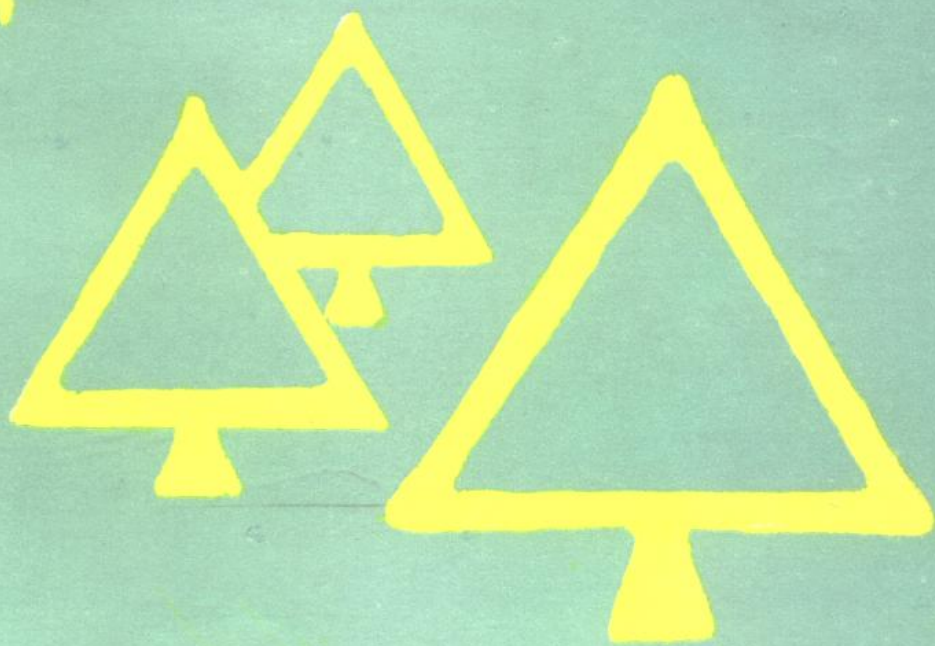


# 环境医学统计学

林琼芳 刘筱娴 梁浩材 余松林 黄渭铭 编著



人民卫生出版社

# 环境医学统计学

林琼芳 刘筱娴 梁浩材 编著  
余松林 黄渭铭

人民卫生出版社

**环境医学统计学**

林琼芳 刘筱娴 梁浩材 编著  
余松林 黄渭铭

人民卫生出版社出版

(北京市崇文区天坛西里10号)

北京市房山区印刷厂印刷

新华书店北京发行所发行

787×1092毫米16开本 22印张 2插页 510千字

1989年7月第1版 1989年7月第1版第1次印刷

印数：00,001—5,000

ISBN 7-117-00328-6/R·329 定价：11.40元

〔科技新书目193—173〕

## 编写说明

环境医学统计是环境医学和环境保护工作者进行环境调查研究、监测、评价和制订卫生标准的重要工具。作者收集了近几年来在教学、科研及现场工作中的实际资料，参考国内外有关文献，应用统计学的基本原理编写此书。编写时，力求理论联系实际，深入浅出，讲清各种统计方法的应用范围和条件，详细列举计算步骤，多举范例，使读者容易理解，学以致用。

本书内容主要包括环境医学调查设计与数据的收集；环境医学数据的分布类型与数学模型的应用；污染水平与人群健康代表值的选择；环境医学研究中质量差异比较的评定方法；多因素的方差分析与回归分析；环境污染物毒性评定的统计分析和环境质量统计评价及控制等方法。可供环境医学工作者、环境保护工作者、卫生防疫人员和医疗保健机构的科研人员参考，大专院校环境医学专业师生也可作参考教材。

