

2009 国家执业医师资格考试用书
遵循新编大纲 贴近实际考试

公卫医师 应试指导 (下)

GONGWEI YISHI
YINGSHIZHIDAO

《国家执业医师资格考试应试指导》专家组 编

- ★ 重视新增内容
- ★ 贴近不同考生
- ★ 精确复习范围
- ★ 提升考试成绩

赠20元
免费学习卡



中国协和医科大学出版社

2009 全国卫生资格考试用书

理论基础知识 临床实践知识

公卫医师 应试指导

(下)

《2009 全国卫生资格考试用书》编委会 编

一、基础知识部分

1. 预防医学基础

2. 流行病学基础

3. 卫生统计学基础



人民卫生出版社 北京

国家执业医师资格考试

(2009 版)

公卫医师应试指导

(下 册)

《国家执业医师资格考试应试指导》专家组 编

中国协和医科大学出版社

目 录

(上 册)

第一部分 基础综合

第一篇 生物化学	(1)	第二单元 血液	(39)
第一单元 蛋白质的结构与功能	(1)	第三单元 血液循环	(44)
第二单元 核酸的结构和功能	(3)	第四单元 呼吸	(52)
第三单元 酶	(5)	第五单元 消化和吸收	(56)
第四单元 糖代谢	(10)	第六单元 能量代谢和体温	(61)
第五单元 生物氧化	(12)	第七单元 尿的生成和排出	(63)
第六单元 脂类代谢	(13)	第八单元 神经系统的功能	(67)
第七单元 氨基酸代谢	(16)	第九单元 内分泌	(74)
第八单元 核苷酸代谢	(19)	第十单元 生殖	(80)
第九单元 遗传信息的传递	(20)	第三篇 医学微生物学	(81)
第十单元 蛋白质生物合成 (翻译)	(23)	第一单元 微生物的基本概念	(81)
第十一单元 基因表达调控	(25)	第二单元 细菌的形态与结构	(82)
第十二单元 信息物质、受体与信号转导	(27)	第三单元 细菌的生理	(85)
第十三单元 重组 DNA 技术	(29)	第四单元 消毒与灭菌	(87)
第十四单元 癌基因与抑癌基因	(30)	第五单元 噬菌体	(90)
第十五单元 血液生化	(30)	第六单元 细菌的遗传与变异	(90)
第十六单元 肝胆生化	(32)	第七单元 细菌的感染与免疫	(92)
第二篇 生理学	(34)	第八单元 细菌感染的检查方法与防治原则	(96)
第一单元 细胞的基本功能	(34)	第九单元 病原性球菌	(98)
		第十单元 肠道杆菌	(102)
		第十一单元 弧菌属	(105)
		第十二单元 厌氧性杆菌	(106)
		第十三单元 棒状杆菌属	(108)

第十四单元 分枝杆菌属	(109)	第八单元 白细胞分化抗原和 黏附分子	(149)
第十五单元 放线菌属和奴卡菌 属	(110)	第九单元 主要组织相容性复合 体及其编码分子	(150)
第十六单元 动物源性细菌	(111)	第十单元 免疫应答	(152)
第十七单元 其他细菌	(113)	第十一单元 黏膜免疫系统	(156)
第十八单元 支原体	(114)	第十二单元 免疫耐受	(157)
第十九单元 立克次体	(114)	第十三单元 抗感染免疫	(158)
第二十单元 衣原体	(115)	第十四单元 超敏反应	(161)
第二十一单元 螺旋体	(115)	第十五单元 自身免疫和自身免 疫性疾病	(164)
第二十二单元 真菌	(116)	第十六单元 免疫缺陷病	(165)
第二十三单元 病毒的基本性 状	(118)	第十七单元 肿瘤免疫	(168)
第二十四单元 病毒的感染与 免疫	(120)	第十八单元 移植免疫	(169)
第二十五单元 病毒感染的检 查方法及防治 原则	(122)	第十九单元 免疫学检测技术	(170)
第二十六单元 呼吸道病毒	(124)	第二十单元 免疫学防治	(172)
第二十七单元 肠道病毒	(127)	第五篇 药理学	(173)
第二十八单元 肝炎病毒	(127)	第一单元 药物效应动力学	(173)
第二十九单元 虫媒病毒	(128)	第二单元 药物代谢动力学	(174)
第三十单元 出血热病毒	(128)	第三单元 胆碱受体激动药	(175)
第三十一单元 疱疹病毒	(129)	第四单元 抗胆碱酯酶药和胆 碱酯酶复活药	(176)
第三十二单元 反转录病毒	(129)	第五单元 M胆碱受体阻断药	(178)
第三十三单元 其他病毒	(131)	第六单元 肾上腺素受体激动 药	(179)
第三十四单元 亚病毒	(132)	第七单元 肾上腺素受体阻断 药	(181)
第四篇 医学免疫学	(133)	第八单元 局部麻醉药	(183)
第一单元 绪论	(133)	第九单元 镇静催眠药	(183)
第二单元 抗原	(134)	第十单元 抗癫痫药和抗惊厥 药	(184)
第三单元 免疫器官	(135)	第十一单元 抗帕金森病药	(186)
第四单元 免疫细胞	(137)	第十二单元 抗精神失常药	(187)
第五单元 免疫球蛋白	(140)	第十三单元 镇痛药	(190)
第六单元 补体系统	(142)		
第七单元 细胞因子	(145)		

