

生物医学研究中统计方法应用和图表绘制

授课人：张力博士

整理：陈靓



(扫码可直接观看视频课程)

在生物学和医学研究中，对数据的合理统计分析并绘制正确清晰的图表，是得出可信的科学结论，并完成高质量论文投稿的基本要求。本次“优选课程”主要针对在统计分析和绘制统计图表上感到困难的研究人员特别是广大研究生，着重介绍统计分析的基本知识，如何选择合理的统计检验，并举例介绍如何使用 GraphPad Prism 软件进行统计分析，并轻松绘制出符合论文发表需要的数据图表的方法。

一、基础篇

1. 统计分析的流程 (收集数据、统计方法选择、统计假设、计算分析、得出结论)

统计假设：

➢ 零假设 (H_0): 假定两组 (或

多组) 数据间无差别

➢ 备择假设 (H_A): 假定数据间存在差别

统计分析的基本思想为：假定 H_0 成立，通过计算 P 值，接受 (拒绝) H_0 ，从而判定组间不存在 (存在) 差别。

2. 数据类型和统计方法的选择

1) 参数 & 非参数化数据

➢ 参数化数据为 (近似) 符合正态分布的样本数据

➢ 非参数化数据为显著性偏离正态分布的样本数据

2) 配对 & 非配对检验

➢ 配对检验：用于同一个样本前后对比 (例如同一个病人给药前和给药后的反应)

➢ 非配对检验：两个 (或多个) 相互独立的样本 (例如两组不同病人分别给药或是安慰剂)

3) 两组 & 多组比较

判定是否需要多组比较，主要看各组间是否为完全独立。

➢ 如果为不同时间点，则必须应用多组比较 (存在联系)。

➢ 如果为严格的相互独立，可使用 t- 检验两两比较，或是多组比较 (视乎需要)。

4) 单一变量 & 二重变量分析

➢ 单一变量：t- 检验或单变量方差分析或 Kruskal 沃利斯检验

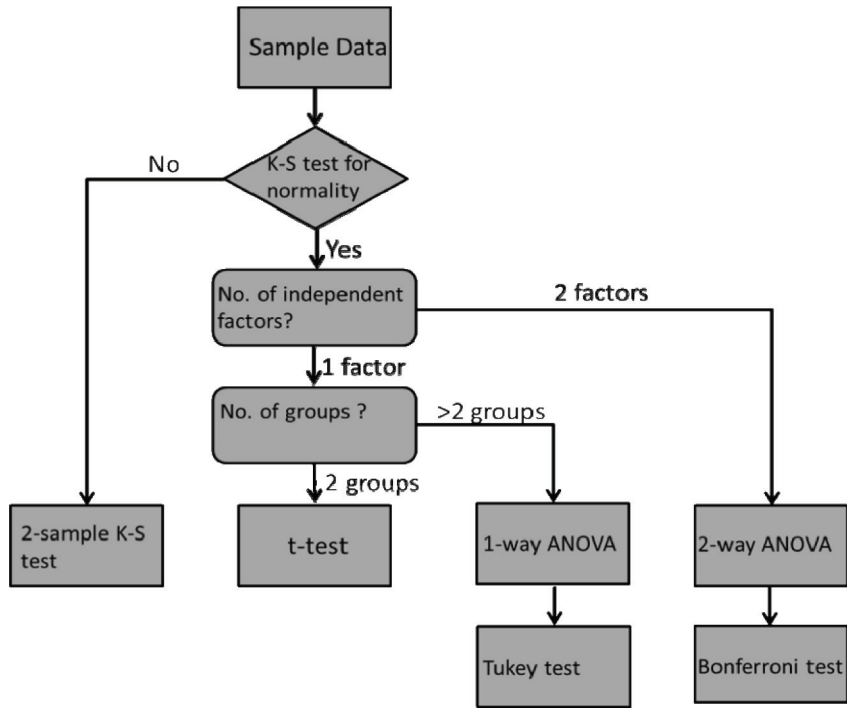
➢ 两个自变量：双因素方差分析

➢ 事后检验 (Post-hoc test)：Kruskal 沃利斯检验 (Dunn' s test)，单变量方差分析 (Tukey' s test)，双因素方差分析 (Bonferroni



test)。

3. 统计方法的选择 (小结流程图)



