

高等医药院校配套教材

卫生统计学习题解答

何大卫 主编

人民卫生出版社

卫生统计学习题解答

R195.1/HDM

生
出
版
社

卫生统计学习题解答

何大卫 主编

杨树勤 主审

编者（按章节顺序）

王琳娜 刘桂芬 郭明英 何大卫

余红梅 武俊青 仇丽霞

人 民 卫 生 出 版 社

图书在版编目 (CIP) 数据

卫生统计学习题解答/何大卫主编. - 北京: 人民卫生出版社, 1997

ISBN 7-117-02788-6

I. 卫… II. 何… III. 卫生学: 统计学 - 高等学校 - 习题
IV. R195.1 - 44

中国版本图书馆 CIP 数据核字 (97) 第 18799 号

2688/01

卫生统计学习题解答

何大卫 主 编

人民卫生出版社出版发行
(100050 北京市崇文区天坛西里 10 号)

人民卫生出版社印刷厂印刷

新华书店经销

850×1168 32 开本 $5\frac{1}{8}$ 印张 139 千字
1997 年 9 月第 1 版 1997 年 9 月第 1 版第 1 次印刷
印数: 00 001—5 000

ISBN 7-117-02788-6/R · 2789 定价: 7.50 元

(凡属质量问题请与本社发行部联系退换)

前 言

实习题是《卫生统计学》(第三版)教材的重要组成部分,学习卫生统计学必须亲自动手完成一定数量的习题,才能较好地掌握卫生统计学的基本概念、原理和方法,有助于培养统计思维方法与分析问题和解决问题能力。提高实习课的教学质量是提高卫生统计学教学质量的关键,为此,早在80年代末,为了深入地学习教材,提高实习课质量,我室曾编写了《卫生统计学》(第二版)习题解答作为内部资料,供教师、学生参考,取得较好效果。《卫生统计学》(第三版)教材出版后,我们把编写第三版习题解答作为集体备课内容,采取分工编写答案,再集体讨论定稿方式编写了这本与《卫生统计学》(第三版)教材配套的习题解答。本《解答》对书中各类习题(计算分析题、最佳选择题、思考题、论述题)全部做了答案,并以《卫生统计学》(第三版)教材所讲述的理论、方法为依据,使之与教材所讲述的内容相呼应,有助于读者加深对教材的理解。个别习题有两种及两种以上解法时,则选择最贴近第三版教材的解法作为答案。

建议初学者请勿死记本《解答》。正确的途径是读者在学习教材的基础上,经认真思考后,亲手做习题,再与答案对照。这样才能深入理解卫生统计学的基本原理和概念,掌握卫生统计方法。

我国著名卫生统计学专家、《卫生统计学》主编、华西医科大学杨树勤教授对本《解答》进行了审定。在编写过程中,得到了本教研室同志的关心、支持与帮助,苏景铭教授及硕士研究生王彤、张晋昕、郜艳晖、闫玉霞、郭静、闫强等参加了讨论并提出了宝贵意见。王琳娜和张岩波同志在计算机上完成了本《解答》稿的编辑、修改。由张岩波同志担任秘书工作并协助绘图、校对,谨致以衷心的感谢。本《解答》引用了第二版教材《解

答》的部分资料，谨向第二版《解答》编者致谢。

参加本《解答》编写的全体同志，力求提高质量，但由于水平所限，一定还有不少缺点、错误，欢迎广大读者批评指正。

编 者

1996年9月于山西医科大学

目 录

第一单元 数值变量资料 (计量资料) 的统计描述	
(第一~三章)	1
计算分析题	1
最佳选择题	8
思考题	9
第二单元 数值变量资料 (计量资料) 的统计推断	
(第四、五章)	15
计算分析题	15
讨论题	35
最佳选择题	38
思考题	39
第三单元 分类资料的统计描述与统计推断	
(第六~八章)	44
计算分析题	44
讨论题	60
最佳选择题	65
思考题	65
第四单元 秩和检验 (第九章)	70
计算分析题	70
最佳选择题	76
讨论题	77
思考题	79
第五单元 直线回归与相关 (第十章)	82
计算分析题	82
最佳选择题	89
讨论题	90
思考题	92

