



# 实验结果的分析与统计



武汉大学 郭喻

发育源性疾病湖北省重点实验室

[guoy@whu.edu.cn](mailto:guoy@whu.edu.cn)





# 讲座提纲

## 一、结果的分析

- ① 统计学意义不是一切，结果的合理性是第一
- ② 阴性与阳性对照
- ③ “律师式思维”与“科学家式思维”

## 二、数据统计

- ① 计量指标及统计方法
- ② 计数指标及统计方法
- ③ 等级指标及统计方法





- **There is no good or bad result, there's only result that you can explain or not.**
- **“Every reaction has a mechanism, but not every mechanism has a reaction.”——Richard B. Turner, 1968**





## ① 统计学意义不是一切，结果的合理性是第一

**方法学：20 cases of healthy newborn infants with Apgar score > 7 at 5 minutes served as control.**

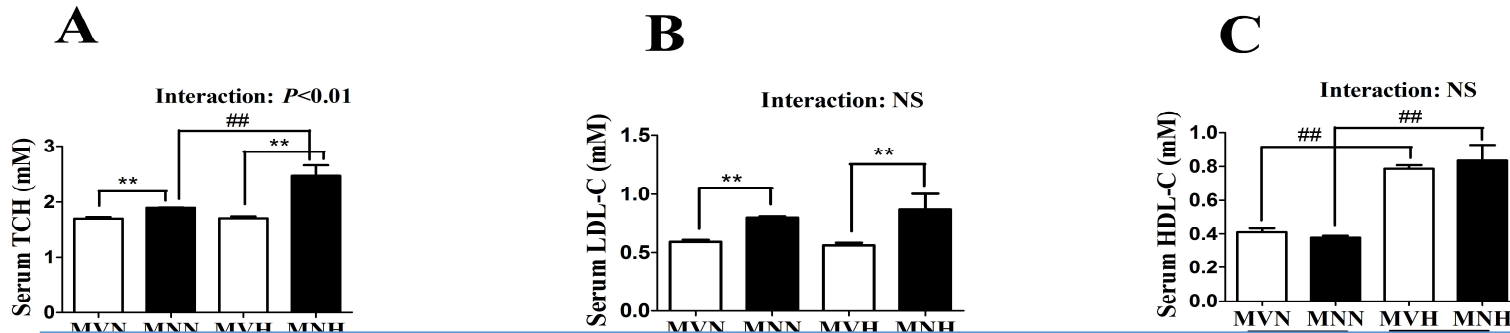
**结果：Apgar score of the study group at 5 minutes after birth was  $4.6 \pm 1.3$  minutes and that of the controls was  $8.5 \pm 2.3$  minutes with a highly significant statistical difference detected between them ( $p\text{-value} < 0.01$ ).**



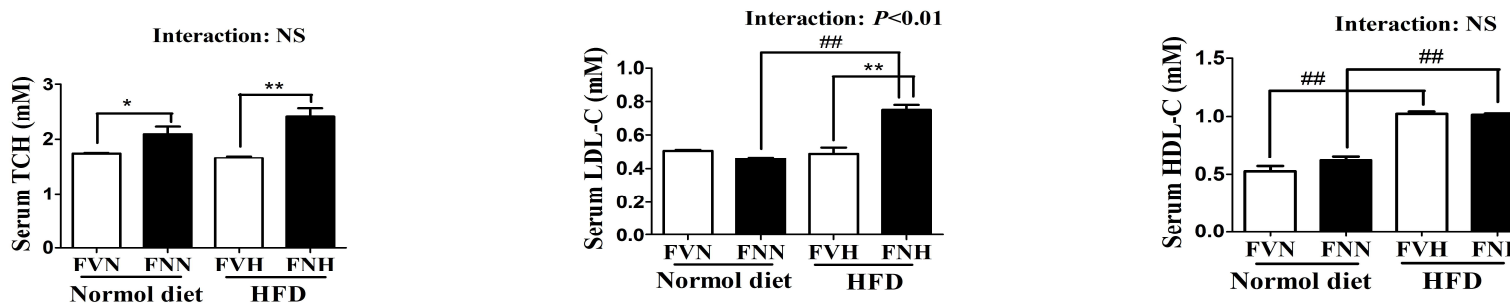


# 举例

## Male



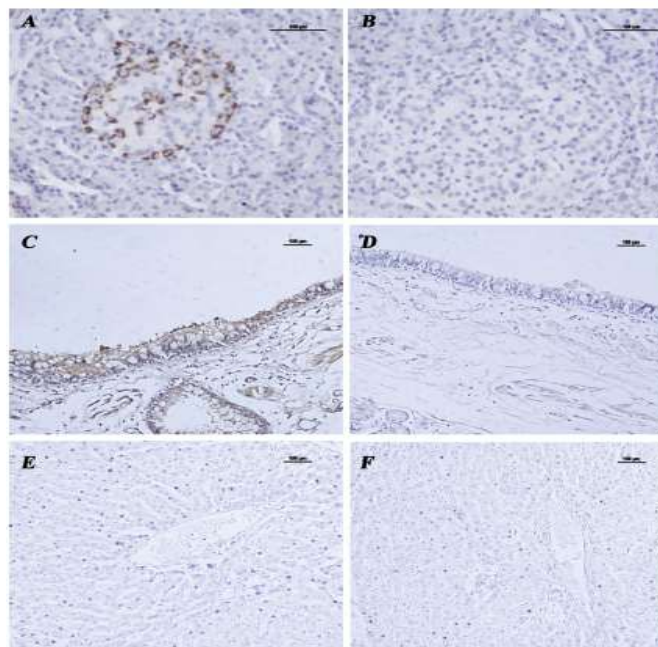
the results presented in figure 1 are not reliable. If we consider only control animals (not exposed to nicotine and with regular diet), they had values of TC around 1.7-1.8 mM, with 0.6 mM in the LDL fraction and 0.4 in the HDL fraction: there are about 0.7-0.8 mM of cholesterol missing here...





