Statistics for Medicine

Massimo Borelli

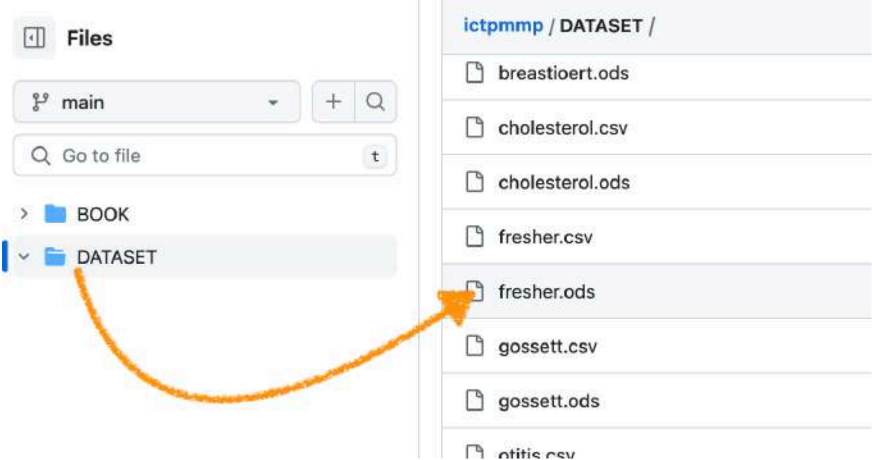
Master of Advanced Studies in Medical Physics

UNIVERSITÀ DEGLI STUDI DI TRIESTE
The Abdus Salam International Centre for Theoretical Physics
IAEA

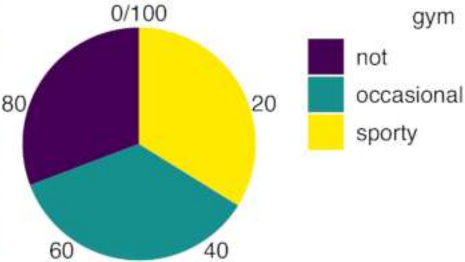
Recap

III **Inferential Statistics**

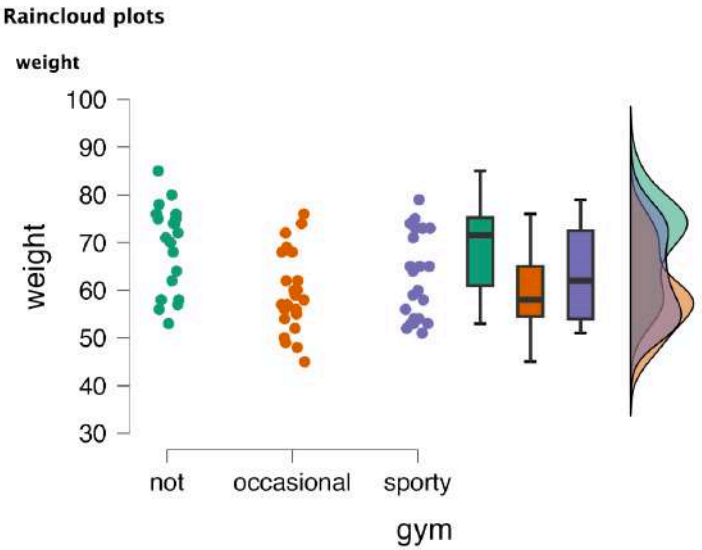
the fresher dataset



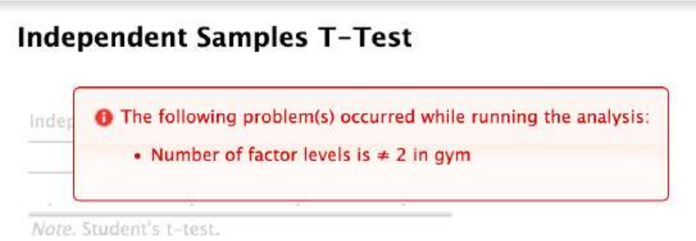
Does gym activity predict weight? /1



Does gym activity predict weight? /2

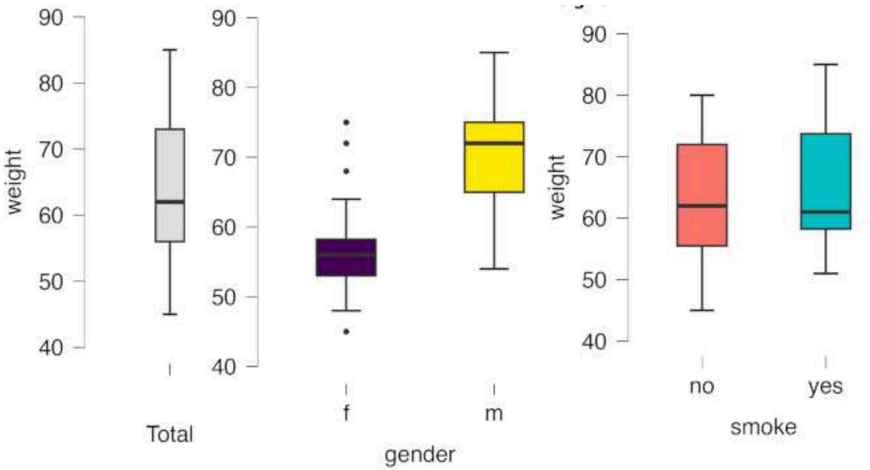


oh no!



