Lecture 42 Wrap Up And Descriptive Statistics Revisited

BIO210 Biostatistics

Xi Chen

Spring, 2022

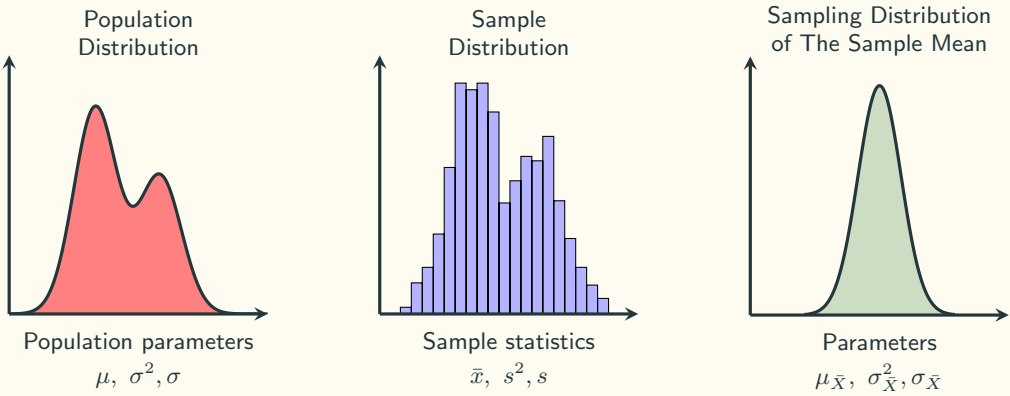
School of Life Sciences
Southern University of Science and Technology

南方科技大学生命科学学院
SUSTech · SCHOOL OF
LIFE SCIENCES

Course Content Review

- Descriptive statistics
- Probability
- Inferential statistics
  - Estimation
  - Hypothesis testing

Three Distributions



1. The exact sample statistics are not of our interest. More important: what the sample represents.

2. How to choose an appropriate test? All you need to ask: what is the sampling distribution of the test statistics.

What's Next?

- Plot the raw data
- Look at the data from all sorts of different angles
- Care about effect sizes
- Practice, read and use what you have learnt
- Bayesian statistics
- Learn a programming language

Anscombe's Quartet

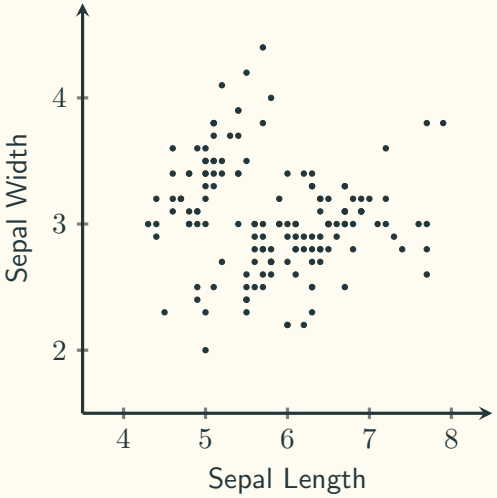
By Francis Anscombe in 1973

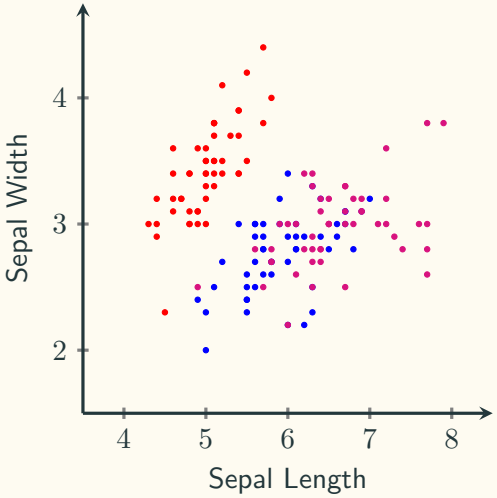
x_1	y_1	x_2	y_2	x_3	y_3	x_4	y_4
10	8.04	10	9.14	10	7.46	8	6.58
8	6.95	8	8.14	8	6.77	8	5.76
13	7.58	13	8.74	13	12.74	8	7.71
9	8.81	9	8.77	9	7.11	8	8.84
11	8.33	11	9.26	11	7.81	8	8.47
14	9.96	14	8.1	14	8.84	8	7.04
6	7.24	6	6.13	6	6.08	8	5.25
4	4.26	4	3.1	4	5.39	19	12.5
12	10.84	12	9.13	12	8.15	8	5.56
7	4.82	7	7.26	7	6.42	8	7.91
5	5.68	5	4.74	5	5.73	8	6.89

$\bar{x} = 9.0$, $\bar{y} = 7.5$, $s_x^2 = 10$, $s_y^2 = 3.75$

Ordinary Least Square regression: $y = 0.5x + 3$

Simpson's Paradox





Effect Size

Two education companies (A & B) have developed their own learning programmes. They both think their programme can improve the test scores of students. Company A has more resource so they recruited many volunteers. Company B has limited resource so they only recruited a small number of volunteers. The results of the test scores are summarised below:

Results from Company A

	Control A	**Programme A**
sample size	1000	1000
mean	99.90	104.81
variance	94.59	96.75

Results from Company B

	Control B	**Programme B**
sample size	20	20
mean	97.85	114.22
variance	96.53	99.23

Question: which programme do you think is more effective?