第十四单元 解热镇痛抗炎药 (192)	第三十三单元 人工合成抗菌药 (221)
第十五单元 钙拮抗药 (193)	第三十四单元 抗真菌及抗病毒药 (225)
第十六单元 抗心律失常药 (195)	第三十五单元 抗结核病药 (227)
第十七单元 治疗充血性心力衰竭的药物 (197)	第三十六单元 抗疟药 (229)
第十八单元 抗心绞痛药 (199)	第三十七单元 抗恶性肿瘤药物 (231)
第十九单元 抗动脉粥样硬化药 (201)	第六篇 医学心理学 (232)
第二十单元 抗高血压药 (202)	第一单元 绪论 (232)
第二十一单元 利尿药 (204)	第二单元 医学心理学基础 (234)
第二十二单元 作用于血液及造血器官的药物 (207)	第三单元 心理卫生 (239)
第二十三单元 组胺受体阻断药 (209)	第四单元 心身疾病 (240)
第二十四单元 作用于呼吸系统的药物 (210)	第五单元 心理评估 (242)
第二十五单元 作用于消化系统的药物 (212)	第六单元 心理治疗 (243)
第二十六单元 肾上腺糖皮质激素类药物 (212)	第七单元 医患关系 (247)
第二十七单元 甲状腺激素及抗甲状腺药 (214)	第八单元 患者的心理问题 (248)
第二十八单元 胰岛素及口服降血糖药 (215)	第七篇 医学伦理学 (252)
第二十九单元 β 内酰胺类抗生素 (216)	第一单元 绪论 (252)
第三十单元 大环内酯类及林可霉素类抗生素 (218)	第二单元 医学道德的规范体系 (256)
第三十一单元 氨基苷类抗生素 (219)	第三单元 医疗活动中的人际关系道德 (261)
第三十二单元 四环素类及氯霉素 (220)	第四单元 预防医学道德 (264)
	第五单元 临床医学实践道德 (267)
	第六单元 医学科研道德 (272)
	第七单元 医学高科技伦理 (275)
	第八单元 医学道德的修养和评价 (278)
	第八篇 卫生法规 (280)
	第一单元 执业医师法 (280)
	第二单元 母婴保健法 (284)
	第三单元 传染病防治法 (286)

第四单元 病原微生物实验室生物安全管理 条例	(293)	第九单元 职业病防治法	(306)
第五单元 艾滋病防治条例	(295)	第十单元 公共场所卫生管理条例	(310)
第六单元 疫苗流通和预防接种管理条例	(298)	第十一单元 学校卫生工作条例	(311)
第七单元 突发公共卫生事件应急条例	(301)	第十二单元 生活饮用水卫生监督管理办法	(313)
第八单元 食品卫生法	(303)	第十三单元 药品管理法	(315)
		第十四单元 处方管理办法	(318)