第六单元 统计表与统计图 (第十一章)	96
列表、制图与分析题	96
思考题	101
第七单元 调查设计与实验设计	
(第十二、十三章)	104
设计与讨论题	104
讨论题	110
思考题	121
第八单元 健康统计 (第十四~十六章)	126
计算分析题	126
最佳选择题	148
思考题	151

第一单元 数值变量资料 (计量资料) 的统计描述 (第一~三章)

计算分析题

1.1 某地 101 例 30~49 岁健康男子血清总胆固醇值 (mmol/L) 测定结果如下:

4.77 3.37 6.14 3.95 3.56 4.23 4.31 4.71 5.69 4.12 4.56 4.37 5.39
6.30 5.21 7.22 5.54 3.93 5.21 6.51 5.18 5.77 4.79 5.12 5.20 5.10
4.70 4.74 3.50 4.69 4.38 4.89 6.25 5.32 4.50 4.63 3.61 4.44 4.43
4.25 4.03 5.85 4.09 3.35 4.08 4.79 5.30 4.97 3.18 3.97 5.16 5.10
5.86 4.79 5.34 4.24 4.32 4.77 6.36 6.38 4.86 5.55 3.04 4.55 3.35
4.87 4.17 5.85 5.16 5.09 4.52 4.38 4.31 4.58 5.72 6.55 4.76 4.61
4.17 4.03 4.47 3.40 3.91 2.70 4.60 4.09 5.96 5.48 4.40 4.55 5.38
3.89 4.60 4.47 3.64 4.34 5.18 6.14 3.24 4.90 3.05

(1) 编制频数分布表并绘制直方图, 简述其分布特征。

1) 找出最大值、最小值, 求全距 (R):

全距 = 最大值 - 最小值 = $7.22 - 2.70 = 4.52$ (mmol/L)

2) 求组距: $i = \text{全距} / \text{组数} = 4.52 / 10 = 0.452 \approx 0.5$ (mmol/L)

3) 分组段, 划记 (表 1-1):

表 1-1 某地 101 例 30~49 岁健康男子血清
总胆固醇值划记表

组段 (mmol/L)	划 记	频 数
2.5~	—	1
3.0~	正下	8
3.5~	正正	9
4.0~	正正正正下	23
4.5~	正正正正正	25
5.0~	正正正正下	17
5.5~	正正	9
6.0~	正—	6
6.5~	下	2
7.0~7.5	—	1
合 计		101

