讲座

简明生物统计方法

王广仪

（吉林医科大学）

第三讲 计量资料的统计推断方法（上）

一、用样本平均数估计总体平均数

用样本平均数 ($\bar{x}$) 估计总体平均数 (μ)，亦常用区间估计法；即根据 $\bar{x}$ 求出 μ 的 95%（或）99% 置信区间。

（一）当样本含量较大 ($n > 30$)，不拘原变量 (x) 的总体分布如何，均可用下式求出置信区间；但如样本较小 ($n < 30$)，则原变量必须服从正态分布，方可用下式计算：

$$\mu_L = \bar{x} - t_{0.95}(\text{或 } t_{0.99})\, S_{\bar{x}} \qquad (3.1)$$

$$\mu_u = \bar{x} + t_{0.95}(\text{或 } t_{0.99})\, S_{\bar{x}} \qquad (3.2)$$

式中 μ_L，μ_u——总体平均数的 95%（或 99%）置信区间的下、上限；

$\bar{x}$——样本平均数，$\bar{x} = \frac{\Sigma x_i}{n}$

$S_{\bar{x}}$——样本标准误差，$S_{\bar{x}} = \sqrt{\frac{\Sigma x_i^2 - (\Sigma x_i)^2/n}{n(n-1)}}$

$t_{0.95}$（或 $t_{0.99}$）与 $t_{0.05}$（或 $t_{0.01}$）同——自由度 $df = n - 1$，$\alpha = 0.05$（或 0.01）的 t 值

例 3-1 随机测得 25 名成年男性（健康者）血浆量（毫升/公斤体重）如下，试估计其总体血浆量平均数。

	46.3	40.6	54.7	46.6	48.7
	52.6	45.4	49.1	49.2	47.5
	46.4	54.1	48.7	48.4	49.0
	44.7	50.4	45.2	50.6	44.8
	52.4	51.8	53.7	48.2	52.5
全距 (R)：	7.9	13.5	9.5	4.0	7.7

本例因原变量基本服从正态分布（可用正态概率纸测定）[1)]，用式 (3.1)、(3.2) 计算总体平均血浆量的 95% 置信区间如下：

$$\bar{x} = (46.3 + 52.6 + \cdots + 52.5)/25 = 48.9$$

$$S_{\bar{x}} = \sqrt{\frac{59,978.9 - (1,221.6)^2/25}{25(25-1)}} = 0.69$$

代入式 (3.1)、(3.2)，

$$\mu_L = 48.9 - 2.064^{2)} \times 0.69 \doteq 47.5$$

$$\mu_u = 48.9 + 2.064 \times 0.69 = 50.3$$

结论：健康成年男性的总体平均血浆量在 47.5 与 50.3（毫升/公斤体重）之间。

（二）为减轻计算负担，如果原变量呈正态分布（如例 3-1），可用"全距"(R)（R = 最大值 − 最小值）估计标准差 (σ) 的方法计算置信区间，与用 (3.1)、(3.2) 式计算结果非常接近。方法如下：

1. 如样本含量 $n > 10$，则把原变量 (x) 随机等分成若干组，使每组的变量个数不超过 10（如果不能等分时，可随机舍弃几个变量）。例 3-1 恰好能随机等分成五组 ($K = 5$)，每组 5 个变量 ($n_j = 5$)，并求出每组变量的"全距"，见例 3-1 的变量数列。

2. 求出"全距"和 (ΣR)：

$$\Sigma R = 7.9 + 13.5 + 9.5 + 4.0 + 7.7 = 42.6$$

3. 代入下式求出置信区间：

$$\mu_L = \bar{x} - A_{0.95}(\text{或 } A_{0.99})\, \Sigma R \qquad (3.3)$$

$$\mu_u = \bar{x} + A_{0.95}(\text{或 } A_{0.99})\, \Sigma R \qquad (3.4)$$

式中 $\bar{x}$——样本平均数（计算方法同前）；

$A_{0.95}$（或 $A_{0.99}$）——查附表 1，n_j K，P 所对应之数值。本例 $n_j = 5$，$K = 5$，$P = 0.95$ 所对应之 $A_{0.95} = 0.035$。

