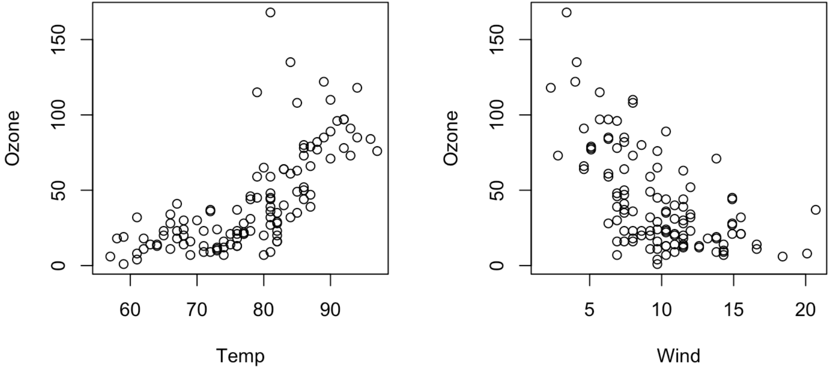
P5.2 Statistics for Medicine

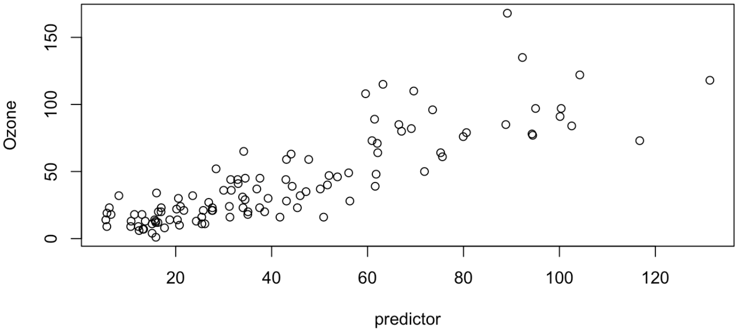
Massimo Borelli

Master of Advanced Studies in Medical Physics

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

the airquality dataset



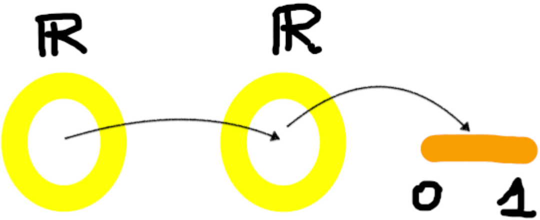



```
Call:
lm(formula = Ozone ~ Solar.R * Temp + I(Temp^2) + Wind + I(Wind^2))

Coefficients:
 (Intercept)      Solar.R         Temp    I(Temp^2)         Wind    I(Wind^2)  Solar.R:Temp
  262.475740    -0.254119    -4.898987     0.036442   -13.029708     0.445797      0.004358
```

the roma dataset

logHE4	logCA125	logCA19-9	logCEA	AgePatient	Menopause	Histology
3.58	4.25	3.33	0.22	34	ante	benign
3.42	5.45	4.84	0.24	21	ante	benign
5.68	4.72	3.20	0.92	64	post	**malignant**
4.14	3.96	3.54	1.76	58	post	**malignant**
3.57	3.03	-0.04	1.03	74	post	benign
3.70	4.11	3.44	0.58	40	ante	benign
7.17	7.58	2.45	0.44	51	ante	**malignant**
3.57	2.48	1.46	0.10	21	ante	benign
3.97	3.64	2.30	0.14	27	ante	benign
4.11	4.03	4.73	0.82	75	post	**malignant**
3.56	4.59	4.17	0.34	37	ante	benign
3.65	2.83	3.00	0.71	30	ante	benign
3.89	5.72	2.42	0.44	71	post	**malignant**



the roma dataset

Logit

From Wikipedia, the free encyclopedia

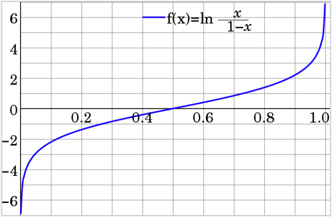
This article discusses the binary logit function only. See discrete choice for a discussion of multinomial logit, conditional logit, nested logit, mixed logit, exploded logit, and ordered logit. For the basic regression technique that uses the logit function, see logistic regression. For standard magnitudes combined by multiplication, see logit (unit).

