-1,n-K,\alpha}$



One-Factor ANOVA F Test Example

You want to see if three different golf clubs yield different distances. You randomly select five measurements from trials on an automated driving machine for each club. At the .05 significance level, is there a difference in mean distance?

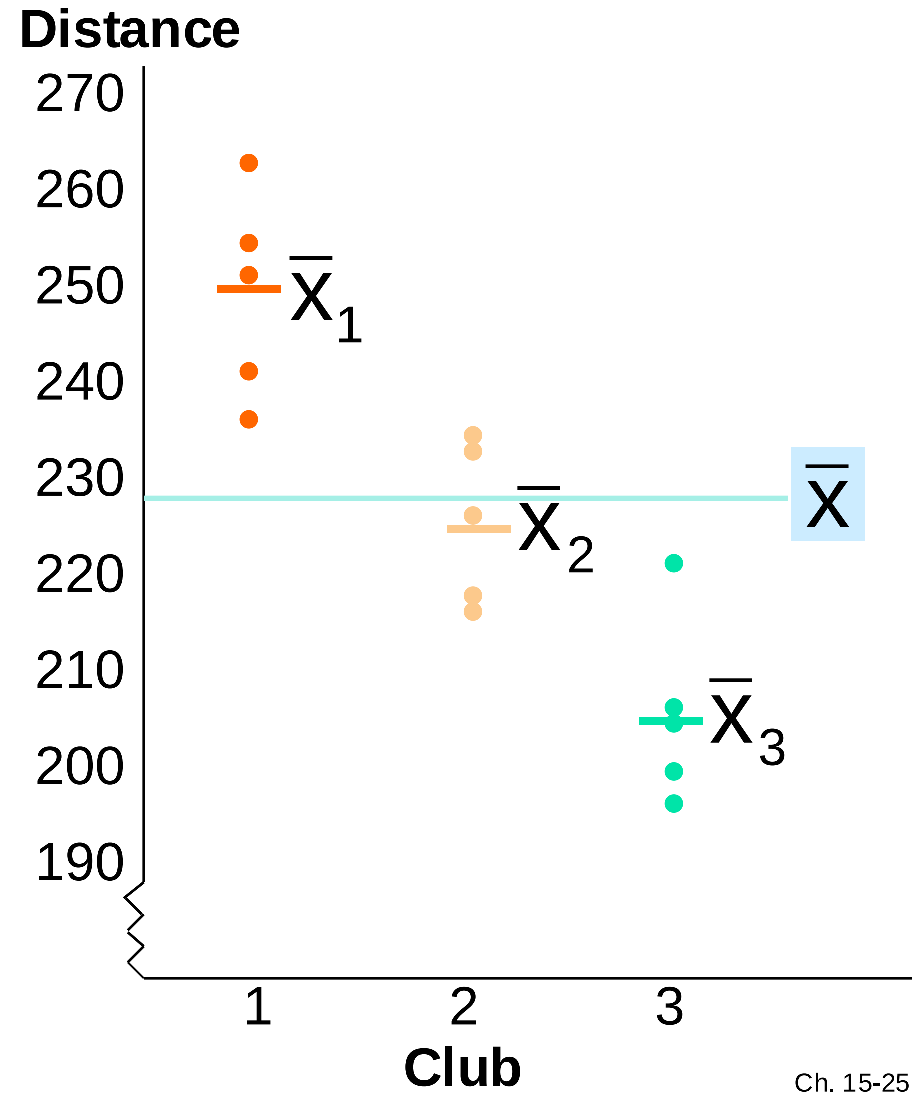
Club 1	Club 2	Club 3
254	234	200
263	218	222
241	235	197
237	227	206
251	216	204

One-Factor ANOVA Example: Scatter Diagram

Club 1	Club 2	Club 3
254	234	200
263	218	222
241	235	197
237	227	206
251	216	204

$\bar{x}_1 = 249.2$	$\bar{x}_2 = 226.0$	$\bar{x}_3 = 205.8$

$\bar{\bar{x}} = 227.0$



One-Factor ANOVA Example Computations

Club 1	**Club 2**	**Club 3**
254	234	200
263	218	222
241	235	197
237	227	206
251	216	204