由表 1-1 及图 1-1 可知, 本例频数分布中间居多, 两侧逐渐减少, 左右基本对称。

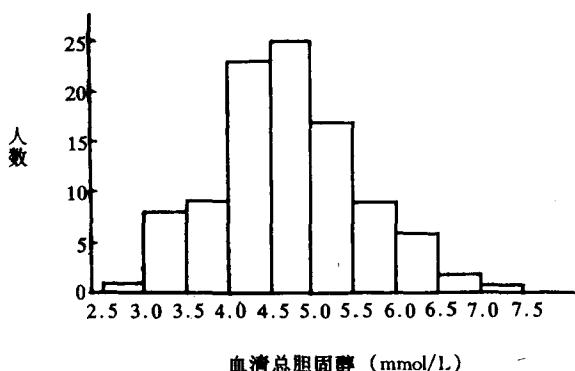


图 1-1 某地 101 例 30~49 岁健康男子血清总胆固醇值 (mmol/L) 分布

表 1-2 某地 101 例 30~49 岁健康男子血清总胆固醇值 (mmol/L) $\bar{X}$ 、 s 计算表

血清总胆固醇值	组中值 X	频数 f	fX	fX^2	累计频数	累计频率 (实际)	$u = (X_u - \bar{X})/s$	累计频率 (理论)
2.5~	2.75	1	2.75	7.563	1	0.0099	-1.97	0.0244
3.0~	3.25	8	26.00	84.500	9	0.0891	-1.40	0.0808
3.5~	3.75	9	33.75	126.563	18	0.1782	-0.83	0.2033
4.0~	4.25	23	97.75	415.438	41	0.4059	-0.27	0.3936
4.5~	4.75	25	118.75	564.063	66	0.6535	0.30	0.6179
5.0~	5.25	17	89.25	468.563	83	0.8218	0.87	0.8078
5.5~	5.75	9	51.75	297.563	92	0.9109	1.43	0.9236
6.0~	6.25	6	37.50	234.375	98	0.9703	2.00	0.9772
6.5~	6.75	2	13.50	91.125	100	0.9901	2.56	0.9948
7.0~7.5	7.25	1	7.25	52.563	101	1.0000	3.13	1.0000
			478.25	2242.315				

注: X_u 为组段上限值

(2) 计算均数 $\bar{X}$ 、标准差 s 、变异系数 CV 。

由上计算表 1-2 可见: $\bar{X} = \Sigma fX / \Sigma f = 478.25 / 101 = 4.735$ (mmol/L)

$$s = \frac{\sqrt{\sum fX^2 - (\sum fX)^2 / \sum f}}{\sqrt{\sum f - 1}}$$

$$= \frac{\sqrt{2342.313 - (478.25)^2 / 101}}{\sqrt{101 - 1}} = 0.882 \text{ (mmol/L)}$$

$$CV = s / \bar{X} \times 100\% = 0.882 / 4.735 \times 100\% = 18.627\%$$

(3) 就所得频数表，计算累积（实际）频率，并查附表 1 得相应的累积理论频率 $\Phi(u)$ ，分别在方格坐标纸上绘制出两条累积频率曲线，观察两者的吻合度。

结果见表 1-2 与图 1-2。由图可见，实际频率与理论频率两条曲线基本吻合，可以认为本资料服从正态分布。

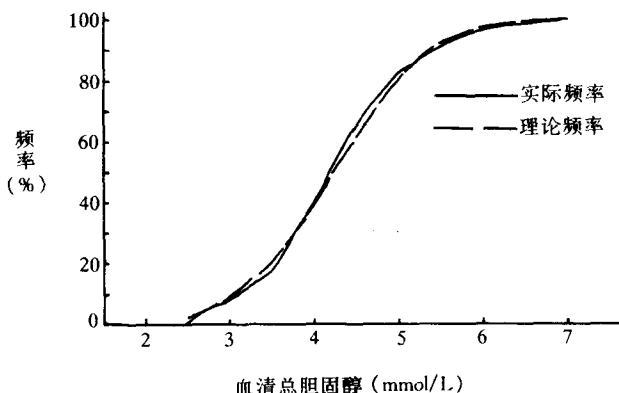


图 1-2 某地 101 例 30~49 岁健康男子血清总胆固醇值累积频率 (%)

(4) 计算中位数 M ，并与均数 $\bar{X}$ 比较。利用前表计算中位数 M

$$M = L + (i / f_{50}) (n \times 50\% - \sum f_L)$$

$$= 4.5 + (0.5 / 25) (101 \times 50\% - 41) = 4.69 \text{ (mmol/L)}$$

本题算术均数为 4.735 (mmol/L)，与中位数 4.69 (mmol/L) 很接近。这也是资料服从正态分布的特征之一。

(5) 计算 $P_{2.5}$ 及 $P_{97.5}$ ，并与 $\bar{X} \pm 1.96s$ 的范围比较。

$$P_{2.5} = 3.0 + (0.5 / 8) \times (101 \times 2.5\% - 1) = 3.095 \text{ (mmol/L)}$$

$$P_{97.5} = 6.5 + (0.5 / 2) \times (101 \times 97.5\% - 98) = 6.619 \text{ (mmol/L)}$$

$$\bar{X} \pm 1.96s = 4.735 \pm 1.96 \times 0.882 = 3.01 \sim 6.46 \text{ (mmol/L)}$$

用百分位数法求得 101 例 30~49 岁健康男子血清总胆固醇值 95% 分布范围 3.095~6.619 (mmol/L), 与正态分布法求得的 95% 分布范围 3.01~6.46 (mmol/L) 基本一致。