本书承周有尚教授审阅，罗五金、严艾荣等同志为本书缮写、制图和校对等工作付出了辛勤劳动，谨此致谢。

作 者

1986年12月

<b>第一章 绪言</b> .....	1
第一节 环境医学与健康统计 .....	1
第二节 几个统计学术语的概念 .....	2
<b>第二章 环境医学调查设计与数据的收集</b> .....	4
第一节 环境医学调查设计的基本程序 .....	4
一、决定目的 .....	5
二、制订计划 .....	6
三、编制调查表 .....	9
四、调查工作的组织 .....	10
第二节 环境污染源及人群健康状况的数据收集 .....	10
一、大气污染源的数据收集 .....	10
二、水体污染的数据收集 .....	12
三、人群健康状况的数据收集 .....	12
四、样本含量的估计 .....	15
第三节 环境污染对健康危害的相关性调查 .....	19
第四节 环境病因的研究方法 .....	21
一、回顾性研究 .....	21
二、前瞻性研究 .....	22
三、历史前瞻性研究 .....	23
四、横向研究 .....	24
第五节 几种常用的随机化分组方法 .....	24
一、完全随机的实验设计 .....	24
二、配对实验设计 .....	25
三、随机区组的实验设计 .....	26
四、拉丁方设计 .....	27
五、复因实验设计 .....	29
六、正交试验设计 .....	29
<b>第三章 环境医学数据分布类型与数学模型的应用</b> .....	31
第一节 测定值的表示方法 .....	31
一、频数分布表的编制 .....	31
二、观察资料的分布形态 .....	32
第二节 连续型概率分布 .....	33
一、正态分布 .....	34
二、对数正态分布 .....	48
三、指数分布 .....	50
第三节 离散型概率分布 .....	54
一、二项分布 .....	54

二、泊松分布	53
三、负二项分布	65
<b>第四章 污染水平的描述方法与人群健康代表值的选择</b>	<b>72</b>
<b>第一节 相对数</b>	<b>72</b>
一、常用相对数的种类	72
二、标准化法	74
三、应用相对数注意事项	78
<b>第二节 平均数</b>	<b>80</b>
一、算术平均数	80
二、几何平均数	83
三、中位数	85
四、移动平均数	87
<b>第三节 变异指标</b>	<b>90</b>
一、全距	90
二、离均差平方和与方差	90
三、标准差	91
四、几何标准差	94
五、百分位数	95
<b>第四节 均数可信区间的估计</b>	<b>97</b>
一、抽样与抽样误差	97
二、总体均数的可信区间估计	98
三、百分位数的可信区间	99
四、率(比)的可信区间	99
<b>第五章 环境医学研究质量差异比较的评定方法</b>	<b>101</b>
<b>第一节 t 检验</b>	<b>101</b>
一、t 值与 t 分布的特点	101
二、t 检验的基本步骤	105
三、样本均数与总体均数差异的显著性检验	106
四、配对资料(或同一批对象实验前后)差异的显著性检验	107
五、两个样本均数差异的显著性检验	109
六、两个样本几何均数差异的显著性检验	111
七、方差不齐时两样本均数差异的显著性检验(t'检验)	113
<b>第二节 F 检验(方差分析一)</b>	<b>115</b>
一、F 检验的基本概念	115
二、F 分布的特点	116
三、单因素多组数据的比较(完全随机设计资料的方差分析)	119
四、双因素多组数据比较(随机区组设计资料的方差分析)	126
五、各组均数间的两两比较	129
<b>第三节 F 检验(方差分析二)</b>	<b>132</b>
一、随机区组设计中缺项的估计	132
二、多因素多组数据比较(拉丁方设计资料的方差分析)	138
三、复因设计资料的方差分析	141

四、正交设计资料的方差分析	144
第四节 $u$ 检验	147
一、 $u$ 检验的意义	147
二、样本率与总体率的比较	147
三、两个样本率的比较	148
四、两个标准化率的比较	149
第五节 $\chi^2$ 检验	152
一、 $\chi^2$ 检验的原理	152
二、 $\chi^2$ 分布和 $\chi^2$ 检验基本步骤	154
三、四格表的 $\chi^2$ 检验	156
四、配对计数资料的 $\chi^2$ 检验	163
五、 $2 \times C$ 表的 $\chi^2$ 检验及分割法	165
六、行 $\times$ 列表的 $\chi^2$ 检验及分割法	168
第六节 非参数统计方法	171
一、符号检验	171
二、秩和检验	175
三、中位数检验法	184
<b>第六章 环境医学数据的相关回归分析</b>	<b>191</b>
第一节 直线相关分析	191
一、相关系数的意义	191
二、相关系数的计算方法	193
三、相关系数的显著性检验	194
四、等级相关分析	197
第二节 直线回归分析	198
一、直线回归方程	198
二、回归方程的显著性检验	201
第三节 直线回归分析的应用	204
一、两个回归方程的比较	204
二、回归模型的确定与应用	210
三、利用回归方程进行预报和控制	212
四、直线回归与调整均值	215
第四节 曲线回归	221
一、指数与对数曲线配合法	221
二、多项式法	226
第五节 多元回归分析	230
一、建立多元回归方程	230
二、多元线性回归方程的效果检验	233
三、复相关系数与偏相关系数	234
第六节 逐步回归分析法	237
一、逐步回归分析法基本思想	237
二、定性资料数量化	238
三、逐步回归的计算方法与步骤	239