$\bar{x}_1 = 249.2$ $n_1 = 5$

$\bar{x}_2 = 226.0$ $n_2 = 5$

$\bar{x}_3 = 205.8$ $n_3 = 5$

$n = 15$

$\bar{x} = 227.0$

$K = 3$

$SSB = 5\,(249.2 - 227)^2 + 5\,(226 - 227)^2 + 5\,(205.8 - 227)^2 = 4716.4$

$SSW = (254 - 249.2)^2 + (263 - 249.2)^2 + \ldots + (204 - 205.8)^2 = 1119.6$

$MSB = 4716.4 \,/\, (3\text{-}1) = 2358.2$

$MSW = 1119.6 \,/\, (15\text{-}3) = 93.3$

$$F = \frac{2358.2}{93.3} = 25.275$$

One-Factor ANOVA Example Solution

$H_0: \mu_1 = \mu_2 = \mu_3$

$H_1: \mu_i$ not all equal

$\alpha = .05$

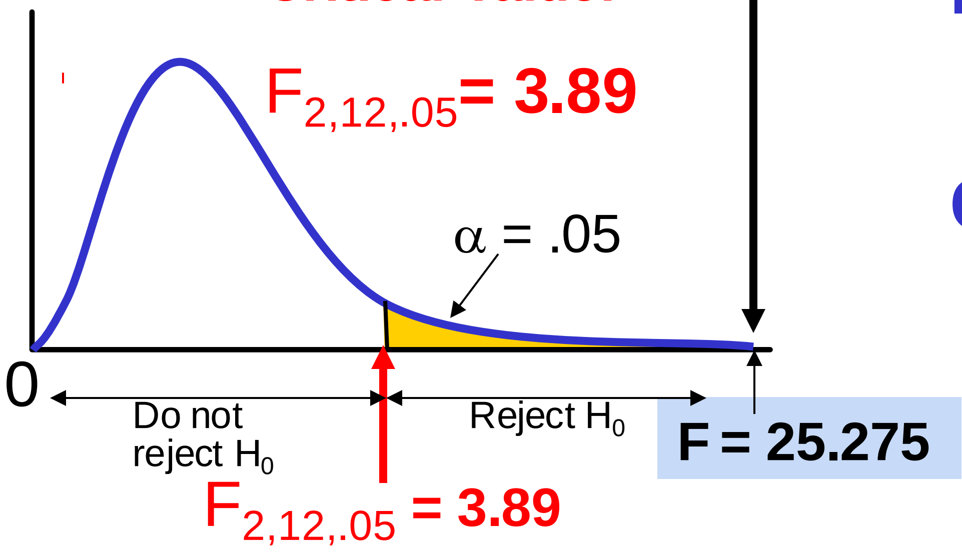
$df_1 = 2 \qquad df_2 = 12$

Test Statistic:

$$F = \frac{MSB}{MSW} = \frac{2358.2}{93.3} = 25.275$$

Critical Value:

$F_{2,12,.05}$ **= 3.89**



Decision:

Reject H_0 at $\alpha = 0.05$

Conclusion:

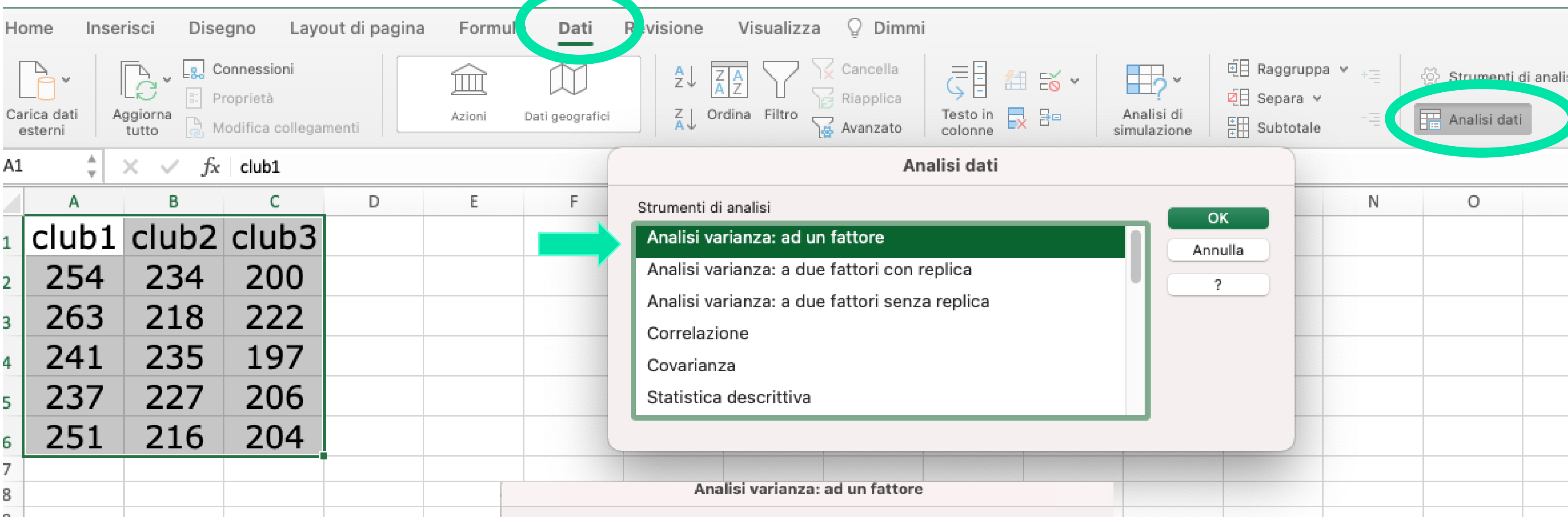
There is evidence that at least one μ_i differs from the rest