对于例 3-1

$$\mu_L = 48.9 - 0.035 \times 42.6 = 47.4$$

$$\mu_u = 48.9 + 0.035 \times 42.6 = 50.4$$

结论同前。

对总体平均数的估计，也有用 $\bar{x} \pm S_{\bar{x}}$ 形式表示的，和置信区间同义。

二、样本平均数的差异显著性试验

根据统计调查或实验结果获得数据的实际情况，常见的样本平均数差异显著性测验，一般可分为：差数平均数、两个（或多个）样本平均数、两组（或多组）样本平均数的差异显著性测验等几种，其测验方法分别介绍如下：

---

1) 见参考资料 [1]，209 页

2) 查 t 值表，$df = 25 - 1$，$\alpha = 0.05$ 的 t 值。见参考资料 [2]，8 页

（一）差数平均数的显著性测验方法

本测验法是用以分析自身比较（或称自身对照）的实验结果的。例如同一批动物在某种处理前、后的测量结果，即为自身对照（不另设对照组），见下例。

例 3-2 用豚鼠肺作支气管灌流实验，用药前、后灌流液经支气管每分钟流出的滴数见表 3-1；问用药前、后的流出滴数有无显著差异？

表 3-1 灌流液经支气管流出量（滴数/分）及其“差数”的顺序号

豚鼠号	用药前	用药后	差数（d=前—后）	顺序号（R_i）
1	30	46	−16	−10
2	38	50	−12	−9
3	48	52	−4	−3
4	48	52	−4	−3
5	60	58	2	1
6	46	64	−18	−11
7	26	56	−30	−12
8	58	54	4	3
9	46	54	−8	−6
10	48	58	−10	−8
11	44	36	8	6
12	46	54	−8	−6
			$\bar{d}=-8$	$R.=10$

1．顺序测验法[1]

（1）把表 3-1 用药前减用药后的“差数”（d），按绝对值大小排列编号，并按原差数的正负号标记；遇有相同数据且分列正、负两边者（如本例有三个“4”和三个“8”），则取其顺序号的平均数［如 2，3，4 号的平均数为 $(2+3+4)/3=3$，余类推］作为各相同数据的顺序号（见表 3-1“顺序号”栏）。

（2）查附表 2 判定测验结果。将顺序号按正负分别相加，取其和（$R.$）的绝对值小者，与附表 2 中 n 和 $P=0.05$（或 $P=0.01$）所对应之数值比较，$R.$ 小时便为 $P<0.05$（或 $P<0.01$）。本例 $R.=10$，它小于 $n=12$ 和 $P=0.05$ 所对应之数值 14，$P<0.05$，差异显著。结论：用药后较用药前，流出量（滴数/分）有明显增加。