第二部分 临床综合

第一篇 症状与体征	(323)	染	(368)
第二篇 呼吸系统疾病	(338)	第三单元 泌尿、男性生殖系统肿瘤	(368)
第一单元 慢性阻塞性肺疾病	(338)	第六篇 女性生殖系统疾病	(370)
第二单元 支气管哮喘	(340)	第一单元 正常妊娠	(370)
第三单元 肺炎	(342)	第二单元 病理妊娠	(372)
第四单元 肺结核	(343)	第三单元 妊娠合并症	(376)
第五单元 肺癌	(346)	第四单元 分娩期并发症	(379)
第三篇 心血管系统疾病	(348)	第五单元 异常产褥	(380)
第一单元 高血压	(348)	第六单元 女性生殖系统炎症	(381)
第二单元 冠状动脉粥样硬化性心脏病	(350)	第七单元 女性生殖器官肿瘤	(382)
第三单元 心脏瓣膜病	(352)	第八单元 生殖内分泌疾病	(385)
第四篇 消化系统疾病	(355)	第七篇 血液系统疾病	(387)
第一单元 食管、胃、十二指肠疾病	(355)	第一单元 贫血	(387)
第二单元 肝脏疾病	(359)	第二单元 白血病	(389)
第三单元 胰腺疾病	(361)	第三单元 输血	(390)
第四单元 肠道疾病	(361)	第八篇 内分泌系统疾病	(392)
第五篇 泌尿、男性生殖系统疾病	(365)	第一单元 甲状腺功能亢进症	(392)
第一单元 肾小球疾病	(365)	第二单元 甲状腺功能减退症	(395)
第二单元 泌尿男性生殖器感		第三单元 单纯性甲状腺肿	(397)
		第四单元 糖尿病	(397)
		第九篇 神经、精神系统疾病	(400)

第一单元 脑血管疾病	(400)	第四单元 伤寒	(437)
第二单元 精神分裂症	(402)	第五单元 细菌性痢疾	(438)
第三单元 心境障碍 (情感性 精神障碍)	(403)	第六单元 霍乱	(438)
第四单元 神经症及癔症	(404)	第七单元 流行性脑脊髓膜炎	(441)
第十篇 运动系统疾病	(407)	第八单元 疟疾	(443)
第十一篇 儿科学疾病	(411)	第九单元 血吸虫病	(445)
第一单元 新生儿与新生儿疾 病	(411)	第十单元 艾滋病	(446)
第二单元 遗传性疾病	(417)	第十三篇 性传播疾病	(448)
第三单元 感染性疾病	(418)	第一单元 淋病	(448)
第四单元 结核病	(422)	第二单元 梅毒	(449)
第五单元 消化系统疾病	(425)	第三单元 生殖道病毒感染	(450)
第六单元 呼吸系统疾病	(427)	第四单元 尖锐湿疣	(450)
第七单元 心血管系统疾病	(430)	第十四篇 其他	(451)
第十二篇 传染病	(432)	第一单元 无菌技术	(451)
第一单元 病毒性肝炎	(432)	第二单元 外科感染	(452)
第二单元 肾综合征出血热	(435)	第三单元 创伤处理原则	(456)
第三单元 流行性乙型脑炎	(436)	第四单元 乳房疾病	(457)
		第五单元 中毒	(460)

第三部分 专业综合

第一篇 流行病学	(473)	断	(524)
第一单元 绪论	(473)	第十一单元 疾病预防策略与 措施	(528)
第二单元 疾病的分布	(474)	第十二单元 传染病流行过程	(531)
第三单元 描述性研究	(481)	第十三单元 传染病暴发调查	(540)
第四单元 队列研究	(486)	第十四单元 艾滋病	(543)
第五单元 病例对照研究	(493)	第十五单元 病毒性肝炎	(548)
第六单元 流行病学实验研究 与随机对照试验	(501)	第十六单元 肺结核	(560)
第七单元 筛检及其评价	(508)	第十七单元 院内感染	(566)
第八单元 系统评价	(511)	第十八单元 常见慢性病	(571)
第九单元 偏倚及其控制	(519)	第十九单元 突发公共卫生事 件流行病学	(580)
第十单元 病因与因果关系推			

(下 册)

第二篇 卫生统计学 (587)	转化 (678)
第一单元 统计学的几个基本概念 (587)	第四单元 影响毒作用的因素 (683)
第二单元 定量资料的统计描述 (587)	第五单元 化学毒物一般毒作用 (688)
第三单元 总体均数的估计和假设检验 (597)	第六单元 化学致癌作用 (699)
第四单元 方差分析 (605)	第七单元 化学毒物致突变作用 (706)
第五单元 分类资料的统计描述 (611)	第八单元 化学毒物生殖和发育毒性作用 (714)
第六单元 率的抽样误差与 Z 检验 (616)	第九单元 管理毒理学 (719)
第七单元 χ^2 检验 (619)	第四篇 环境卫生学 (729)
第八单元 秩和检验 (624)	第一单元 绪论 (729)
第九单元 直线回归与相关 (630)	第二单元 环境与健康的关系 (729)
第十单元 统计表和统计图 (637)	第三单元 大气卫生 (737)
第十一单元 统计设计 (含调查设计和实验设计) (640)	第四单元 水体卫生 (751)
第十二单元 医学常用人口统计指标 (646)	第五单元 饮用水卫生 (757)
第十三单元 寿命表 (648)	第六单元 土壤卫生 (770)
第十四单元 生存分析 (650)	第七单元 住宅和办公场所卫生 (777)
第三篇 卫生毒理学 (666)	第八单元 公共场所卫生 (782)
第一单元 卫生毒理学基本概念 (666)	第九单元 城乡规划卫生 (784)
第二单元 化学毒物的生物转运 (672)	第十单元 环境质量评价 (787)
第三单元 化学毒物的生物	第十一单元 家用化学品卫生 (792)
	第十二单元 环境卫生学基本技能 (797)
	第五篇 劳动卫生与职业病 (799)