ANOVA -- Single Factor: Excel Output

EXCEL: data | data analysis | ANOVA: single factor

SUMMARY						
Groups	*Count*	*Sum*	*Average*	*Variance*		
Club 1	5	1246	249.2	108.2		
Club 2	5	1130	226	77.5		
Club 3	5	1029	205.8	94.2		
ANOVA						
Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	**4716.4**	**2**	**2358.2**	**25.275**	4.99E-05	**3.89**
Within Groups	**1119.6**	**12**	**93.3**			
Total	5836.0	14				

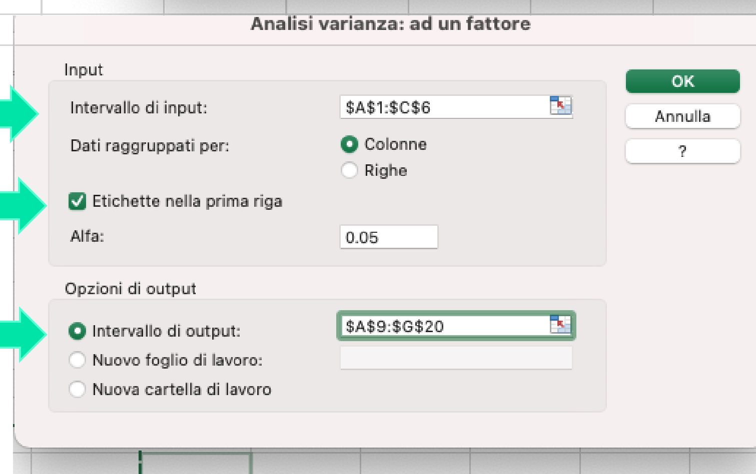
ANOVA -- Single Factor: Excel Output



Selezione delle unità

Etichette

Dove si vuole l'output



ANOVA -- Single Factor: Excel Output

Analisi varianza: ad un fattore

RIEPILOGO

Gruppi	*Conteggio*	*Somma*	*Media*	*Varianza*
club1	5	1246	249.2	108.2
club2	5	1130	226	77.5
club3	5	1029	205.8	94.2

ANALISI VARIANZA

Origine della variazione	*SQ*	*gdl*	*MQ*	*F*	*Valore di significatività*	*F crit*
Tra gruppi	4716.4	2	2358.2	25.275	0.000	3.885
In gruppi	1119.6	12	93.3			
Totale	5836	14				

Examples with R

- **Exercise 1** - Open dataset PlantGrowth (included in R):

1. compute some summary statistics;
2. compute mean of weight by group;
3. represent graphically variable weight by group with a box plot;
4. compute ANOVA. Provide an interpretation of the obtained result.

- **Exercise 2** - Open dataset Iris (included in R):

1. compute some summary statistics;
2. compute mean of Petal.Length by Species;
3. compute ANOVA of Petal.Length by Species. Provide an interpretation of the obtained result.

Exercise to do at home

Let's consider the following table which shows the information relating to the result of a test conducted with three types of different treatments:

	Treatment		
	A	B	C
	162	142	126
	142	156	122
	165	124	138
	145	142	140
	148	136	150
	174	152	128
Sampling mean	156	142	134
Sampling variance	164.4	131.2	110.4

Compute:

1. the deviance between treatments;
2. variance between treatments;
3. the deviance within treatments;
4. the variance within treatments;
5. The ANOVA table for this problem
6. For a significance level of α=0.05, check whether the means for the three treatments are equal.

Try also to solve the exercise with Excel and R.