四、确定回归方程	246
五、应用回归方程进行预测预报	248
<b>第七章 环境污染物毒性评定的统计分析</b>	251
<b>第一节 急性毒性试验的半数致死量统计分析</b>	251
一、概率单位法	252
二、面积法	259
三、移动平均法	261
四、序贯法	263
五、两个半数致死量差异的显著性检验	264
<b>第二节 急性毒性试验统计的应用及慢性毒性试验的评定</b>	265
一、半数致死量的应用	265
二、毒物蓄积作用	266
三、毒物联合作用	267
<b>第三节 毒性试验设计的要求</b>	268
<b>第四节 繁殖试验及显性致死突变试验的统计指标</b>	268
一、繁殖试验的统计指标	268
二、显性致死突变试验的统计指标	269
<b>第八章 环境质量统计的评价和控制</b>	273
<b>第一节 环境质量指数的统计分析评价</b>	273
一、环境质量单项指数	273
二、综合的环境质量指数	273
<b>第二节 环境质量控制图</b>	274
一、环境质量控制图的作用	274
二、环境质量控制图的原理	275
三、各种控制图的制作方法	275
四、控制图的观察和分析	280
<b>第三节 人体正常值范围的估计</b>	281
一、人体正常值范围资料的收集	281
二、资料的统计处理	282
<b>附录一 统计用表</b>	291
附表1 随机数字表	291
附表2.1 两样本率比较时所需样本含量(双侧)	292
附表2.2 两样本率比较时所需样本含量(单侧)	293
附表3 配对比较(t检验)时所需对子含量	294
附表4 两样本均数比较(t检验)时所需样本含量	295
附表5.1 显著水平与把握度的t值表( $t_{\alpha} + t_{1-\beta}$ )(单侧检验)	296
附表5.2 显著水平与把握度的t值表( $t_{\alpha} + t_{1-\beta}$ )(双侧检验)	296
附表6.1 正态分布表	297
附表6.2 正态分布表(续)	298
附表7 正态性D检验界值表	299
附表8.1 百分率的可信限( $1 \leq n \leq 50$ )	300
附表8.2 百分率的可信限( $1 \leq n \leq 50$ )	301



附表8.3	百分率的可信限 ( $1 \leq n \leq 50$ )	302
附表8.4	百分率的可信限 ( $50 \leq n \leq 100$ )	303
附表8.5	百分率的可信限 ( $100 \leq n \leq 1000$ )	304
附表9	泊松分布的可信限	305
附表10	t 值表	306
附表11.1	F 值表 (双侧检验, 方差齐性检验用)	307
附表11.2	F 值表 (双侧检验, 方差齐性检验用)	308
附表12.1	F 值表 (方差分析用) $P = 0.05$	309
附表12.2	F 值表 (方差分析用) $P = 0.05$ (续)	310
附表13.1	F 值表 (方差分析用) $P = 0.01$	311
附表13.2	F 值表 (方差分析用) $P = 0.01$ (续)	312
附表14	q 值表	313
附表15	$\chi^2$ 值表	314
附表16	符号等级检验表 (Wilcoxon 成对比较用)	315
附表17	等级秩和临界值 (双侧检验)	316
附表18	等级秩和临界值 (单侧检验)	317
附表19.1	H 值与概率对照表	318
附表19.2	H 值与概率对照表 (续)	319
附表20	M 值的界限值 ( $P = 0.05$ )	320
附表21.1	相关系数界值表	321
附表21.2	相关系数界值表 (续)	322
附表22	r 与 z (0 至 3) 之对照表	323
附表23	等级相关系数界值表	324
附表24.1	百分率与概率单位换算表	325
附表24.2	百分率与概率单位换算表	326
附表24.3	百分率与概率单位换算表	327
附表24.4	百分率与概率单位换算表	327
附表25	加权系数	328
附表26.1	用移动平均法计算 $LD_{50}$ 及可信区间	329
附表26.2	用移动平均法计算 $LD_{50}$ 及可信区间	330
附表26.3	用移动平均法计算 $LD_{50}$ 及可信区间	331
附表26.4	用移动平均法计算 $LD_{50}$ 及可信区间	332
附表26.5	用移动平均法计算 $LD_{50}$ 及可信区间	333
附表26.6	用移动平均法计算 $LD_{50}$ 及可信区间	334
附表26.7	用移动平均法计算 $LD_{50}$ 及可信区间	335
附表26.8	用移动平均法计算 $LD_{50}$ 及可信区间	336
附录二	统计符号表	337
附录三	英汉统计学名词对照	340