第一单元 绪论 (799)	第七篇 妇女保健学 (992)
第二单元 劳动过程的心理 与生理变化 (802)	第一单元 妇女保健概论 (992)
第三单元 人类工效学原理 与应用 (807)	第二单元 女童保健 (995)
第四单元 毒物与职业中毒 (815)	第三单元 青春期保健 (997)
第五单元 粉尘与尘肺 (836)	第四单元 婚前保健 (1001)
第六单元 物理因素对健康 的影响 (847)	第五单元 孕产期保健 (1007)
第七单元 职业性致癌因素 与职业性肿瘤 (860)	第六单元 节育保健 (1019)
第八单元 妇女劳动卫生 (862)	第七单元 更年期保健 (1020)
第九单元 职业性有害因素 的评价与控制 (864)	第八单元 妇科常见病防治 (1024)
第六篇 营养与食品卫生学 (875)	第八篇 儿童保健学 (1034)
第一单元 宏量营养素与能 量 (875)	第一单元 体格生长发育 (1034)
第二单元 矿物质 (885)	第二单元 神经心理发育 (1037)
第三单元 维生素 (892)	第三单元 合理营养 (1038)
第四单元 植物化学物 (904)	第四单元 免疫规划 (1043)
第五单元 各类食品的营养 价值 (909)	第五单元 心理卫生 (1045)
第六单元 特殊人群营养 (920)	第六单元 儿童各年龄期保健 (1046)
第七单元 营养与营养相关 疾病 (936)	第七单元 社区儿童保健 (1048)
第八单元 社区营养 (940)	第八单元 儿童意外伤害及预 防 (1049)
第九单元 食品污染及其预 防 (944)	第九篇 学校/青少年卫生 (1053)
第十单元 各类食品的卫生 及管理 (962)	第一单元 儿童少年生长发育 (1053)
第十一单元 食物中毒及其 预防 (971)	第二单元 影响生长发育的因 素 (1066)
第十二单元 食品卫生监督 管理 (988)	第三单元 生长发育调查和评 价 (1071)
	第四单元 儿童少年心理卫生 (1080)
	第五单元 儿童少年健康监测 与常见病预防 (1090)
	第六单元 体育锻炼与健康 (1105)
	第七单元 教育过程卫生 (1110)
	第八单元 学校建筑和设备卫 生 (1114)
	第九单元 学校健康教育和健 康促进 (1122)

第十单元 学校卫生监督	(1129)	第六单元 健康教育与健康促	
第十篇 社会医学	(1133)	进效果评价	(1195)
第一单元 绪论	(1133)	第七单元 社区健康教育与健	
第二单元 医学模式与健康观	(1134)	康促进	(1198)
第三单元 社会因素与健康	(1137)	第八单元 学校健康教育与健	
第四单元 社会医学研究	(1140)	康促进	(1201)
第五单元 社会卫生状况与社		第九单元 医院健康教育与健	
会卫生策略	(1146)	康促进	(1203)
第六单元 健康危险因素评价	(1148)	第十单元 工作场所健康教育	
第七单元 生命质量评价	(1155)	与健康促进	(1205)
第八单元 社区卫生服务	(1159)	第十一单元 高血压病的健康	
第九单元 社会病防制	(1163)	教育与健康促进	(1208)
第十单元 卫生服务评价	(1174)	第十二单元 成瘾性行为的健	
第十一篇 健康教育与健康促进	(1178)	康教育与健康促	
第一单元 基本概念	(1178)	进	(1210)
第二单元 健康相关行为	(1181)	第十三单元 艾滋病健康教育	
第三单元 健康传播	(1184)	与健康促进	(1212)
第四单元 健康教育与健康促		第十四单元 突发公共卫生事	
进的计划设计	(1188)	件中的健康教育	
第五单元 健康教育与健康促		与健康促进	(1214)
进计划的实施	(1191)		