二、进阶篇

在本篇的讲解中，我们从各个实例，看应该如何使用 GraphPad Prism 软件，进行统计分析并绘制出完美的数据图表。

1) 例一：正态分布检验

样本 A: 2,3,3,4,4,4,5,5,6

样本 B: 2,4,4,4,6,6,6,6,6

非正态分布数据，应选择散点图类型。

2) 例二：方差分析结果的解读—F 值

	野生型 (Wild type)	基因敲入 (Knock-in)	基因敲除 (Knock-out)
动物前肢肌力	6.5	16.8	1.2
	4.3	22.5	0.8
	2.1	7.8	3.5
	7.9	16.2	1.9
	10.2	8.4	0.8
	11.1	6.9	5.8

One-way analysis of variance:

P value	0.0023
P value summary	**
Are means significant difference? (P < 0.01)	Yes
Number of groups	3
F	9.393
R square	0.5560

- 上述结果在论文中应表述为： $F(2, 15) = 9.393, P < 0.01$
 自由度 $dF = \text{组数} - 1$ 组内变异度 $= N - \text{组数}$
- F 值反映的是所有样本总体的组间差异显著性
- 单独两组间的比较需要利用事后检验 (而非 t-test)

3) 例三：双变量方差分析的解读：交叉效应

训练次数	野生型 (Wild type)	基因敲除 (Knock-out)
1	6.7 ± 1.1	7.3 ± 1.0
2	12.4 ± 2.3	9.5 ± 2.2
3	17.5 ± 4.9	10.4 ± 2.5
4	21.3 ± 5.1	12.3 ± 3.1
5	22.6 ± 4.7	13.6 ± 4.0

- 交叉效应：不同基因型动物间随着训练次数其行为改善的程度不同。

Source of Variation	P value summary	Significant?
Interaction 1. 交叉效应	***	Yes
Column Factor 2. 组间差异(基因型)	***	Yes
Row Factor 3. 行间差异(训练次数)	***	Yes

Source of Variation	Df	Sum-of-squares	Mean square	F
Interaction	4	372.1	93.01	8.148
Column Factor	1	786.5	786.5	68.90
Row Factor	4	1695	423.8	37.12
Residual	95	1084	11.42	

本例的结论：野生型和基因敲除动物之间存在显著的行为学差异 ($F_{\text{genotype}}(1, 95) = 68.90, P < 0.001$)，同时不同基因型动物的行为学随训练次数增加的改善情况也有显著差异 ($F_{\text{Interaction}}(4, 95) = 8.418, P < 0.001$)。

4) 例四：卡方检验 (比率的检验)

	患癌症	未患癌症
野生型 (Wild type)	11	19
基因敲除 (Knock-out)	12	8

P value	0.0089
P value summary	**
One- or two-sided	Two-sided
Statistically significant? (alpha<0.05)	Yes

本例的结论：两组动物间癌症发病率存在显著差别。

5) 例五：图表的格式操作（以复合折线图为例）

Format Graph

- Appearance: Mean and Error Plot: SD
- Data set: Example 3:Knock-out
- Show symbols: Color: [Black] Shape: [Square] Size: 4
- Show error bars: Dir.: Above Style: T Thickness: 2 pt
- Show connecting line/curve: Thickness: 1 pt
- Plot on: Left Y axis Right Y axis

Export Graph

- File format: TIF (Tagged Image File)
- Resolution (dpi): 300
- Color Model: CMYK
- Size: Actual Make width 9.26 cm
- Where to export: Folder: C:\Users\Lee Zhang\OneDrive\论文\投稿\submit-mp

Annotations:

- 选择数据组 (Select data set)
- 选择数据展现方式 (Select data display method)
- 数据标记格式 (Data marker format)
- 误差棒格式 (Error bar format)
- 折线格式 (Line format)
- 图例 (Legend)
- 坐标轴 (Axis)
- 建议选择 TIF 格式 (Recommended to choose TIF format)
- 分辨率至少 300dpi, 推荐 600dpi (Resolution at least 300dpi, recommended 600dpi)
- 此处应该输入投稿要求的图片宽度 (This is where you should enter the image width required by the journal)
- 文件名称和存储目录 (File name and storage directory)