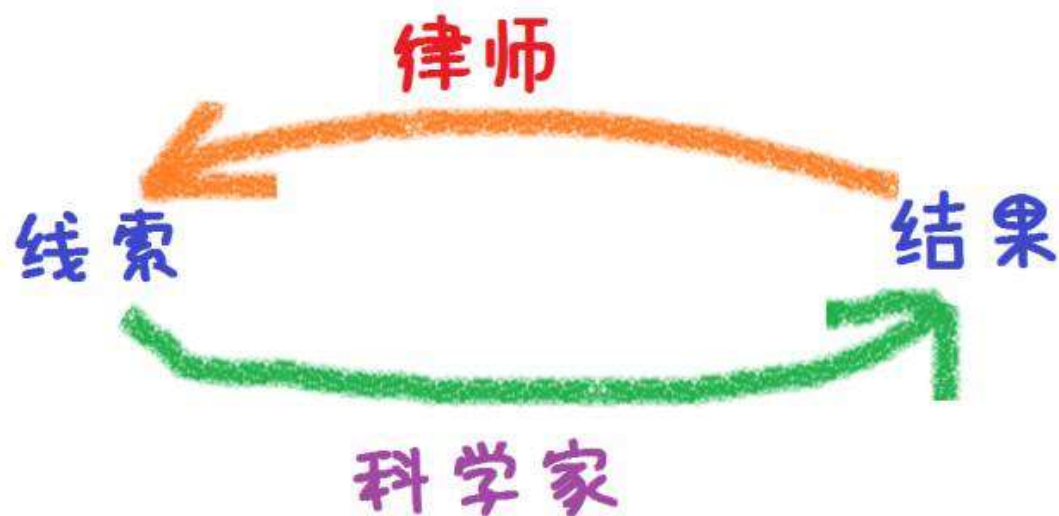
## ② 阴性与阳性对照

- **阳性对照**：过表达、含量丰富组织、有确切疗效的药物等，用于反映方法学或体系的有效性
- **阴性对照**：溶媒、空载体、安慰剂等，用于排除假阳性或心理因素。