Huge Amount of Data In Modern Biology

Nature, 381: 620-3 (1996)

A human Mad protein acting as a BMP-regulated transcriptional activator

Fang Liu, Akiko Hata, Julie C. Baker*, Jacqueline Doody, Juan Cárcamo, Richard M. Harland* & Joan Massagué

Cell Biology Program and Howard Hughes Medical Institute, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, New York 10021, USA
* Department of Molecular and Cell Biology, Division of Biochemistry and Molecular Biology, University of California, Berkeley, California 94709, USA

THE TGF-β/activin/BMP cytokine family signals through serine/threonine kinase receptors, but how the receptors transduce the signal is unknown. The *Mad* (*Mothers against decapentaplegic*) gene from *Drosophila*[1] and the related *Sma* genes from *Caenorhabditis elegans*[2] have been genetically implicated in signalling by members of the bone-morphogenetic-protein (BMP) subfamily. We have cloned *Smad1*, a human homologue of *Mad* and *Sma*. Microinjection of *Smad1* messenger RNA into *Xenopus* embryo animal caps mimics the mesoderm-ventralizing effects of BMP4. Smad1 moves into the nucleus in response to BMP4. Smad1 has transcriptional activity when fused to a heterologous DNA-binding domain, and this activity is increased by BMP4 acting through BMP-receptor types I and II. The transactivating activity resides in the conserved carboxy-terminal domain of Smad1 and is disrupted by a nonsense mutation that corresponds to null mutations found in *Mad* and in the related gene *DPC4*, a candidate tumour-suppressor gene in human pancreatic cancer[3]. Additionally, we show that DPC4 contains a transcriptional activation domain. The results suggests that the Smad proteins are a new class of transcription factors that mediate responses to the TGF-β family.

Nature, 577: 566-571 (2020)

Article

TGF-β orchestrates fibrogenic and developmental EMTs via the RAS effector RREB1

https://doi.org/10.1038/s41586-019-1897-5
Received: 4 October 2018
Accepted: 5 November 2019
Published online: 8 January 2020
There are amendments to this paper

Jie Su[1], Sophie M. Morgani[2,3], Charles J. David[1,4], Qiong Wang[1,7], Ekrem Emrah Er[1], Yun-Han Huang[1,4], Harihar Basnet[1], Yilong Zou[1,4,8], Weiping Shu[1], Rajesh K. Soni[5], Ronald C. Hendrickson[6], Anna-Katerina Hadjantonakis[2] & Joan Massagué[1]*

Epithelial-to-mesenchymal transitions (EMTs) are phenotypic plasticity processes that confer migratory and invasive properties to epithelial cells during development, wound-healing, fibrosis and cancer[1–4]. EMTs are driven by SNAIL, ZEB and TWIST transcription factors[5,6] together with microRNAs that balance this regulatory network[7,8]. Transforming growth factor β (TGF-β) is a potent inducer of developmental and fibrogenic EMTs[4,9,10]. Aberrant TGF-β signalling and EMT are implicated in the pathogenesis of renal fibrosis, alcoholic liver disease, non-alcoholic steatohepatitis, pulmonary fibrosis and cancer[4,11]. TGF-β depends on RAS and mitogen-activated protein kinase (MAPK) pathway inputs for the induction of EMTs[12–19]. Here we show how these signals coordinately trigger EMTs and integrate them with broader pathophysiological processes. We identify RAS-responsive element binding protein 1 (RREB1), a RAS transcriptional effector[20,21], as a key partner of TGF-β-activated SMAD transcription factors in EMT. MAPK-activated RREB1 recruits TGF-β-activated SMAD factors to *SNAIL*. Context-dependent chromatin accessibility dictates the ability of RREB1 and SMAD to activate additional genes that determine the nature of the resulting EMT. In carcinoma cells, TGF-β–SMAD and RREB1 directly drive expression of SNAIL and fibrogenic factors stimulating myofibroblasts, promoting intratumoral fibrosis and supporting tumour growth. In mouse epiblast progenitors, Nodal–SMAD and RREB1 combine to induce expression of SNAIL and mesendoderm-differentiation genes that drive gastrulation. Thus, RREB1 provides a molecular link between RAS and TGF-β pathways for coordinated induction of developmental and fibrogenic EMTs. These insights increase our understanding of the regulation of epithelial plasticity and its pathophysiological consequences in development, fibrosis and cancer.

Huge Amount of Data In Modern Biology

	Nature, 381: 620-3 (1996)	Nature, 577: 566-571 (2020)
Figures & panels	4 figures (8 panels)	4 figures (33 panels) + 10 supplementary figures (82 panels) + 2 GEO submissions
Experiments	Cell culture, Xenopus cap, gene reporter assay, immunofluorescence	2 mouse models, RNAi screen, ChIP-seq, RNA-seq, organoid culture, ...
Statistics	No	A lot !