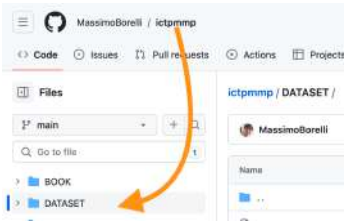
P5.2 Statistics for Medicine

Massimo Borelli

Master of Advanced Studies in Medical Physics

UNIVERSITÀ DEGLI STUDI DI TRIESTE — ICTP International Centre for Theoretical Physics



Finite random variables

The `otitis` dataset
(number of episodes of otitis media in the first two years of life)

0	1	2	3	4	5	6
0.129	0.264	0.271	0.185	0.095	0.039	0.017

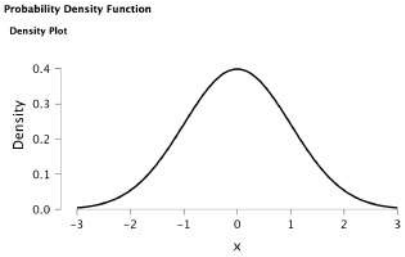
- **probability mass / discrete density** function

Example (graph)

Draw the barplot with JASP

Commonly used random variables

the Normal distribution



the Normal distribution

the Normal distribution

Example (B. Rosner, example 5.22)

The cerebral blood flow (CBF) in the general population is, approximately, normally distributed with mean $\mu = 75$ and standard deviation $\sigma = 17$ ($\sigma^2 = 289$). Which could be the percentage of persons having a CBF < 40?

the Normal distribution

Example (B. Rosner, example 5.23)

Glaucoma is characterized by intraocular pressure greater than 20 mmHg, while in normal population intraocular pressure X has mean $\mu = 16$ and standard deviation $\sigma = 3$.
How much it could be $P(12 \leq X \leq 20)$?

the Normal distribution

Example (again intraocular pressure)

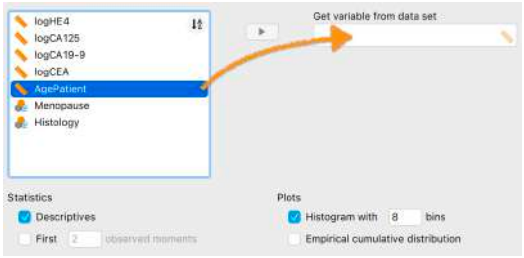
Can you find the upper and the lower fifth percentile of the intraocular pressure, as above defined ($\mu = 16, \sigma = 3$)?

the Normal distribution: the QQ plot

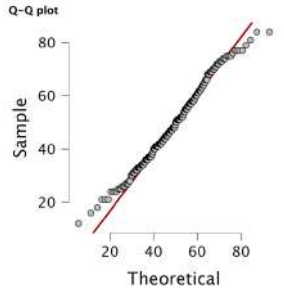
S. Najaf Zadeh
the `roma` dataset

- `Histology`
- `AgePatient`
- Menopause status
- four biomarkers (log transformed):
  - `logHE4`,
  - `logCA125`
  - `logCA19.9`
  - `logCEA`

the Normal distribution

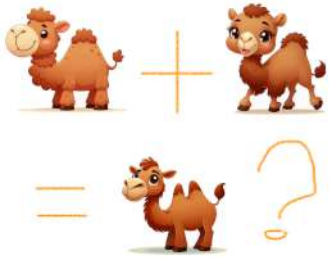


the Normal distribution



caveat: the Normal distribution

Do two dromedaries make a camel?



caveat: the Normal distribution

Do two dromedaries make a camel?

Bernard Rosner

> .. *linear combination of normal random variables are often of specific concern. It can be shown that any linear combination of normal random variables is itself normally distributed.*

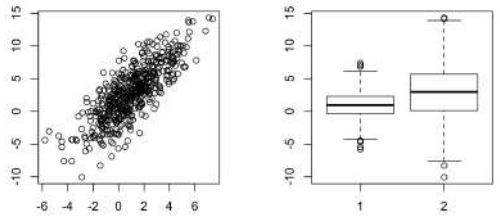
Martin Bland:

> ... *If we add two variables from Normal distributions together, even with different means and variances, the sum follows a Normal distribution.*

caveat: the Normal distribution

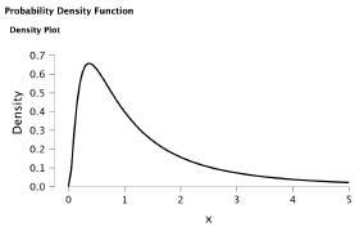
Do two dromedaries make a camel?

caveat: the Normal distribution



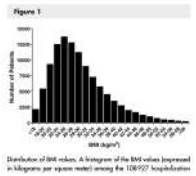
Edward L. Melnick, Aaron Tenenbein. 2012
Misspecifications of the Normal Distribution
https://www.tandfonline.com/doi/pdf/10.1080/00031305.1982.10483052

the log-Normal distribution



Eckhard Limpert, et al.
Log-normal Distributions across the Sciences: Keys and Clues
https://academic.oup.com/bioscience/article/51/5/341/243981

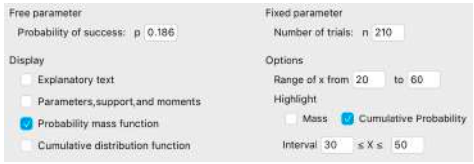
the logNormal distribution



Example (Gregg Fonarow et al. - summarizing body mass index)

Suppose that you are required to lead a pilot study concerning radiation dosimetry in 25 obese patients. How do you think you are going to describe the data? Using the mean and the standard deviation, or the median and the quartiles? What are here the difficulties?

the binomial distribution



Example (probability)

Suppose that you collect a new sample of 210 women with the same symptoms of those enrolled in `roma`. Obviously, only by chance you will observe exactly '39' malignancies. Can you compute the probability to observe a number of malignancy between 30 and 50?

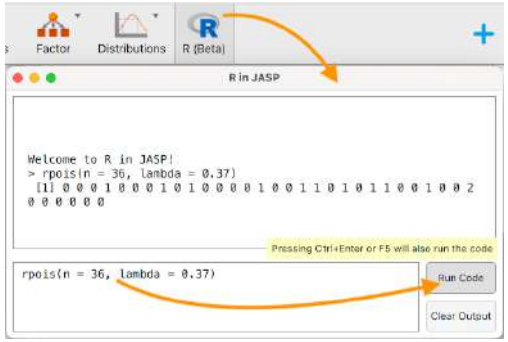
the Poisson distribution

1				1	2
2	3	1	2	1	
	1		2		1
1	1	2		4	1
1		1		3	1
	2	1	1		

Example (probability)

Use JASP to discover in a $\lambda = 0.37$ Poisson distribution how many, in probability, cells could have a value greater or equal than 2.

the Poisson distribution



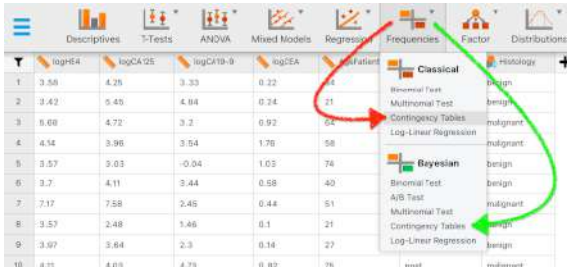
Statistics for Medicine

Massimo Borelli

Master of Advanced Studies in Medical Physics

UNIVERSITÀ DEGLI STUDI DI TRIESTE — ICTP International Centre for Theoretical Physics

Frequencies



Frequencies

	Menopause		
Histology	ante	post	Total
benign	106	65	171
malignant	12	27	39
Total	118	92	210

Table: Menopausal status and malignancy in ovarian cancer

Odds Ratio

	Menopause		
Histology	ante	post	Total
benign	**106**	**65**	171
malignant	**12**	**27**	39
Total	118	92	210

Example (Odds Ratio)

Explore the output of the Odds Ratio (2×2 only) checkbox in the Statistics menu of the contingency table of `Histology` (Rows) versus `Menopause` (Columns).

Bayes Theorem

$$P(malignant|ante) = \frac{P(ante|malignant)}{P(ante)} \cdot P(malignant)$$

	Menopause		
Histology	ante	post	Total
benign	106	65	171
malignant	**12**	27	**39**
Total	**118**	92	**210**

- prevalence
- sensitivity and specificity
- predictive values
- ROC curves
- ...