### ③ “律师式思维”与“科学家式思维”

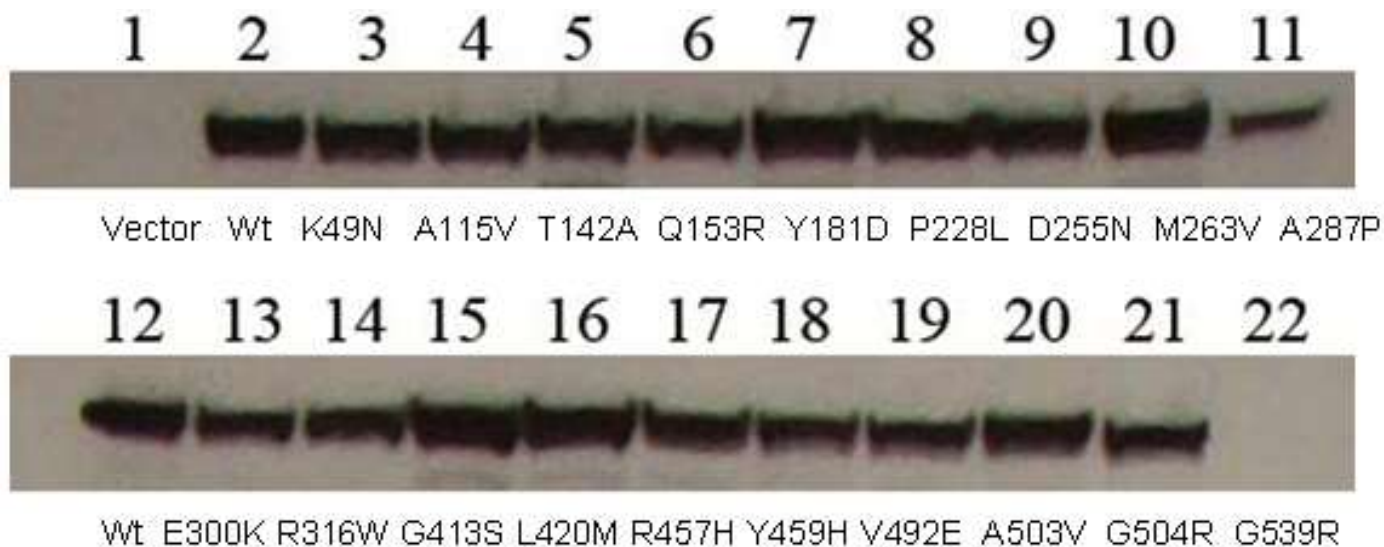


转载自公众号“实验万事屋”文章——  
《为什么你们的课题辣么Low？主要还是因为……》





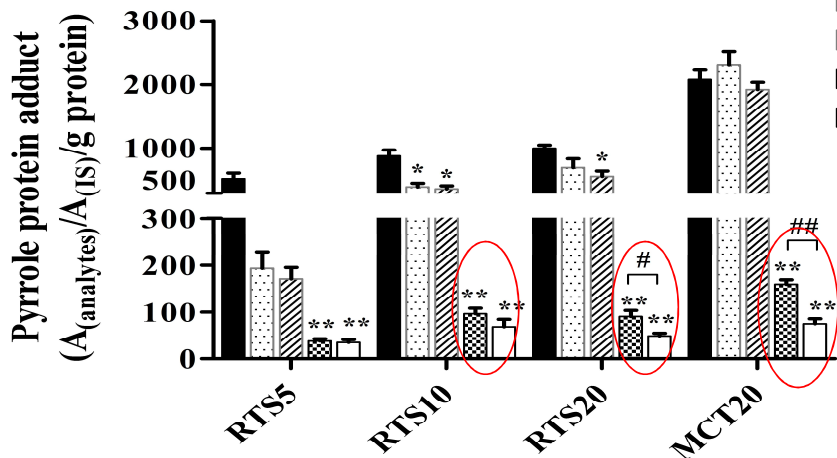
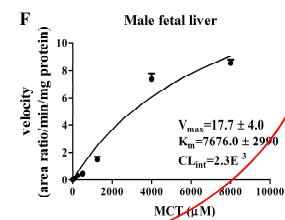
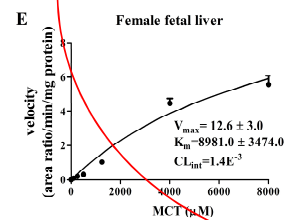
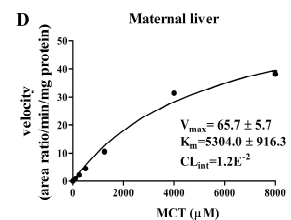
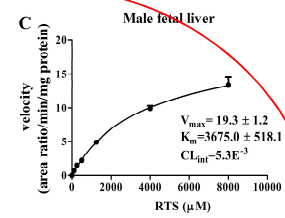
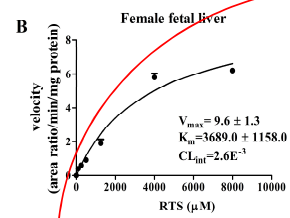
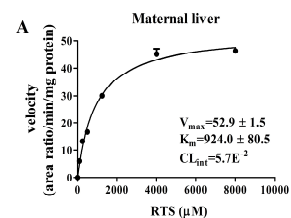
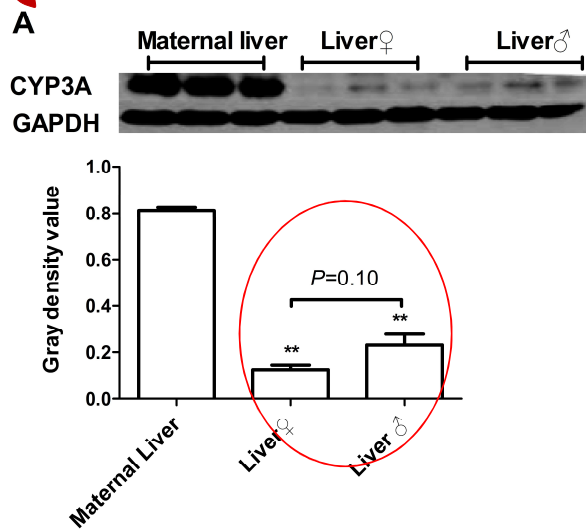
## 举例





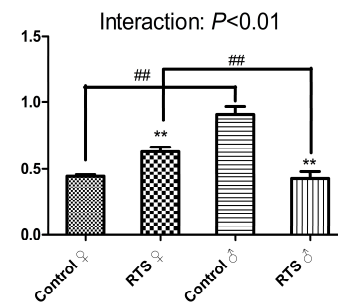
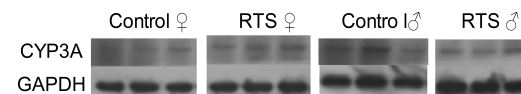


# 举例



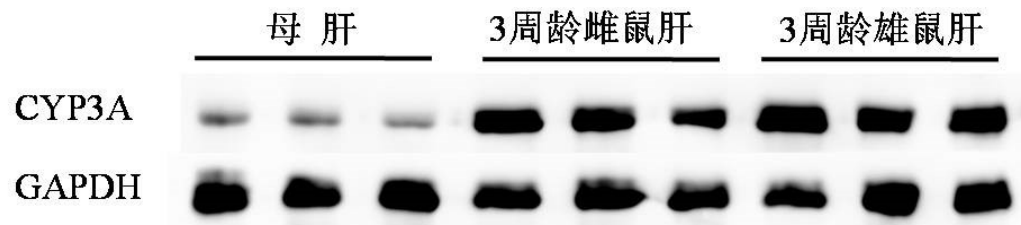
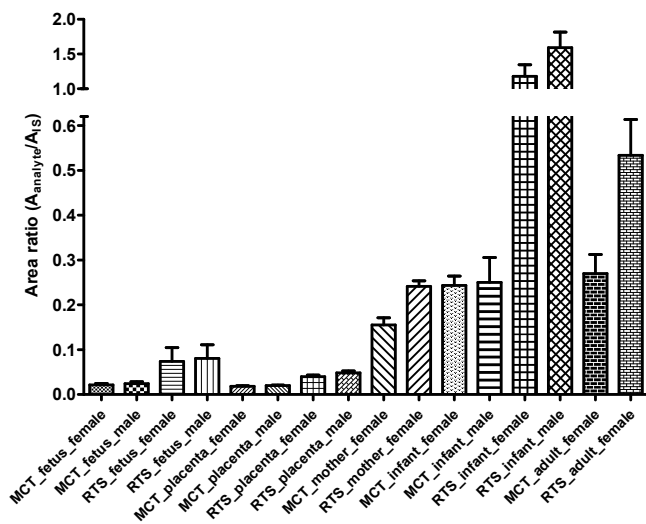
Maternal Liver