In statistics, the **logit** (/ˈloʊdʒɪt/ *LOH-jit*) function is the quantile function associated with the standard logistic distribution. It has many uses in data analysis and machine learning, especially in data transformations.

Mathematically, the logit is the inverse of the standard logistic function $\sigma(x) = 1/(1+e^{-x})$, so the logit is defined as

$$\operatorname{logit}(p) = \sigma^{-1}(p) = \ln\left(\frac{p}{1-p}\right) \quad \text{for} \quad p \in (0,1).$$

Because of this, the logit is also called the **log-odds** since it is equal to the logarithm of the odds $\frac{p}{1-p}$ where p is a probability. Thus, the logit is a type of function that maps probability values from $(0,1)$ to real numbers in $(-\infty, +\infty)$



Plot of logit(p) in the domain of 0 to 1, where the base of logarithm is e.

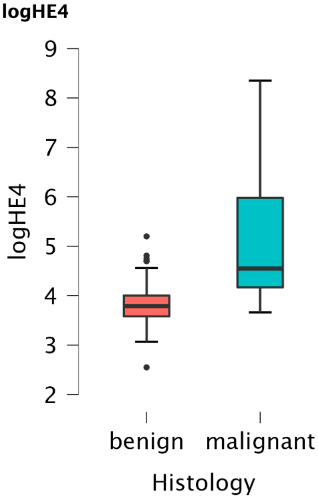
the `roma` dataset

The **standard logistic function** is the logistic function with

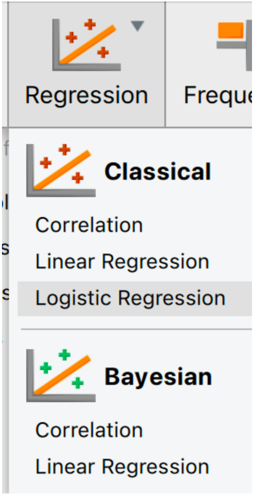
$$f(x) = \frac{1}{1+e^{-x}} = \frac{e^x}{e^x+1} = \frac{1}{2} + \frac{1}{2}\tanh\left(\frac{x}{2}\right).$$

the roma dataset

	logHE4	
	benign	malignant
Minimum	2.550	3.660
Maximum	5.200	8.350



the roma dataset

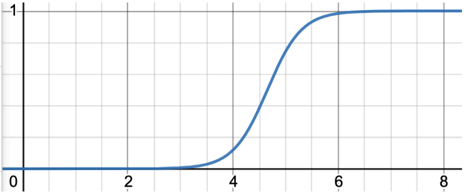


the roma dataset

				Wald Test		
	Est.	St. Error	z	Wald	df	p
(Intercept)	−14.28	2.38	−6.00	35.98	1	< .001
logHE4	3.07	0.57	5.38	28.94	1	< .001

$$f(x) = -14.28 + 3.07x$$

$$y = \frac{\exp\left(f(x)\right)}{\left(1 + \exp\left(f(x)\right)\right)}$$



repeated measures

Alice	Ellen
73.60	73.80

repeated measures

	Alice	Ellen
1	73.60	73.80
2	73.40	73.50
3	74.10	74.60
4	73.50	73.80
5	73.20	73.60

```
Two Sample t-test

data:  alice and ellen

t = -1.2227, df = 8, p-value = 0.2562
alternative hypothesis: true difference in means
is not equal to 0

95 percent confidence interval:
 -0.865794  0.265794
sample estimates:
mean of x mean of y
    73.56     73.86
```

repeated measures

	Alice	Ellen		Alice	Ellen
1	73.60	73.80	12	74.10	74.60
2	73.40	73.50	13	73.60	73.80
3	74.10	74.60	14	73.40	73.60
4	73.50	73.80	15	74.10	74.40
5	73.20	73.60	16	73.50	73.70
6	74.00	74.40	17	73.20	73.50
7	73.60	73.80	18	74.00	74.40
8	73.30	73.50	19	73.60	73.90
9	74.20	74.30	20	73.30	73.60
10	73.60	73.90	21	74.20	74.50
11	73.40	73.60	-	-	-

```
Two Sample t-test

data:  peso by gemella

t = -2.4594, df = 40, p-value = 0.01834
alternative hypothesis: true difference in means
is not equal to 0

95 percent confidence interval:
 -0.51183215 -0.05007261
sample estimates:
mean in group alice mean in group ellen
           73.66190            73.94286
```