`https://github.com/MassimoBorelli/ictpmmp/tree/main/BOOK`

professor Luigi Rigon

the Bayes factor: JASP core business!

- Alice has a balanced urn with 5 winning black balls and 5 white balls ($p = 0.5$)
- Bob has a tricky urn with 6 winning black balls and 4 white balls ($p = 0.6$).

(binomial scheme, extractions with replacement)
we observe 115 successes over 200 draws, but without knowing if they are generated from Alice's or Bob's urn.

the Bayes factor: JASP core business!

$$P(X = 115|\ Alice\ ) = \binom{200}{115} \cdot 0.5^{115} \cdot 0.5^{200-115} \approx 0.006$$

$$P(X = 115|\ Bob\ ) = \binom{200}{115} \cdot 0.6^{115} \cdot 0.4^{200-115} \approx 0.044$$

the Bayes factor: JASP core business!

$$\frac{P(X = 115|\ Bob\ )}{P(X = 115|\ Alice\ )} \approx \frac{.044}{.006} \approx 7.4$$

it is much more likely that the balls have been drawned by Bob's urn: about seven times higher

the **Bayes factor**:

$$\frac{P(D|M_1)}{P(D|M_2)} = \frac{P(M_1|D)}{P(M_2|D)} \cdot \frac{P(M_2)}{P(M_1)}$$

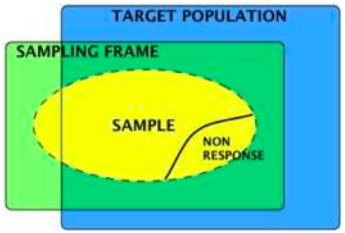
P5.2 Statistics for Medicine

Massimo Borelli

Master of Advanced Studies in Medical Physics

UNIVERSITÀ DEGLI STUDI DI TRIESTE — ICTP International Centre for Theoretical Physics

what about inference



Mismatching variability with reliability

Dispersion

- [x] S.E.mean
- [x] Std.deviation
- [ ] Coefficient of Variation
- [ ] MAD
- [ ] MAD Robust
- [ ] IQR
- [ ] Variance
- [ ] Range
- [ ] Minimum
- [ ] Maximum

Figure: The old version of JASP

Mismatching variability with reliability

the dataset `cholesterol.ods`

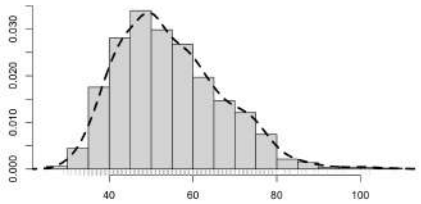
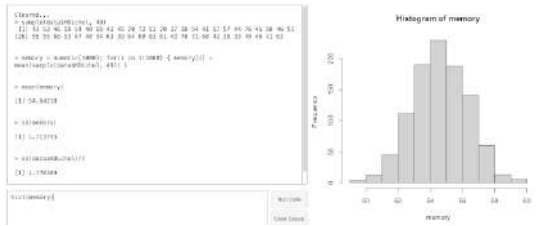


Figure: `hdlchol`: mean? std. deviation?

Mismatching variability with reliability



Mismatching variability with reliability

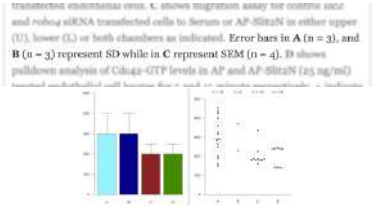
$$\frac{\sigma}{\sqrt{n}}$$

	HDLchol
Valid	1025
Mean	54.685
Std. Error of Mean	0.387
Std. Deviation	12.392

Mismatching variability with reliability

Yu-Kang Tu and Mark Gilthorpe.
The most dangerous hospital or the most dangerous equation?
https://bmchealthservres.biomedcentral.com/articles/10.1186/1472-6963-7-185

Mismatching variability with reliability



Tatsuki Koyama.
Beware of Dynamite
https://biostat.app.vumc.org/wiki/pub/Main/TatsukiRcode/Poster3.pdf