第二篇 卫生统计学

第一单元 统计学的几个基本概念

一、资料的类型

(一) 定量资料 (quantitative data) 定量资料或数值变量 (numerical variable), 亦称计量资料 (measurement data), 其变量值是定量的, 表现为数值大小, 一般有度量衡单位, 如调查某地某年七岁女童的身体发育状况, 以人为观察单位, 每个人的身高 (cm)、体重 (kg)、血压 (kPa)、坐高指数 (%), 坐高/身高) 等均属定量资料。

(二) 分类资料 (categorical data) 分类资料或分类变量 (categorical variable) 亦称计数资料 (count data)、定性资料 (qualitative data), 其观察值是定性的, 表现为互不相容的类别或属性, 有两种情况:

1. 无序分类 (unordered categories) 包括: ①二项分类: 如观察用某药治疗某病患者的治疗结果, 以每个患者为观察单位, 结果分为治愈与未愈两类。两类间互相对立; ②多项分类: 如观察某人群的血型, 以人为观察单位, 结果分为 A 型、B 型、AB 型与 O 型, 为互不相容的多个类别。

2. 有序分类 (ordinal categories) 各类之间有程度的差别, 给人以“半定量”的概念, 亦称等级资料 (ordinal data)。如测定某人群血清反应, 以人为观察单位, 结果可分 -、 $\pm$ 、+、++ 四级; 又如观察用某药治疗某病患者的治疗结果, 以每个患者为观察单位, 结果分为治愈、显效、好转、无效四级。

二、总体和样本

总体是根据研究目的确定的同质观察单位某种变量值的集合。由总体中随机地抽取部分观察单位, 其变量值组成样本。总体的指标, 用希腊字母表示; 样本的指标用拉丁字母表示。

三、参数和统计量

描述总体特征的指标称为参数 (parameter)。由样本计算得出的指标称为统计量 (statistic)。统计学关心的常常是总体参数的大小, 其依据是统计量的大小及其性质。

四、概率

概率是描述某随机事件发生可能性大小的数值, 常用 P 表示。随机事件的概率在 0 和 1 之间。若某随机事件发生的可能性很大, 概率接近于 1, 若某事件发生可能性小, 则概率接近于 0, 若某随机事件发生的可能性很小, 如其发生的概率 P 小于 0.05 或 0.01, 此时, 称为小概率事件, 可以认为在一次抽样中不会发生。

第二单元 定量资料的统计描述

第一节 定量资料的频数分布

频数分布的类型

频数分布可分为对称和偏态分布两种类型。所谓对称分布是指集中位置在中间, 左右两侧频

数大体对称,如表 2-2-1 所示。所谓偏态分布是指集中位置偏向一侧,频数分布不对称。根据集中位置偏的方向,又可将偏态分布分为正偏态(或左偏态)和负偏态(或右偏态)。表 2-2-2 资料的分布属正偏态分布。

统计分析描述时,针对不同类型的频数分布类资料,采用不同的统计方法。

表 2-2-1 110 名 7 岁男童身高 (cm) 频数分布

身高组段 (1)	频数 (2)	频率 (%) (3)	累计频数 (4)	累计频率 (%) (5)
110 ~	1	0.91	1	0.91
112 ~	3	2.73	4	3.64
114 ~	9	8.18	13	11.82
116 ~	9	8.18	22	20.00
118 ~	15	13.64	37	33.64
120 ~	18	16.36	55	50.00
122 ~	21	19.09	76	69.09
124 ~	14	12.73	90	81.82
126 ~	10	9.09	100	90.91
128 ~	4	3.64	104	94.55
130 ~	3	2.73	107	97.27
132 ~	2	1.82	109	99.09
134 ~ 136	1	0.91	110	100.00
合计	110	100.00	-	-

第二节 集中位置的描述

常用的描述定量资料分布规律的指标有两类,一类是描述数据分布的集中趋势或中心位置的指标——平均数 (average), 另一类是描述数据分布的离散程度或变量变化的变异程度的指标。

广义的平均数包含各种表示数据分布集中趋势的指标。下面介绍在卫生领域最常用的三种平均数指标: 算术均数、几何均数、中位数。

一、算术均数 (arithmetic mean)