2．t 测验法

差数平均数的 t 测验公式为：

$$t=\frac{|\bar{d}-\mu_d|}{S_{\bar{d}}} \qquad (3.5)$$

$$df=n-1$$

式中 $\bar{d}$——差数平均数（计算公式及含义同 $\bar{x}$）；

$S_{\bar{d}}$——差数（d）的标准误差（计算公式及含义同 $S_{\bar{x}}$）；

μ_d——差数（d）的总体平均数，根据无效假设应为 0。

对于例 3-2，

$$\bar{d}=(-16-12+\cdots-8)/12=-8$$

$$S_{\bar{d}}=\sqrt{\frac{1,968-(96)^2/12}{12(12-1)}}=3.02$$

$$t=\frac{|-8-0|}{3.02}=2.65$$

$$df=12-1=11$$

查 t 值表，$t=2.65>t_{0.05(df=11)}=2.201$[2]，$P<0.05$，差异显著。结论同前。

3．本例亦可用“全距”求置信区间法进行显著性测验。方法如下：

（1）先将“差数”（d）随机等分成 3 组，并求出 ΣR：

	−16	2	−8
	−12	−18	−10
	−4	−30	8
	−4	4	−8
全距（R）：	12	34	18

$$\Sigma R=12+34+18=60$$

（2）查附表 1，$n_j=4$，$K=3$，$P=0.95$ 所对应之 $A_{0.95}=0.104$，则有

$$A_{0.95}\Sigma R=0.104\times 60=6.24$$

如果差数平均数（$\bar{d}$）的绝对值大于 $A_{0.95}\Sigma R$（或 $A_{0.99}\Sigma R$）时，便为 $P<0.05$（或 $P<0.01$）。本例，差数平均数绝对值 $|-8|>6.24$，$P<0.05$，差异显著。结论同前。

（二）两个样本平均数的差异显著性测验

两个样本如系配对分组，则适用（一）所介绍的方法。这里指的两个样本（包括实验组与对照组或两个实验组）系随机分组，样本含量（n_1，n_2）可以相等，也可以不等。

例 3-3 两组雌鼠分别给以高蛋白和低蛋白饲料，8 周后（自生后 28 天至 84 天）测量所增体重见表 3-2；问两组所增体重有无显著差异？

1．顺序测验法[3]

方法同“两列样本百分率差异显著性测验”所用之顺序测验法。

（1）排列编号，见表 3-2“顺序号栏”。

（2）查表判定显著性测验结果。求出 $R_{.1}$（或 $R_{.2}$）即可查附表 3，$K=2$，$n_1=7$，$\alpha=0.05$ 所对应之数值为：37.2(1.975)；67.8(5.025)，换算[4]后得：

1) 见参考资料［3］，402 页。
2) 见参考资料［2］，8 页。
3) 见参考资料［3］，405 页。
4) 见附表 3 说明之（2）。

表 3-2 两组雌鼠 8 周后所增体重(克)及其顺序号

低蛋白组 (x_{i1})	高蛋白组 (x_{i2})	顺序号*	
		R_{i1}	R_{i2}
70	83	1	2
85	97	3	5
94	104	4	7
101	107	6	8.5
107	113	8.5	10
118	119	11	12
132	123	16	13
	124		14
	129		15
	134		17
	146		18
	161		19
$\bar{x}_{\cdot 1}=101$	$\bar{x}_{\cdot 2}=120$	$R_{\cdot 1}=49.5$	$R_{\cdot 2}=140.5$

* $R_{\cdot 1}$, $R_{\cdot 2}$ 可用 $R_{\cdot 1}+R_{\cdot 2}=\frac{(n_1+n_2)(n_1+n_2+1)}{2}$

加以验证本例，$n_1=7$，$n_2=12$

$$49.5+104.5=\frac{(7+12)(7+12+1)}{2}=190$$

$R_{\cdot j;0.05(L)}=47.1$； $R_{\cdot j;0.05(u)}=92.9$， 本例 $R_{\cdot 1}=49.5$，介于 47.1 与 92.9 之间，$P>0.05$，差异不显著。结论：尚不能断定两组雌鼠所增体重有显著差别。

2．t 测验法

两个样本平均数的 t 测验公式：

$$t=\frac{|\bar{x}_{\cdot 1}-\bar{x}_{\cdot 2}|}{S_{(\bar{x}_{\cdot 1}-\bar{x}_{\cdot 2})}} \quad (3.6)$$

式中

$$S_{(\bar{x}_{\cdot 1}-\bar{x}_{\cdot 2})}=\sqrt{\frac{\Sigma x_{i1}^2-(\Sigma x_{i1})^2/n_1+\Sigma x_{i2}^2-(\Sigma x_{i2})^2/n_2}{n_1+n_2-2}\times\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}$$

$$df=n_1+n_2-2$$

对于例 3-3，

$$S_{(\bar{x}_{\cdot 1}-\bar{x}_{\cdot 2})}=\sqrt{\frac{73{,}959-(707)^2/7+177{,}832-(1{,}440)^2/12}{7+12-2}\times\left(\frac{1}{7}+\frac{1}{12}\right)}=10.04$$