(6) 分别考察 $\bar{X} \pm 1s$ 、 $\bar{X} \pm 1.96s$ 、 $\bar{X} \pm 2.58s$ 范围内的实际频数与理论分布是否基本一致 (表 1-3)。

表 1-3 某地 101 例 30~49 岁健康男子血清总胆固醇值
理论分布与实际分布比较

$\bar{X} \pm us$	血清总胆固醇值	实际分布		理论分布
		人数	%	%
$\bar{X} \pm 1s$	3.85~5.62	72	71.29	68.27
$\bar{X} \pm 1.96s$	3.01~6.46	97	96.04	95.00
$\bar{X} \pm 2.58s$	2.46~7.01	100	99.01	99.00

由上表, $\bar{X} \pm 1s$ 范围内, 实际分布与理论分布略有不同, 而 $\bar{X} \pm 1.96s$ 、 $\bar{X} \pm 2.58s$ 范围内, 实际分布与理论分布基本一致。

(7) 现测得一 40 岁男子的血清总胆固醇值为 6.993 (mmol/L), 若按 95% 正常值范围估计, 其血清总胆固醇值是否正常? 估计该地 30~49 岁健康男子中, 还有百分之几的人血清总胆固醇值比他高?

前计算得 95% 正常值范围为 3.01~6.46 (mmol/L), 现测得一 40 岁男子的血清总胆固醇值为 6.993 (mmol/L), 在 95% 范围以外, 故属于异常。

$$u = (X - \mu) / \sigma = (6.993 - 4.735) / 0.882 = 2.56$$

因 $\Phi(2.56) = \Phi(-2.56)$, 查表 1 得 $\Phi(-2.56) = 0.0052$

估计该地 30~49 岁健康男子中约有 0.52% 的人血清总胆固醇值比他高。

1.2 某地卫生防疫站, 对 30 名麻疹易感儿童经气溶胶免疫一个月后, 测得其血凝抑制抗体滴度资料如表 1-4 第(1)(2)栏。

(1) 试计算其平均滴度。

由表 1-4 得, $G = \lg^{-1}(\Sigma f \lg X / \Sigma f) = \lg^{-1}(50.5730/30) = \lg^{-1}1.6858 = 48.5$

该站 50 名麻疹易感儿童经气溶胶免疫一个月后，测得血凝抑制抗体平均滴度为 1:48.50。

表 1-4 平均滴度计算表

抗体滴度	人数 f	滴度倒数 X_1	$\lg X_1$	$f \lg X_1$
(1)	(2)	(3)	(4)	(5) = (2) × (4)
1:8	2	8	0.9031	1.8062
1:16	6	16	1.2041	7.2247
1:32	5	32	1.5051	7.5257
1:64	10	64	1.8062	18.0618
1:128	4	128	2.1072	8.4288
1:256	2	256	2.4082	4.8165
1:512	1	512	2.7093	2.7093
合 计	30			50.5730

(2) 有人发现本例用抗体滴度稀释倍数和直接用滴度（原书误为倒数）算得对数值的标准差相同（表 1-5），为什么？

表 1-5 滴度对数值计算表

抗体滴度 X_2	人数 f	$\lg X_2$	$f \lg X_2$
1:8	2	-0.9031	-1.8062
1:16	6	-1.2041	-7.2247
1:32	5	-1.5051	-7.5257
1:64	10	-1.8062	-18.0618
1:128	4	-2.1072	-8.4288
1:256	2	-2.4082	-4.8165
1:512	1	-2.7093	-2.7093
合 计	30		-50.5730

1) 由表 1-4 中数据计算标准差为： $s_{\lg X_1} = \lg^{-1} 0.4444 = 2.7823$

2) 由表 1-5 中数据计算标准差为: $s_{\lg X_2} = \lg^{-1} (0.4444) = 2.7823$

直接用抗体滴度的对数 $\lg X_2$ 与稀释倍数的对数 $\lg X_1$ 计算标准差是相等的, 因为由上表可见 $\lg X_2 = \lg 1 - \lg X_1 = -\lg X_1$, 而 $\lg X_1$ 与 $-\lg X_1$ 的离散程度是相同的, 所以用抗体滴度稀释倍数和直接用滴度算得对数值的标准差是相同的。