算术均数简称均数 (mean)。总体均数 (population mean) 用希腊字母 μ 表示, 样本均数常用 $\bar{X}$ 表示。均数描述一组数据在数量上的平均水平, 其计算方法如下:

1. 直接法 将所有数据 $X_1, \dots, X_n$ 直接相加, 再除以总例数 n , 即

$$\bar{X} = \frac{X_1 + \dots + X_n}{n} = \frac{\sum X}{n} \quad (2.1)$$

式中 $\sum$ 为希腊字母, 读作 sigma, 为求和的符号。

如例 2.1, 某市 1995 年 110 名 7 岁男孩身高均数为

$$\bar{X} = \frac{114.4 + 119.2 + \dots + 132.8}{110} = 121.72 \text{ (cm)}$$

2. 加权法 (weighting method) 当资料中出现相同观察值时, 可将相同观察值的个数 (即

频数 f , 亦称权重) 与该观察值 X 的乘积代替相同观察值的逐个相加。加权法求均数的计算公式为:

$$\bar{X} = \frac{\sum fX}{n} \quad (2.2)$$

对于已绘制成频数表的资料, 可将每组段的组中值 [即 (下限 + 上限) / 2] 代替该组段观察值的实际取值, 用加权法求算术均数。如据表 2-2-1, 用加权法可求某市 1995 年 110 名 7 岁男孩身高均数如下:

$$\bar{X} = \frac{111 \times 1 + 113 \times 3 + 115 \times 9 + \cdots + 135 \times 1}{110} = 121.95 \text{ (cm)}$$

其中 111, 113, 115, ..., 135 是各组段的组中值; 1, 3, 9, ..., 1 是相应各组段的频数。

算术均数 ($\bar{X}$) 适用于频数分布对称的数据。当数据分布对称时, $\bar{X}$ 位于分布的中心, 它是频数分布最集中的位置。但若数据分布不对称, 则 $\bar{X}$ 不位于分布的中心, 也不在分布的集中趋势位置, 它不反映数据的平均水平, 此时可用下面介绍的两种平均数指标。

二、几何均数 (geometric mean)

几何均数适用于原始数据分布不对称, 但经对数转换后对称分布的资料, 其特征是数据之间呈倍数关系或等比数列。医学上血清滴度资料常用几何均数描述其分布的集中趋势。

样本几何均数常用 G 表示, 其计算公式是:

$$G = \sqrt[n]{X_1 X_2 \cdots X_n}$$

利用对数的性质, 上述公式可表达为:

$$G = \lg^{-1} \left(\frac{\sum \lg X}{n} \right) \quad (2.3)$$

可以发现, 几何均数 G 是对数转换后的数据 ($\lg X$) 的算术均数 $\frac{\sum \lg X}{n}$ 的反对数。当资料中出现相同观察值时, 也可用加权法计算几何均数如下:

$$G = \lg^{-1} \left(\frac{\sum f \lg X}{n} \right) \quad (2.4)$$

计算 X 的几何均数时, 可以先求 $1/X$ 的几何均数 G' , 再求 $G = 1/G'$ 。

例 2.2 5 名学龄儿童的麻疹血凝抑制抗体滴度为 1:25, 1:50, 1:50, 1:100, 1:400, 求几何均数。

$$G' = \lg^{-1} \left(\frac{\lg 25 + 2 \lg 50 + \lg 100 + \lg 400}{5} \right) = 75.79 \quad G = 1:75.79$$

三、中位数 (median)

任何分布的定量数据均可用中位数描述其分布的集中位置。样本中位数常用 M 表示。所谓中位数, 是指将一组观察值从小到大排序后居于中间位置的那个数值。对于原始数据, 可用下述公式计算中位数。

$$M = \frac{1}{2} (X_{(\frac{n}{2})} + X_{(\frac{n}{2}+1)}) \quad \text{当 } n \text{ 为偶数时} \quad (2.5)$$

$$M = X_{(\frac{n+1}{2})} \quad \text{当 } n \text{ 为奇数时} \quad (2.6)$$

其中 $X_{(i)}$ 是从小到大排序后居于第 i 位的观察值, 即 $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$ 。

例 2.3 8 名新生儿的身高 (cm) 为 55, 58, 54, 50, 53, 51, 54, 52。求中位数。

将 8 个数排序得 50, 51, 52, 53, 54, 54, 55, 58

本例 $n=8$ $X_{(\frac{n}{2})} = X_{(4)} = 53$ $X_{(\frac{n}{2}+1)} = X_{(5)} = 54$

$$M = \frac{1}{2} (X_{(\frac{n}{2})} + X_{(\frac{n}{2}+1)}) = \frac{1}{2} (X_{(4)} + X_{(5)}) = \frac{1}{2} (53 + 54) = 53.5$$