$$t=\frac{|101-120|}{10.04}=1.89$$

$$df=7+12-2=17$$

$t=1.89<t_{0.05(df=17)}=2.110$[1]，$P>0.05$，差异不显著。结论同前。

当原变量 (x) 偏离正态分布较远，且样本含量较小时，不宜直接应用 t 测验法；需先经变量代换，使其接近正态分布，然后再用。另当两总体标准差有明显差异 ($\sigma_1 \neq \sigma_2$) 时，亦不能应用 t 测验法，需应用近似的方法。但不拘原变量情况如何，皆适用“顺序测验法”。如样本含量过少时，可将“顺序号” (R_{ij}) 进行标准正态代换[2]，能相应提高测出效率。

附表 1 "全距"法计算均数置信区间用表

K	P	n_j								
		2	3	4	5	6	7	8	9	10
1	.95	6.36	1.30	0.719	0.505	0.402	0.336	0.291	0.256	0.232
	.99	31.92	3.00	1.360	0.865	0.673	0.514	0.430	0.379	0.338
2	.95	0.879	0.316	0.206	0.154	0.125	0.106	0.093	0.084	0.076
	.99	2.110	0.474	0.312	0.227	0.179	0.150	0.131	0.116	0.105
3	.95	0.360	0.156	0.104	0.079	0.065	0.056	0.049	0.044	0.040
	.99	0.660	0.273	0.150	0.112	0.091	0.077	0.068	0.060	0.054
4	.95	0.210	0.096	0.065	0.050	0.042	0.036	0.032	0.028	0.026
	.99	0.350	0.142	0.092	0.070	0.057	0.048	0.043	0.038	0.035
5	.95	0.140	0.066	0.046	0.035	0.030	0.025	0.022	0.020	0.018
	.99	0.226	0.092	0.063	0.049	0.040	0.034	0.030	0.027	0.025

1) 见参考资料 [2]，8 页。
2) 见参考资料 [4]，14 页。

附表 2　差数平均数顺序测验法用表

n	$P=0.05$	$P=0.01$
6	0	—
7	2	—
8	4	0
9	6	2
10	8	3
11	11	5
12	14	7
13	17	10
14	21	13
15	25	16

$n>15$ 时，用下式算出表内数值：

$$\frac{n(n+1)}{4}-1.96\,(\text{或}\,2.58)\sqrt{\frac{n(n+1)(n+1/2)}{12}}$$

附表 3　完全随机分组样本平均数差异显著性顺序测验法用表示例

$$P\{R_{\cdot j;\,\alpha(L)}<R_{\cdot j;}<R_{\cdot j;\,\alpha(u)}\}=1-\alpha$$

n_j	α	$K=2$	$K=3$	$K=4$	$K=5$
6	0.05	26.8(1.590); 51.2(4.410)	25.1(1.413); 52.8(4.589)	24.3(0.984); 53.7(4.690)	23.8(1.245); 54.2(4.755)
	0.01	22.9(1.145); 55.1(4.815)	21.6(0.994); 56.4(5.007)	20.8(0.900); 57.3(5.098)	20.3(0.842); 57.7(5.158)
7	0.05	37.2(1.975); 67.8(5.025)	35.2(1.781); 69.8(5.219)	34.1(1.672); 70.9(5.328)	33.4(1.602); 71.6(5.398)
	0.01	32.3(1.493); 72.7(5.507)	30.7(1.330); 74.3(5.671)	29.6(1.234); 75.4(5.772)	29.0(1.166); 76.0(5.834)
8	0.05	49.3(2.368); 86.7(5.632)	47.0(2.160); 89.0(5.840)	45.6(2.043); 90.4(5.957)	44.8(1.968); 91.2(6.032)
	0.01	43.4(1.851); 92.6(6.148)	41.4(1.676); 94.6(6.324)	40.2(1.568); 95.8(6.432)	39.4(1.502); 96.6(6.498)
9	0.05	63.3(2.767); 107.7(6.467)	60.5(2.545); 110.5(6.453)	58.9(2.422); 112.1(6.578)	57.9(2.342); 113.1(6.658)
	0.01	56.3(2.218); 114.7(6.782)	53.9(2.032); 117.1(6.968)	52.4(1.917); 118.6(7.083)	51.5(1.847); 119.5(7.153)
10	0.05	79.1(3.171); 130.9(6.829)	75.8(2.937); 134.2(7.063)	73.9(2.807); 136.1(7.194)	72.7(2.723); 137.3(7.278)
	0.01	70.9(2.592); 139.1(7.408)	68.1(2.396); 141.9(7.604)	66.4(2.274); 143.6(7.726)	65.3(2.200); 144.7(7.800)