Placenta<sub>♀</sub>  
Placenta<sub>♂</sub>  
Liver<sub>♀</sub>  
Liver<sub>♂</sub>

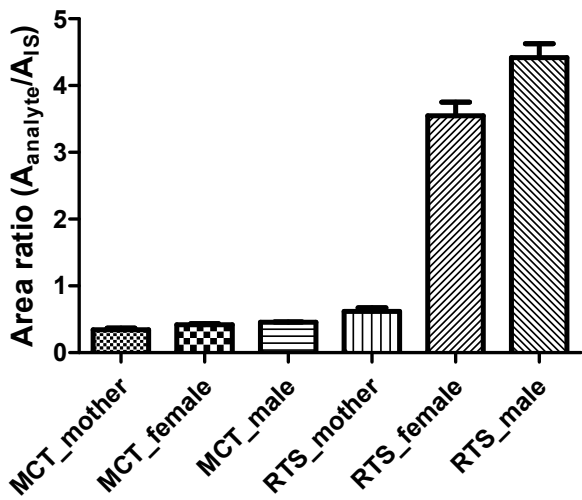




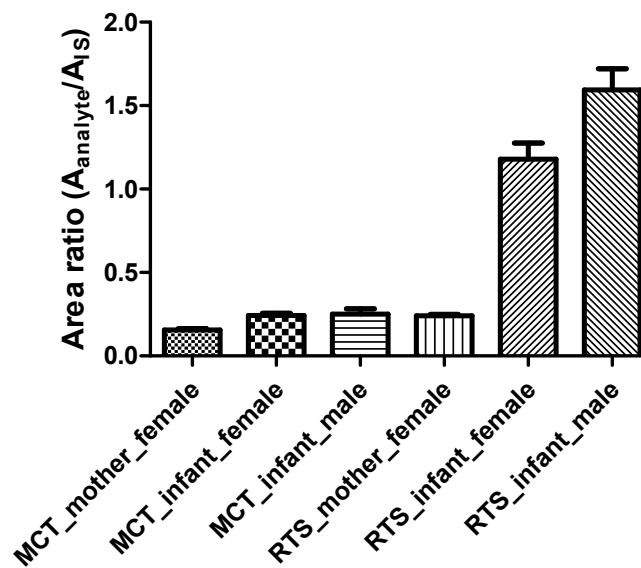
举例



李霞数据 20170413



李霞数据 20161227





# 讲座提纲

## 一、结果的分析

- ① 统计学意义不是一切，结果的合理性是第一
- ② 阴性与阳性对照
- ③ “律师式思维”与“科学家式思维”

## 二、数据统计

- ① 计量指标及统计方法
- ② 计数指标及统计方法
- ③ 等级指标及统计方法





## 结果的统计

- 根据数据类型选择正确的统计方法
- 两两比较选什么？
- 交互作用（两因素和重复测量）
- 三因素方差分析





## (1) 计量资料 (定量指标)

用定量的方法测定所获得的资料，如血压、血糖、血细胞数等。常带有度量衡单位，如cm、mg、ml等。

Table 1 Effects of sodium ferulate (SF) on serum indexes of glycerol-induced renal injury in mice

Groups	Dosages (mg·kg <sup>-1</sup> )	BUN (mmol·L <sup>-1</sup> )	Cr (μmol·L <sup>-1</sup> )	NAG (U·L <sup>-1</sup> )	GST (mmol·min <sup>-1</sup> ·L <sup>-1</sup> )
Control		11.4 ± 1.8	47 ± 23	207 ± 9	31 ± 6
SF	0	27.7 ± 5.4**	144 ± 37**	248 ± 24**	57 ± 6**
	100	25.2 ± 9.3**	76 ± 34 <sup>△△</sup>	231 ± 23*	55 ± 6**
	150	25.3 ± 9.7**	97 ± 38* <sup>△</sup>	229 ± 20	53 ± 11**
	200	16.3 ± 4.0 <sup>△△</sup>	64 ± 11 <sup>△△</sup>	222 ± 17 <sup>△</sup>	53 ± 9**

$\bar{x} \pm s$ ,  $n = 8$ . \* $P < 0.05$ , \*\* $P < 0.01$  vs control group;  $\Delta P < 0.05$ ,  $\Delta\Delta P < 0.01$  vs SF 0 mg·kg<sup>-1</sup> group.





## 常用统计方法

### (1) $t$ 检验

用于两组均数、前后对比或配对对比差值均数的显著性检验，包括完全随机化设计的两样本均数的  $t$  检验和配对  $t$  检验。

可用Excel 统计软件直接获得P值

其判断标准为：

$t < t_{0.05}$ ，则  $P > 0.05$ ，差异无显著性意义。

$t \geq t_{0.05}$ ，则  $P \leq 0.05$ ，差异有显著性意义。

$t \geq t_{0.01}$ ，则  $P \leq 0.01$ ，差异有高度显著性意义。

$$t = \frac{|\bar{X}_1 - \bar{X}_2|}{S_{\bar{X}_1 - \bar{X}_2}}$$

$t$  检验

$$t = \frac{|d|}{S_{\bar{d}}}$$

配对  $t$  检验

### (2) 方差分析

用于多组样本均数间差别的显著性检验。

可用SPSS 统计软件直接获得F值





## (2) 计数资料（定性指标）

将观察单位按某种属性或类别分组计数，得到各组观察单位的个数，如患病与未患病的人数；某人群中 **O**、**A**、**B**、**AB** 各种血型的人数。率(如阳性率、死亡率)是最常用的指标。





## 常用统计方法

### ① 2×2 表的 $\chi^2$ 检验

a. 基本公式: 
$$\chi^2 = \sum \frac{(A-T)^2}{T} \quad T = \frac{\text{行合计} \times \text{列合计}}{\text{总合计}}$$