对于频数表资料求中位数, 可用公式

$$M = L_M + \frac{i_M}{f_M} \left(\frac{n}{2} - \sum f_L \right) \quad (2.7)$$

其中, L_M : 中位数所在组下限; i_M : 中位数所在组的组距; f_M : 中位数所在组的频数; $\sum f_L$: 中位数所在组前一组的累计频数。

例 2.4 某市大气中 SO_2 的日平均浓度 ($\mu g/m^3$) 见表 2-2-2, 求中位数。

表 2-2-2 某市大气中 SO_2 日平均浓度

浓度 ($\mu g/m^3$) (1)	频数 f (2)	累计频数 $\sum f$ (3)	累计频率 % (4)
25 ~	39	39	10.8
50 ~	67	106	29.4
75 ~	64	170	47.1
100 ~	63	233	64.5
125 ~	45	278	77.0
150 ~	30	308	85.3
175 ~	17	325	90.0
200 ~	9	334	92.5
225 ~	7	341	94.5
250 ~	6	347	96.1
275 ~	5	352	97.5
300 ~	3	355	98.3
325 ~ 350	6	361	100.0

由总例数 n 的一半数值可知中位数 M 位于组段 100 ~ 125 之间, 则用公式 (2.7) 计算中位数如下:

$$M = 100 + \frac{125 - 100}{63} \left(\frac{361}{2} - 170 \right) = 104.17 (\mu g/m^3)$$

中位数不仅可用于任何分布的定量资料, 还可用于开口资料 (即无确切最大值或无确切最小值的资料), 且当数据分布对称时, 理论上中位数等于算术均数, 当数据分布对数转换后对称时, 理论上中位数等于几何均数。但由于进一步统计处理的方法较少, 故实际工作中对于能用算术均数或几何均数描述集中趋势的资料还是尽量使用算术均数或几何均数。

第三节 离散程度的描述

集中趋势是数据分布的一个重要特征, 但单有集中趋势指标还不能很好描述数据的分布规律。例如, 设有三组同年龄、同性别儿童体重 (kg) 数据如下:

甲组	26	28	30	32	34
乙组	24	27	30	33	36
丙组	26	29	30	31	34

三组的均数均为 30, 但他们的分布特征不尽相同, 三组中 5 个数据参差不齐程度 (即变异度) 不相同。为了比较全面地描述数据分布的规律, 除了需要有描述集中趋势的指标外, 还需要引入描述数据分布离散程度的指标。

描述离散程度的指标有多种, 这里介绍常用的几种: 极差、方差、标准差、四分位数间距及变异系数。

一、极差和四分位数间距

(一) 极差 (range) 极差亦称为全距, 用字母 R 表示, 描述数据分布的范围。极差大, 说明数据分布较离散。如上述三组数据中: 甲组数据的极差 $R = 34 - 26 = 8$, 乙组数据的极差 $R = 36 - 24 = 12$, 甲组数据分布较乙组集中。

用极差说明数据分布的离散程度, 方法简单明了, 但缺点是: ①除了最大与最小值外, 不能反映组内其他数据的变异; ②样本较大时抽到较大值与较小值的可能性也较大, 因而样本极差也较大, 故样本含量相差较大时, 不宜用极差来比较分布的离散度。

(二) 四分位数间距 (inter-quartile range)

1. 百分位数 (percentile) 百分位数是一种位置指标。样本的第 x 百分位数常记为 P_x , 它是指把数据从小到大排列后处于第 x 百分位置的数值。它将全部数据分成两部分, 在不包括 P_x 的全部数据中有 $x\%$ 的数据小于 P_x , 有 $1 - x\%$ 的数据大于 P_x 。其计算公式:

$$P_x = L_x + \frac{i_x}{f_x} (n \times x\% - \sum f_L) \quad (2.8)$$

式中, L_x : 第 x 百分位数所在组下限; i_x : 第 x 百分位数所在组的组距; f_x : 第 x 百分位数所在组的频数; $\sum f_L$: 第 x 百分位数所在组前一组的累计频数。