附表 3 说明：

(1) 本表系根据下式算出：

$$R_{\cdot j;\,0.05(\text{或}0.01)(L)}=\mu_{(R\cdot j)}-t_{0.05(\text{或}0.01)}\sigma_{(R\cdot j)};$$

$$R_{\cdot j;\,0.05(\text{或}0.01)(u)}=\mu_{(R\cdot j)}+t_{0.05(\text{或}0.01)}\sigma_{(R\cdot j)}$$

式中

$$\mu_{(R\cdot j)}=\frac{n_j(n+1)}{2},\quad n=\sum_{j=1}^{K}n_j$$

$$\sigma_{(R\cdot j)}=\sqrt{\frac{\mu_{(R\cdot j)}(n-n_j)}{6}}$$

$t_{0.05(\text{或}0.01)}$——查本文附表 4

(2) 查表方法举例。如本文例 3-3，$n_j=8$，$K=4$，本应直接由表内查出 $R_{\cdot j;0.05(\text{或}0.01)(L)}$ 与 $R_{\cdot j;0.05(\text{或}0.01)(u)}$ 的数值，但因制表篇幅限制，表内通过下式给出近似值，即

$$R_{\cdot j;0.05(l)}=45.6(2.043);$$

$$R_{\cdot j;0.05(u)}=90.4(5.957)。$$

实即，

$$\begin{aligned}R_{\cdot j,0.05(L)}&=45.6+2.043(n-2n_j)\\&=45.6+2.043(32-2\times 8)\\&=78.3\end{aligned}$$

$$\begin{aligned}R_{\cdot j,0.05(u)}&=90.4+5.957(n-2n_j)\\&=90.4+5.957(32-2\times 8)\\&=185.7\end{aligned}$$

余类推。

(3) 表内数值略有误差，如遇 $R_{\cdot j}$ 正靠近临界值，则以按 (1) 的公式计算结果为准。

(4) 遇有较多同等级"顺序号"(如百分率的顺序测验时)，计算 $\sigma_{(R\cdot j)}$ 应用下例校正式*：

$$\sigma_{(R\cdot j)}=\sqrt{\frac{n_j(n-n_j)}{n(n-1)}\left(\frac{n(n-1)(n+1)-\Sigma[f(f-1)(f+1)]}{12}\right)}$$

式中　f——相同等级顺序号个数

* 校正式不但计算繁冗，且不宜制便查表，经过大量抽样实验结果说明不用校正式亦不影响预期测验效果

(下转第 40 页)

配制凝胶所用的几种试剂的纯度，对电泳分离效果影响颇大。故两种单体均须重结晶后再用[3]。缓冲液和染液中如有不溶物存在，皆应先行过滤除去。试液(1)、(2)、(3)、(4)应贮放于低温避光处，但不宜放置过久(不超过二个月)。过硫酸铵则应每周新鲜配制。

3. 缓冲系统的影响

我们曾比较过两种缓冲液系统(0.4*M* Tris-0.2*M* NaAc-0.02*M* Na_2-EDTA, pH 7.8 及 0.89*M* Tris-0.89*M* 硼酸-0.025*M* Na_2-EDTA pH 8.3) 使用前者，得到的色带松散，模糊变宽；用后者，则色带窄细清晰，提高了分辨效果。