1.3 50 例链球菌咽峡炎患者的潜伏期如表 1-6, 说明用均数、中位数或几何均数, 何者的代表性较好? 并作计算。

表 1-6 50 例链球菌咽峡炎患者的潜伏期的中位数计算表

潜伏期 (小时)	病例数 f	累计频数
12~	1	1
24~	7	8
36~	11	19
48~	11	30
60~	7	
72~	5	
84~	4	
96~	2	
108~120	2	
合 计	50	

本例目测频数分布为偏态分布, 长尾拖向右侧, 故为正偏态, 宜用中位数及几何均数表示其平均水平。

如上表, 经计算中位数、几何均数、算术均数分别为:

$M = 54.55$ (小时), $G = 54.08$ (小时), $\bar{X} = 58.56$ (小时), 显然, 算术均数受长潜伏期的影响使其偏大, 中位数 M 与几何均数 G 接近, 故描述链球菌咽峡炎患者潜伏期的集中趋势指标使用中位数 M 或几何均数 G 均可。

1.4 某市 1974 年为了解该地居民发汞的基础水平, 为汞污

染的环境监测积累资料，调查了留住该市一年以上，无明显肝、肾疾病，无汞作业接触史的居民 238 人，发汞含量如表 1-7：

表 1-7 238 人发汞含量频数计算表

发汞值 ($\mu\text{mol/kg}$)	人数 f	组中值 X	fX	fX^2	累计频数	累计频率
1.5~	20	2.5	50.0	125.00	20	8.40
3.5~	66	4.5	297.0	1336.50	86	36.10
5.5~	60	6.5	390.0	2535.00	146	61.34
7.5~	48	8.5	408.0	3468.00	194	81.50
9.5~	18	10.5	189.0	1984.50	212	89.08
11.5~	16	12.5	200.0	2500.00	228	95.80
13.5~	6	14.5	87.0	1261.50	234	98.32
15.5~	1	16.5	16.5	272.25	235	98.74
17.5~	0	18.5	0.0	0.00	235	98.74
19.5~21.5	3	20.5	61.5	1260.75	238	100.00
合 计	238		1699.0	14743.50		

(1) 说明此频数分布的特征：可见发汞值的频数分布高峰位于第 2 个组段。前 4 个组段的频数占总频数的 81.5%，长尾拖向右侧，呈极度正偏态。

(2) 计算均数 $\bar{X}$ 和中位数 M ，何者较大？为什么？何者用于说明本资料的集中位置较合适？

$$\bar{X} = \Sigma fX / \Sigma f = 1699 / 238 = 7.139 \text{ } (\mu\text{mol/kg})$$

$$M = L + (i / f_{50}) (n \times 50\% - \Sigma f_L)$$

$$= 5.5 + 2/60 (238 \times 50\% - 86) = 6.6 \text{ } (\mu\text{mol/kg})$$

由计算结果得知， $\bar{X} > M$ ，其原因为本例呈正偏态分布，均数计算结果受到少数较大发汞值的影响，使得 $\bar{X}$ 偏向大发汞值一边。本例用中位数描述偏态资料的集中趋势较好，它不受两端较大值和极小值的影响。

(3) 选用何种指标描述其离散程度较好?