$$t = \frac{m_1 - m_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

key idea



	weight
Mean	63.523
Variance	92.128

	weight
Mean	63.523
Variance	92.128

	weight	
gender	f	m
Mean	56.625	70.212
Variance	41.081	50.735

	weight	
smoke	no	yes
Mean	63.098	65.071
Variance	89.410	106.379

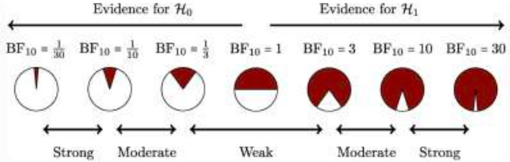
1. the one-way Anova

- consider as a Fixed Factor the gym physical activity
  - ordered: not < occasional < sporty
- as Dependent Variable the weight

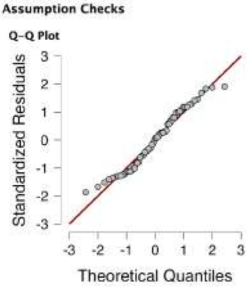
the one-way Anova

Cases	Sum of Squares	df	Mean Square	F	p
gym	1020.400	2	510.200	6.488	**0.003**
Residuals	4875.816	62	78.642		

Models	P(M)	P(M|data)	BF_M	BF_{10}	error %
gym	0.500	0.933	**14.018**	1.000	
Null model	0.500	0.067	0.071	0.071	0.015



Normality assumptions



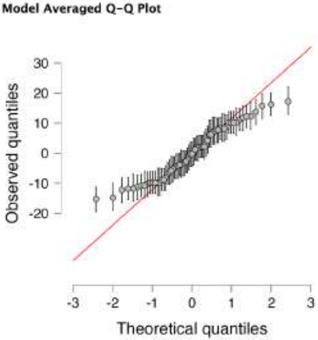
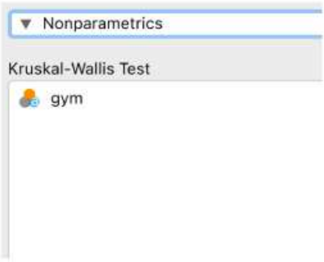


Table: Equality of Variances (Levene's)

F	df1	df2	p
0.470	2.000	62.000	0.627

Nonparametric one-way Anova

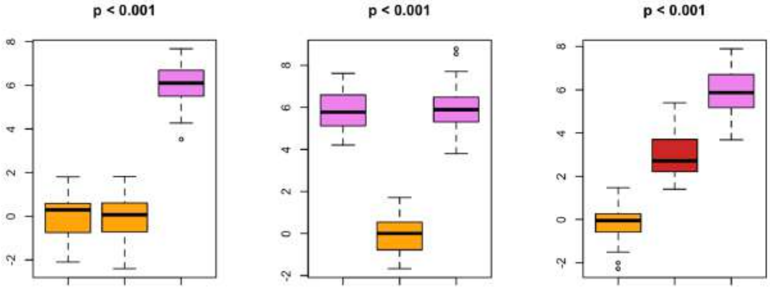


the multiple comparison issue /1

Consider the following simulations:

$$a = \mathrm{rnorm}(n = 30, \mathrm{mean} = 0, \mathrm{sd} = 1)$$
$$b = \mathrm{rnorm}(n = 30, \mathrm{mean} = 0, \mathrm{sd} = 1)$$
$$c = \mathrm{rnorm}(n = 30, \mathrm{mean} = 6, \mathrm{sd} = 1)$$

the multiple comparison issue /2



the multiple comparison issue /3

bad idea: to make multiple t-tests on all possible pairs of means

$\alpha = 0.05$

$$1 - (1 - \frac{5}{100}) \cdot (1 - \frac{5}{100}) \cdot (1 - \frac{5}{100}) =$$

$$= 1 - (1 - \frac{5}{100})^3 = 0.143$$

multiple tests increases the probability of a type-I error.

Carlo Bonferroni 'radical' solution (Bernoulli $1 + nh < (1 + h)^n$)
$n = 3$ groups, $n \cdot (n - 1)/2 = 3$ comparisons, then
$h = \alpha/3 = 0.05/3 = 0.017$.