4. RNA 样品纯度

RNA 样品中未除尽的杂质(如蛋白质，多糖等)皆可造成区带模糊，拖尾，底色洗不清，甚至有的带被歪曲等现象。因此，改进样品提取方法，努力提高其纯度是十分重要的。

5. 铺样量

如果铺样量过多或体积过大，都会影响分离效果。我们认为，每个凝胶柱铺样 10—60 微克，体积 5—30 微升为宜。样品纯度高时，铺样量可适当放大，仍能取得好的分离效果。

其它，如电泳需在低温环境下进行，电流电压要稳定以及精细的操作等，对于得到好的能重复的结果，也是必要的。

参 考 资 料

[1] Slater, G.: *Fed. Proc*., **24**, 225, 1965.

[2] Dale, G. et al.: *Lancet*., **1**, (7547), 847, 1968.

[3] Loening, U. E.: *Biochem. J*., **102**, 251, 1967.

[4] Peacock, A. C. et al.: *Biochem*., **7**, (2), 668, 1968.

[5] Watcher, R. et al.: Methods in Enzymol., **XXI**, part D, p. 167.

[6] Pollard, D. et al.: *Anal. Biochem*., **42**, 286, 1971.

[7] Zubay, G.: *J. Mol. Biol*., **4**, 347, 1962.

[8] Kirby, K.: Methods in Enzymol., **XII**, Nucleic Acid, part B, p. 87.

[9] Bishop, D. H. L. et al.: *J. Mol. Biol*., **26**, 373, 1967.

[10] Athuton, K. et al.: *Anal. Biochem*., **57**, 403, 1974.

[本文于 1975 年 3 月 15 日收到]

(上接第 57 页)

五、结 束 语

血红蛋白的分子杂交技术虽然建立不久，但它已广泛地应用于这种蛋白质的许多研究领域，得到不少成果。分子杂交开始于血红蛋白，现在已延伸到其他蛋白质、酶和核酸。不过，我们对血红蛋白分子杂交的认识并没有完结。目前看来，血红蛋白不仅在人工条件下可以杂交，就是在自然的生理条件下也能发生亚单位的交换，它们的过程和关系可归纳成 (7) 式：

$$
\begin{array}{ccccccc}
 & & 2\alpha\beta\alpha'\beta'\text{……天然杂交分子} & & & & \\
 & & \Updownarrow & & & & \\
\alpha_2\beta_2 & \rightleftharpoons & 2\alpha\beta & \rightleftharpoons & 2\alpha & + & 2\beta \\
 & & + & & + & & + \\
\alpha'_2\beta'_2 & \rightleftharpoons & 2\alpha'\beta' & \rightleftharpoons & 2\alpha' & + & 2\beta' \\
 & & & & & \Updownarrow & \\
 & & & & & \alpha'_2\beta_2+\alpha_2\beta'_2\text{……人工杂交分子} &
\end{array}
\tag{7}
$$

人工分子杂交的机制尚未完全弄清，但是，可以相信，随着分子杂交理论的深入和其应用面的扩大，这一问题会很快得到解决。天然分子杂交是最近才逐步产生的新知识，估计今后它会解决更多理论和实践中的问题。

(上接封3)

附表 4 *K* 组(包括对照组)比较用 *t* 值表 ($df \to \infty$)

	$t_{0.05}$	$t_{0.01}$
2	1.96	2.58
3	2.21	2.79
4	2.35	2.92
5	2.44	3.00
6	2.51	3.06
7	2.57	3.11
8	2.61	3.15
9	2.65	3.19
10	2.69	3.22

参 考 资 料

[1] 中国科学院数学研究所统计组编：常用数理统计方法，科学出版社，1973。

[2] 中国科学院数学研究所概率统计室编：常用数理统计表，科学出版社，1974。

[3] Steel, R. G. D. & Torrie, J. H.: Principles and procedures of statistics, 1960.

[4] 王广仪：简易医用数理统计方法初稿，吉林省医学科学情报室、吉林医大科研处编：《学术活动资料》，1964。

(待续)