例如, 据表 2-2-2 资料求第 25 百分位数 P_{25} 。

由表 2-2-2 第 (4) 列可知 P_{25} 在组段 50 ~ 75 之间, 按公式 (2.8) 计算如下:

$$P_{25} = 50 + \frac{75 - 50}{67} (361 \times 25\% - 39) = 69.1$$

2. 四分位数 下四分位数即第 25 百分位数, 常用 Q_L 表示; 上四分位数即第 75 百分位数, 常用 Q_U 表示。

3. 四分位数间距 四分位数间距是指上、下四分位数的间距, 即 Q_U 与 Q_L 间的差距, 它是从小到大排列后中间一半数据所在的范围。四分位数间距越大, 数据分布的离散程度越大。

例 2.5 据表 2-2-2 资料求某市大气中 SO_2 日平均浓度分布的四分位数间距。

前面已求得第 25 百分位数 $P_{25} = 69.1$, 即下四分位数 $Q_L = 69.1$ 。同样方法可求第 75 百分位数, 由表 2-2-2 第 (4) 列可知 P_{75} 在组段 125 ~ 150 之间, 上四分位数 $Q_U = P_{75} = 125 + [(150 - 125) / 45] \times (361 \times 75\% - 233) = 146.0$ 。

$$\text{四分位数间距} = Q_U - Q_L = 146.0 - 69.1 = 76.9$$

四分位数间距可用于各种分布的资料, 但常用于偏态分布资料。用四分位数间距作为描述数据分布离散程度的指标, 比极差稳定, 但仍未考虑到每个数据的大小。

二、方差与标准差

(一) 方差 (variance) 方差是常用的描述数据分布离散程度指标。样本方差与总体方差

分别用 S^2 与 σ^2 表示。总体方差往往是未知的, 常用样本方差来估计, 样本方差的计算公式为:

$$S^2 = \frac{\sum (X - \bar{X})^2}{n - 1} \quad (2.9)$$

上式中 $\sum (X - \bar{X})^2$ 被称为离均差平方和, 它是每个数据 X 与算术平均数 $\bar{X}$ 之差的平方和, 它描述了每个数据 X 相对于 $\bar{X}$ 分布的集中程度。若数据分布相对于 $\bar{X}$ 很集中, 则 $\sum (X - \bar{X})^2$ 很小; $\sum (X - \bar{X})^2$ 大, 则意味着数据分布分散。样本容量越大, $\sum (X - \bar{X})^2$ 也越大, 故当样本容量不等时, 不能据 $\sum (X - \bar{X})^2$ 的大小来比较数据分布的离散程度。

方差是将离均差平方和除以 $(n - 1)$ 所得, 它相当于平均每个数据的离均差的平方值, 从而克服了离均差平方和受样本含量影响的缺点。故方差可用于不同含量样本数据分布离散程度的比较。方差越大, 数据分布离散程度越大。

方差计算公式中的 $n - 1$ 被称为自由度 (degree of freedom), 它描述了当 $\bar{X}$ 选定时, n 个 X 中能自由变动的 X 的个数。

例如, 甲组 5 名儿童的体重的方差

$$S^2 = \frac{(26 - 30)^2 + (28 - 30)^2 + (30 - 30)^2 + (32 - 30)^2 + (34 - 30)^2}{5 - 1} = 10 \text{ (kg)}^2$$

(二) 标准差 (standard deviation) 方差的单位是原测量单位的平方, 这不利于进一步统计处理, 为此人们常用方差的算术平方根——标准差 (S) 替代方差描述数据分布的离散程度。标准差 S 的计算公式为:

$$S = \sqrt{\frac{\sum (X - \bar{X})^2}{n - 1}} \quad (2.10)$$

标准差的计算

1. 直接法 用代数知识, 可得

$$\sum (X - \bar{X})^2 = \sum X^2 - \frac{(\sum X)^2}{n}$$

故标准差的公式也可写为:

$$S = \sqrt{\frac{\sum X^2 - \frac{(\sum X)^2}{n}}{n - 1}} \quad (2.11)$$

公式中, $\sum X^2$ 是数据的平方和, $(\sum X)^2$ 是数据的和的平方。如上述甲组 5 名儿童体重的离均差平方和为

$$(26^2 + 28^2 + 30^2 + 32^2 + 34^2) - (26 + 28 + 30 + 32 + 34)^2 / 5 = 40$$

$$\text{标准差 } S = \sqrt{\frac{40}{5 - 1}} = 3.16$$

2. 加权法 频数表资料可用加权法计算, 公式如下:

$$S = \sqrt{\frac{\sum fX^2 - \frac{(\sum fX)^2}{n}}{n - 1}} \quad (2.12)$$

其中 X 是各组段的组中值, f 是相应的频数。