A: 实际数(观察数); T: 理论数

### b. 四格表专用公式:

表2-2 实验组与对照组的有效例数和无效例数

	治疗效果		合计
	显效	无效	
甲组	a	b	a+b
乙组	c	d	c+d
合计	a+c	b+d	n = a+b+c+d

$$\chi^2 = \frac{(ad-bc)^2 n}{(a+b)(c+d)(a+c)(b+d)}$$







## ② $2 \times K$ 表的 $\chi^2$ 检验

如果行或列中有一个分组数大于2时，统称为  $2 \times K$  表，其  $\chi^2$  检验的通用公式和步骤均同四格表，但因较繁，故采用  $2 \times K$  表专用公式。

$$\chi^2 = n \left( \sum \frac{A^2}{n_R n_C} - 1 \right)$$

n: 总例数 A: 每格实际数

$n_R$ 、 $n_C$  分别为每格实际数A同行、同列的合计数。

公式中自由度  $\nu = 1$ ，故  $\chi^2_{0.05} = 3.84$ ， $\chi^2_{0.01} = 6.63$ 。

计算出  $\chi^2$  值后，按下面规律作出判断：

- ①  $\chi^2 > 6.63$ ，则  $P < 0.01$ ，差别有高度显著性意义。
- ②  $\chi^2 > 3.84$ ，则  $P < 0.05$ ，差别有显著性意义。
- ③  $\chi^2 < 3.84$ ，则  $P > 0.05$ ，差别无显著性意义。





### (3) 等级资料 (半定量资料)

将观察单位按不同程度分组计数，得到各组观察单位的个数，如疾病痊愈、显效、改善、无效和恶化的人数；组织病变程度为+++、++、+、±、-的人数；疼痛缓解为高度、中度、轻度、无效的人数等。其准确性介于计量资料和计数资料之间。





## 常用统计方法

等级资料多属于偏态资料，其分析常用非参数统计法，如等级序值法、秩和检验等进行显著性检验。

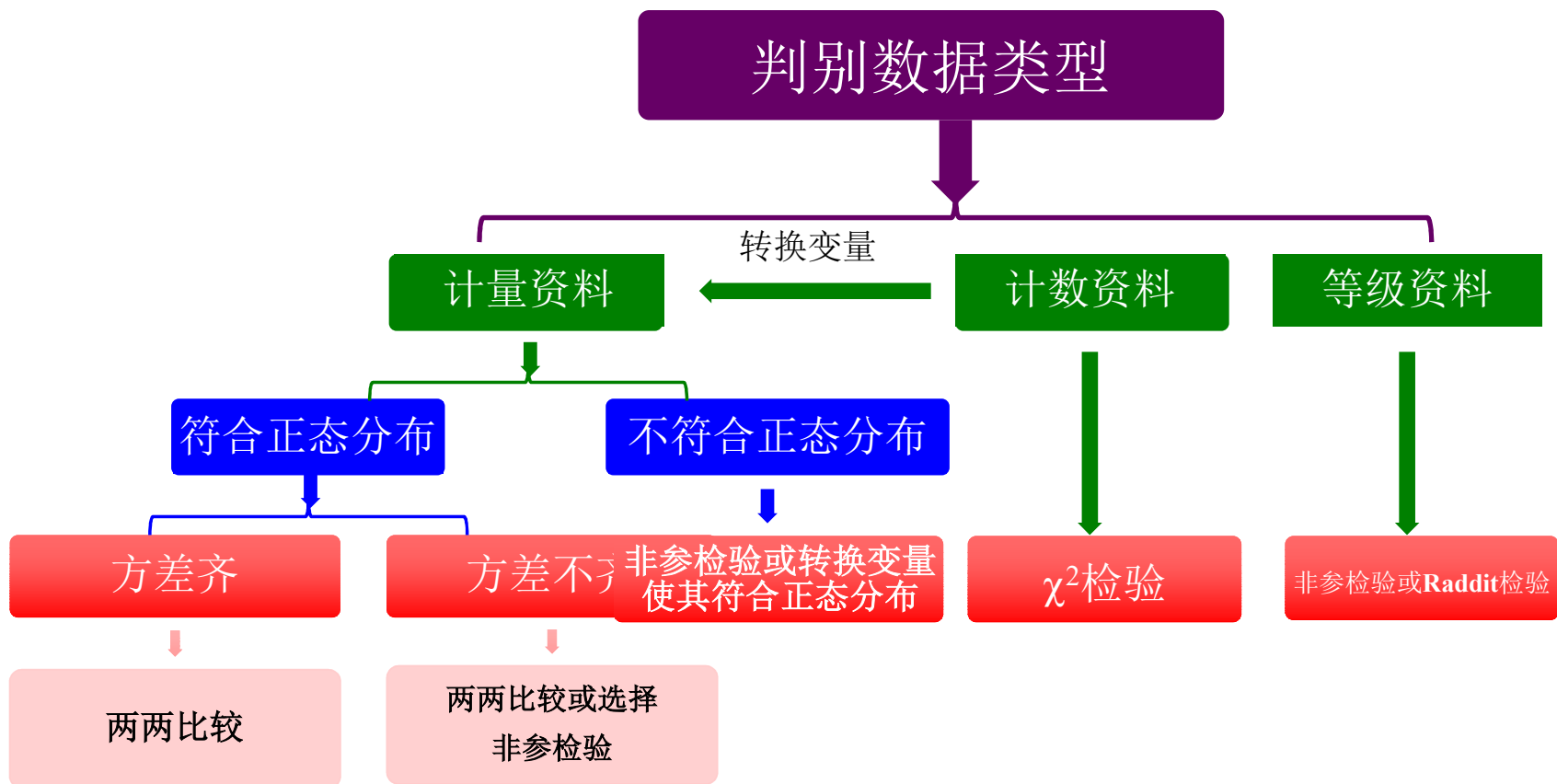
表2. 模型对照组与各给药组秩和检验时各组秩次和

组别	例数 (只)	Mankin' s评分(分)				总分 <sup>#</sup>
		结构	细胞	基质染色	潮线完整性	
模型对照组	8	169.5	192	174	169	214
丹参组	8	131	112	103	130	100.5
HA组	8	129.5	88	101	114.5	111.5
丹参和HA组	8	98	124	150	114.5	102