Table: Post Hoc Comparisons - gym

		Mean Difference	SE	t	p_{tukey}
not	occasional	9.665	2.711	3.565	0.002
	sporty	6.373	2.740	2.326	0.060
occasional	sporty	−3.292	2.645	−1.245	0.432

Table: Post Hoc Comparisons - gym

		Prior Odds	Postrr Odds	$BF_{10,U}$	error %
not	occsnl	0.587	21.169	36.039	4.422×10^{-7}
	sporty	0.587	1.313	2.236	0.008
occsnl	sporty	0.587	0.328	0.558	0.006

ALERT: very difficult with JASP

the `tooth.ods` dataset

? heteroskedasticity ?

Can you detect wether `il1b` is a predictor of `areainfl` ?

explore one-way anova and look to raincloud plot

Statistics for Medicine

Massimo Borelli

Master of Advanced Studies in Medical Physics

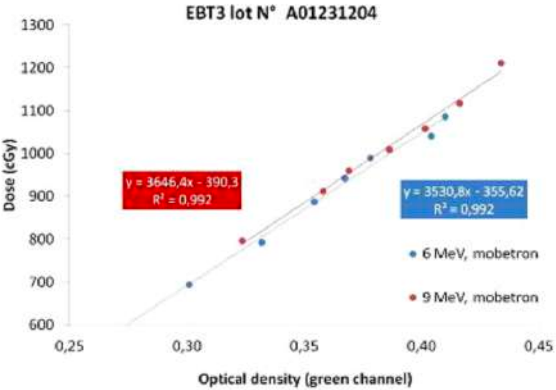
UNIVERSITÀ DEGLI STUDI DI TRIESTE

ICTP The Abdus Salam International Centre for Theoretical Physics

IAEA

In vivo dosimetry and shielding disk alignment verification by EBT3 GAFCHROMIC film in breast IOERT treatment

Mara Severgnini,[1a] Mario de Denaro,[1] Marina Bortul,[2] Cristiana Vidali,[3] Aulo Beorchia[3]
Department of Medical Physics;[1] Department of Surgery;[2] Department of Radiation Oncology;[3] A.O.U. Ospedali Riuniti, Trieste, Italy



a calibration procedure

246 *Anthropological Miscellanea.*

ANTHROPOLOGICAL MISCELLANEA.

REGRESSION *towards* MEDIOCRITY *in* HEREDITARY STATURE.

By FRANCIS GALTON, F.R.S., &c.

[WITH PLATES IX AND X.]

THIS memoir contains the data upon which the remarks on the Law of Regression were founded, that I made in my Presidential Address to Section H, at Aberdeen. That address, which will appear in due course in the Journal of the British Association, has already been published in "Nature," September 24th. I reproduce here

the fresher.ods dataset

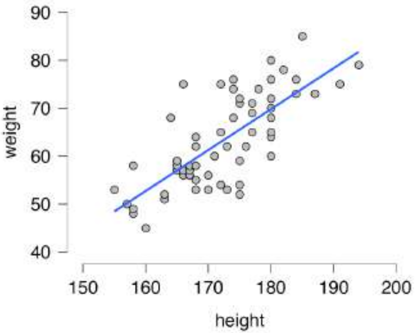


Table: Pearson's Correlations

Variable		height	...
weight	Pearson's r	0.744	...
	p-value	$< .001$	...

Table: classical linear regression - Fixed effects

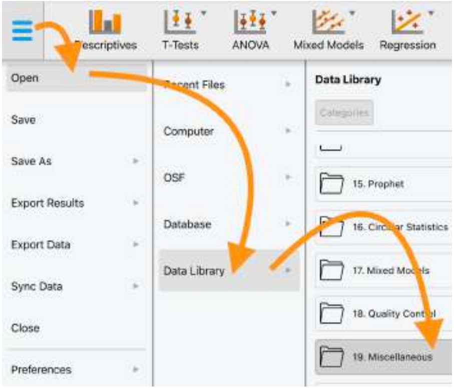
Model		Unstndr	Std Err	Stndrd	t	p
H_0	(Intercept)	63.523	1.191		53.357	$< .001$
H_1	(Intercept)	−83.891	16.677		−5.030	$< .001$
	height	0.854	0.096	0.744	8.850	$<$ **.001**

Table: Random component

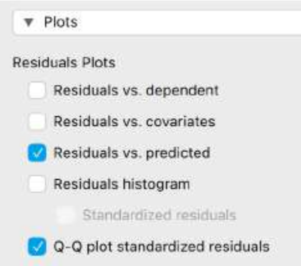
Model	R	R^2	Adjusted R^2	RMSE
H_0	0.000	0.000	0.000	9.598
H_1	0.744	0.554	0.547	**6.459**

the Anscombe's Quartet

$$y = 0.5x + 3 \; ; \; R^2 = \frac{2}{3}$$



diagnostic tools




United States
Environmental Protection
Agency

Office of Environmental
Information
Washington, DC 20460

EPA/240/B-06/003
February 2006

EPA

Data Quality Assessment: Statistical Methods for Practitioners

EPA QA/G-9S