选用四分位数间距描述其离散程度较好。

(4) 估计该地居民发汞值的 95% 参考值范围。

本资料应选用单侧 95% 上界值。本例是正偏态分布, 而且样本含量较大, $n = 238$, 可以保证获得一个较为稳定的分布, 故采用百分位数法计算的参考值范围较为合适。

$$\begin{aligned} P_{95} &= L + (i/f_{95}) (n \times 95\% - \Sigma f_L) \\ &= 11.5 + (2/16)(238 \times 95\% - 212) = 13.2625 (\mu\text{mol/kg}) \end{aligned}$$

最佳选择题

1.5 各观察值相加(或减)同一数后 b。

- a. 均数不变, 标准差改变 b. 均数改变, 标准差不变
c. 两者均不变 d. 两者均改变

1.6 用均数与标准差可全面描述 c 资料的特征。

- a. 正偏态分布 b. 负偏态分布
c. 正态分布和近似正态分布 d. 对称分布

1.7 比较身高和体重两组数据变异度大小宜采用 a。

- a. 变异系数(CV) b. 方差(s^2)
c. 极差(R) d. 标准差(s)

1.8 描述一组偏态分布资料的变异度, 以 d 指标较好。

- a. 全距(R) b. 标准差(s)
c. 变异系数(CV) d. 四分位数间距($Q_U \sim Q_L$)

1.9 正态分布曲线下, 横轴上, 从均数 μ 到 $\mu + 1.96$ 倍标准差的面积为 d。

- a. 95% b. 45%
c. 97.5% d. 47.5%

1.10 标准正态分布曲线下中间 90% 的面积所对应的横轴尺度 u 的范围是 a。

- a. -1.645 到 $+1.645$ b. $-\infty$ 到 $+1.645$
c. ∞ 到 $+1.282$ d. -1.282 到 $+1.282$

1.11 若正常人血铅含量近似对数正态分布, 拟用 300 名正

常人血铅值确定 99% 参考值范围,最好采用公式 d 计算。

$$a. \bar{X} \pm 2.58s \quad b. \lg^{-1}(\bar{X} \lg X + 2.58s_{\lg X})$$

$$c. P_{99} = L + (i/f_{99})(300 \times 99/100 - \Sigma f_L)$$

$$d. \lg^{-1}(\bar{X} \lg X + 2.33s_{\lg X})$$

思考题

1.12 某种疫苗通过皮下注射,对 20 名观察对象进行免疫,21 天后观察结果,分别采用 3 种原始记录形式,结果如表 1-8:

表 1-8 用某种疫苗对 20 名对象作皮下注射免疫结果

观察对象	抗体滴度	目测判断 抗体水平	免疫效果 分 类	观察对象	抗体滴度	目测判断 抗体水平	免疫效果 分 类
1	1:40	++	有效	11	1:80	+++	有效
2	1:20	+	无效	12	1:160	+++	有效
3	1:160	+++	有效	13	1:160	+++	有效
4	1:40	++	有效	14	1:80	+++	有效
5	1:320	+++	有效	15	1:40	++	有效
6	1:80	+++	有效	16	1:40	++	有效
7	<1:20	±	无效	17	1:20	+	无效
8	<1:20	±	无效	18	1:80	+++	有效
9	1:40	++	有效	19	1:40	++	有效
10	1:40	++	有效	20	1:160	+++	有效

(1) 表中 3 种记录各属何种类型的变量?

表 1-9 抗体滴度整理表

抗体滴度	人数
<1:20	2
1:20	2
1:40	7
1:80	4
1:160	4
1:320	1
合 计	20

抗体滴度（表 1-9）为数值变量（计量）资料；目测判断抗体水平（表 1-10）为有序分类变量（等级资料）；免疫效果分类（表 1-11）为二项分类变量（计数资料）。

（2）怎样对这些资料进行整理、分析？

<1:20 是两个不确定值，统计描述应采用百分位数法。

表 1-10 目测判断抗体水平整理表

抗体水平	人数
±	2
+	2
++	7
+++	4
++++	5
合 计	20

统计描述采用相对数。

表 1-11 免疫效果分类

效果分类	人数
有效	16
无效	4
合 计	20

统计描述采用相对数。

1.13 数值变量资料频数表的组段是否越细越好？

不是。组段数目一般以 10~15 为宜，组段数目过多，计算较繁；组段数较少，则误差较大，且不易看出资料的分布形态。分组数目要适当，否则会失去分组的意义。

1.14 均数、几何均数和中位数的适用范围有何异同？

相同点：均为表示计量资料集中趋势的指标。