# 总结





## 错误的示范:

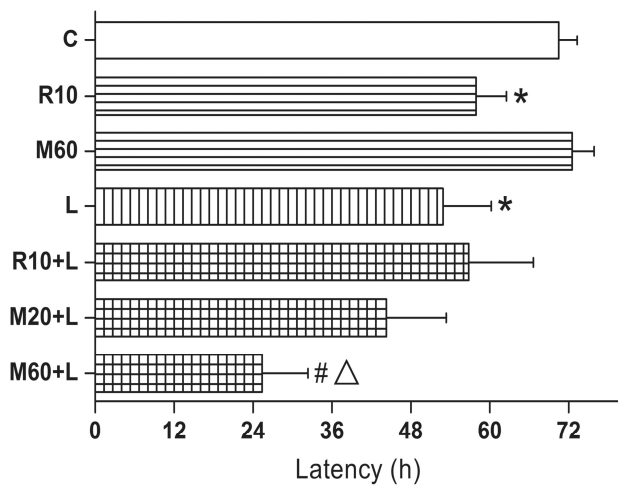
表1. 乙醇对孕13d小鼠胎仔数、吸收胎数和着床点数的影响

组别	平均胎仔数 (个)	平均吸收胎数 (个)	平均着床点数 (个)
正常对照组 (n=15)	10.6±1.9	0.4±0.6	11.0±1.9
乙醇小剂量 (n=14)	10.1±2.5	0.6±0.9	10.7±2.5
乙醇中剂量 (n=11)	9.5±3.2	1.0±2.0	10.6±1.6
乙醇大剂量 (n=9)	9.6±2.9	1.1±2.3	10.7±1.8



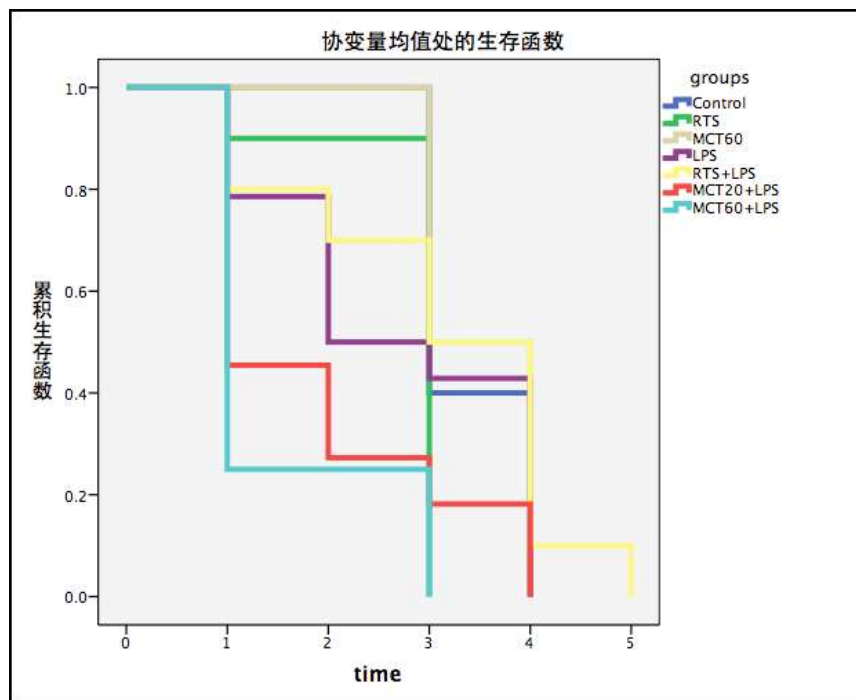


举例



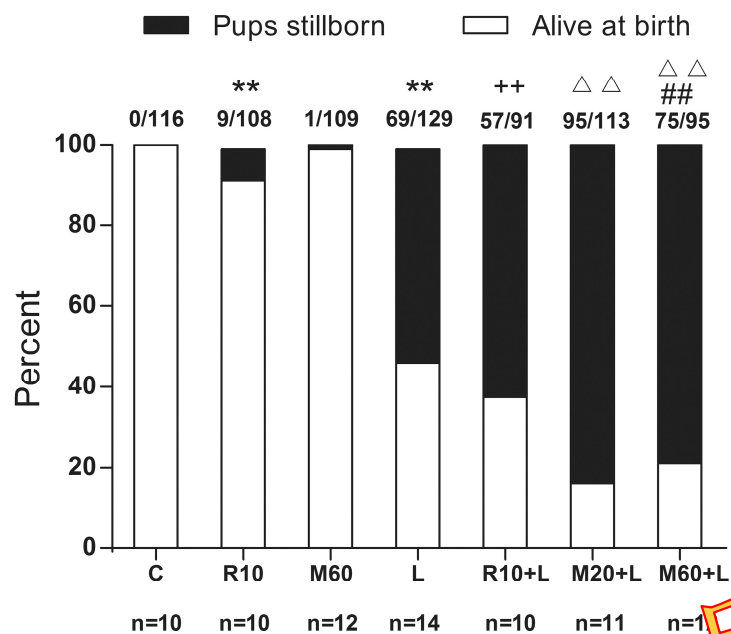
One way-ANOVA

Log rank test

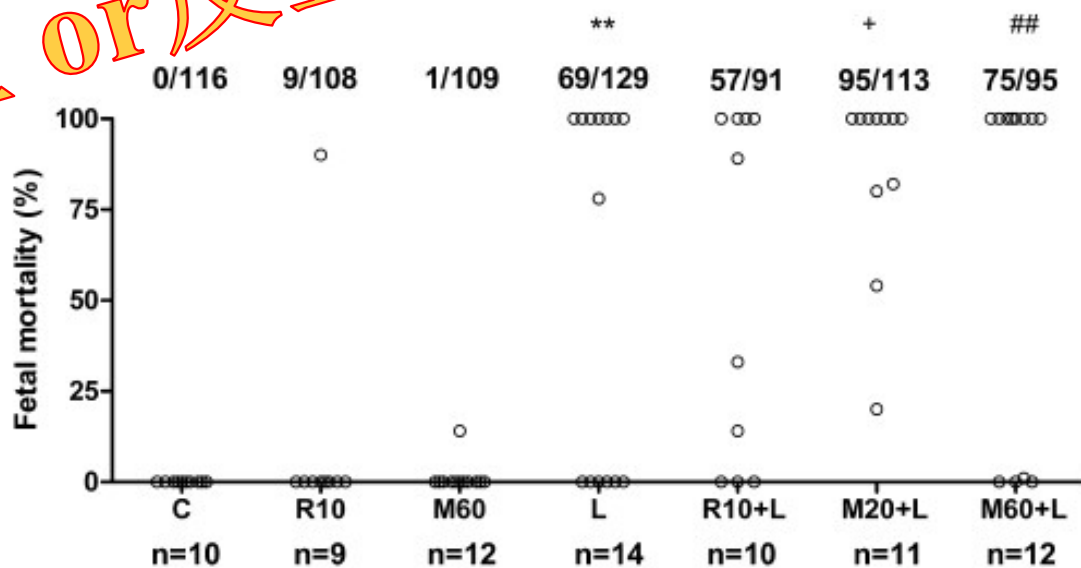




思考

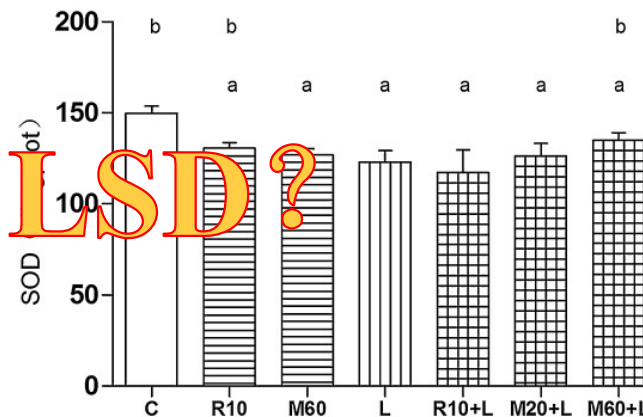
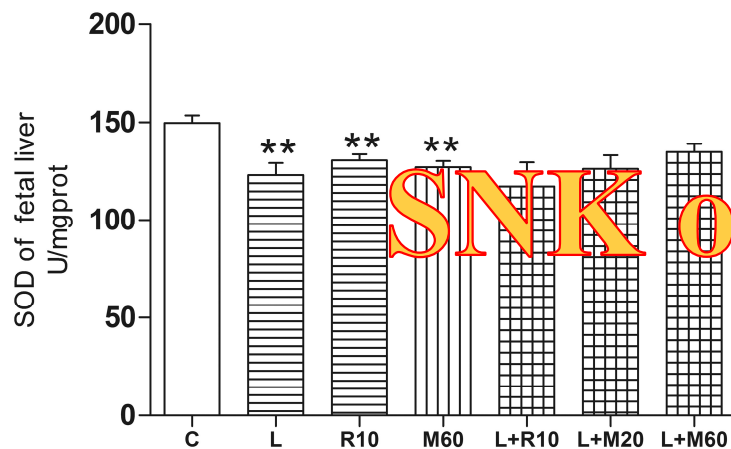


X<sup>2</sup>检验 or 反正弦平方根?





思考



SNK or LSD?

酶活性				
分组	N	alpha = 0.05 的子集		
		1	2	3
Student-Newman-Keuls <sup>a</sup>				
Mother	3	.620533		
Female	3		3.546667	
Male	3			4.419333
显著性		1.000	1.000	1.000

将显示同类子集中的组均值。

a. 将使用调和均值样本大小 = 3.000。







## 总结

方差齐时，两两比较选择：

**Dunnett-t** 和固定一个组比，如全部和对照组相比

**SNK、Turkey** 任意组间两两比较

**Bonferroni** 检验水准调整为 $\alpha'=\alpha/n-1$ ，避免犯I类错误概率

**LSD** 最敏感，最易出统计学意义，文章中最少用

方差不齐时，两两比较选择：

**Tamhane's T2, Dunnett's T2, Games-Howell, Dunnett's C**都是用于方差不齐时。

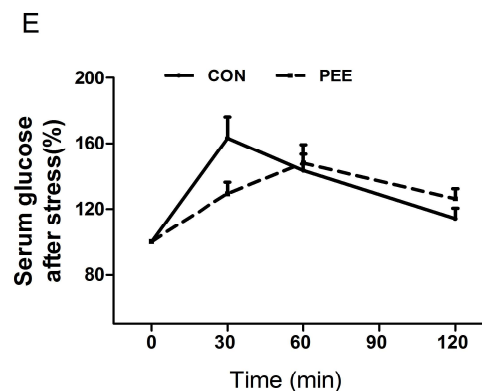
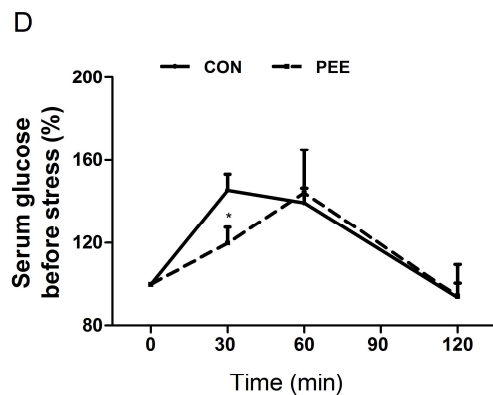
**Games-Howell**是最powerful的方法，而且在组间样本量不一致的时候也能保证正确性，唯一的问题是当样本量太小的时候容易显著。其它3个都很保守，注意控制I类错误概率。





## 总结

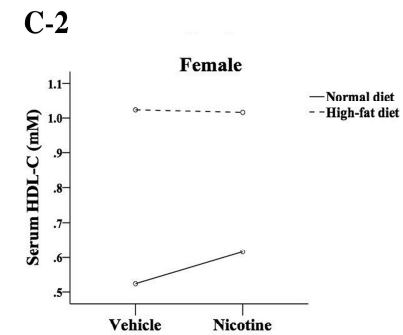
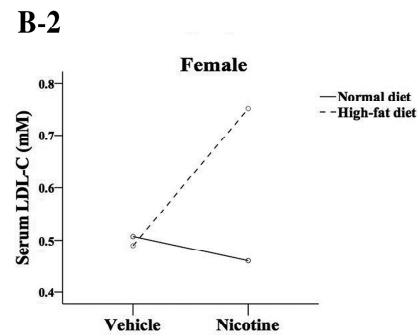
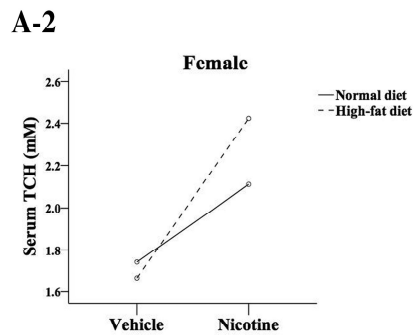
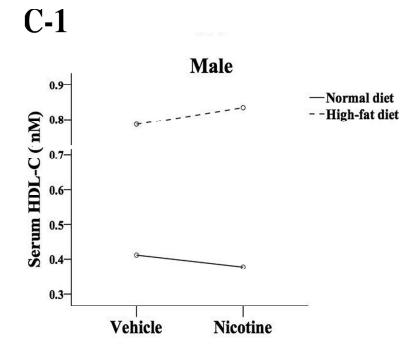
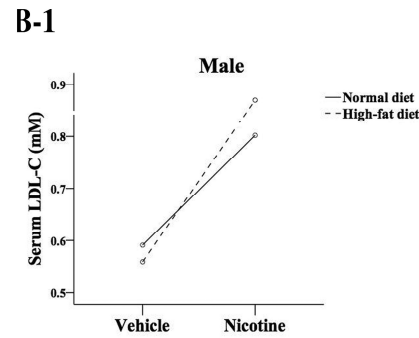
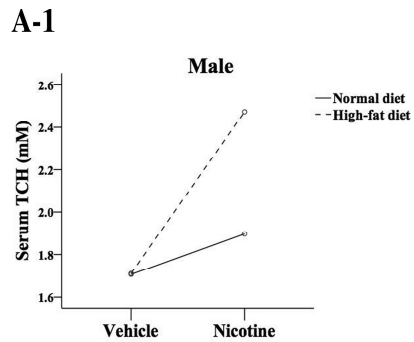
- 两因素方差分析（析因分析）可以用来分析两个因素的不同水平对结果是否有显著影响，以及两因素之间是否存在交互效应。
- 重复测量指同一受试对象的某一观察指标在不同时间点上进行多次测量的检验，常用来分析不同处理在不同时间点上的变化情况。





# 总结

三因素方差分析与两因素基本相同，但当三个因素的析因分析结果存在交互效应时，单独效应的分析比较繁琐，如果要分析某一个因素的单独效应，需要固定其他两个因素的所有不同组合。





## 其他统计分析

- Poisson分布
- 多分类Logistic回归分析

表 2 专业喜爱程度影响因素的多元 Logistic 回归结果

因素	B	标准误	Wald	P值	OR值(OR值95%CI)
就读原因(以专业调剂为对照)					
自身兴趣	1.825	0.842	4.702	0.030	6.202(1.192~32.271)
学习方法(以听课自学为参照)					
仅靠听课	-2.193	0.927	5.598	0.018	0.112(0.018~0.686)
仅靠自学	-1.529	0.912	2.812	0.094	0.217(0.036~1.294)
课程设置(以不合理为参照)					
合理	2.274	1.343	2.867	0.090	9.719(0.699~135.131)
部分合理	2.065	0.908	5.178	0.023	7.886(1.332~46.705)

注：对专业喜爱程度参考类别为不喜欢，各因素对专业无所谓与不喜欢的影响无明显统计学差异，故未在表格中显示。





## 实例分析

- 醉酒实验中一般根据动物入睡时间和睡眠时间来判断醉酒程度，但有的动物不睡，有的动物睡不醒，可用什么方法进行统计？





*Thank You*