# 第一章 绪 言

## 第一节 环境医学与健康统计

环境医学是研究环境污染与人群健康关系的科学，它是伴随着我国社会主义现代化建设的进程而发展起来的。保护环境和发展生产力是相辅相成、互相制约、互相促进的。人类在周围环境中生存和发展，环境条件的改变，对机体将产生有利或不利的影晌。例如，环境污染对机体产生不利影晌。环境污染是指由于人类各种活动，向环境中排放各种有害的物质，并且依据这些物质的数量、浓度和污染的持续时间，引起居民身体、精神和财产等方面直接或间接受到恶劣影晌。或在很大区域范围，妨害人类和生物的生产和生活，使公共卫生恶化。环境保护主要是保护大气、水域、土壤（包括森林、草原）三大方面。同时还包括保护矿藏资源、风景区、自然保护区、野生动物资源、城市环境等。因此，环境保护不只限于治理“三废”的狭窄概念。

二十世纪以来，工业生产突飞猛进，社会规模日益扩大，对自然环境影晌愈加深刻，由于环境的污染而显著影晌人民的生命和健康。特别是五十年代以来，严重公害事件接连发生，惊动了工业国家。在被称为当代危机之一的“公害危机”面前，在群众反公害斗争的冲击下，这些国家用高昂的代价解决环境问题。六十年代环境科学诞生，环境医学也是其中重要组成部分。

我国在社会主义现代化过程中，也产生了严重的环境污染问题。据有关部门报告，1979年大气严重污染的城市有22个。国家规定城市每平方公里的降尘量为6~8吨/月，但北京市达39吨，首钢工业区285吨，有些城市高达600~1000吨。全国1979年受严重污染河流230多条，40多个城市地下水受酚、氰、砷等有害物质污染，有的城市全市无净水，群众每年有 $\frac{1}{3}$ 以上时间喝发臭的污水。北京、上海、天津三大直辖市的中心地带噪音强度超过容许强度一倍以上，比纽约、东京、伦敦吵闹得多。据肿瘤学家认为：80%以上的肿瘤可能由环境因素所致。解放三十年来，全国癌症死亡率增长1.46倍，特别是近十多年，增长更快，在死亡病因中由第九位上升到第二位。由于水、粮、菜等被污染，引起畸变和遗传基因退化，畸胎和痴呆逐年增加。至于对工农业生产的损失，更是不胜枚举，对人民健康和生产都带来严重后果。

面对这些环境所引起的问题，环境医学不仅研究上述环境污染对人民健康的损害，也要研究利用和改造环境，化害为利，以保护环境、造福人民、造福子孙后代。在环境医学研究工作中，除了应用物理、化学和生物等有关学科的知识和方法手段外，也广泛应用了统计研究的方法。环境医学统计主要在数量和质量紧密联系的基础上，研究有关环境保护的数据收集、整理和分析，阐明环境污染对居民健康的影响，为制订环境保护规划和计划，加强环境保护系统的管理，评价环境保护工作的质量和效果提供客观依据。

环境医学统计学除了具备一般统计学书籍所述的共同基本内容外，还要结合环境医学的特点。例如：据研究，空气污染物浓度的频率分布通常是高度偏态，它符合对数正

态分布或其它一些不对称分布等。环境质量调查也较复杂，因为污染物浓度受排出时间、地点等人为因素的影响，因而抽取有代表性的样本，必须有详细而周密的设计。

环境医学统计是我们监测环境质量，调查环境污染，评价污染程度，研究对人体健康的影响，预测环境污染动态的有力武器。环境保护人员除了应掌握理化和生物检验知识和技能外，要很好掌握统计学知识和技能，把它作为开展工作的左右手。对于环境保护部门的领导者来说，也应掌握统计学知识，运用统计取得的各种信息，进行科学管理、评价和质量控制，克服盲目性，更好地按科学规律办事。

## 第二节 几个统计学术语的概念

### (一) 总体 (population) 和样本 (sample)

总体指被研究事物的全体，又称母体。样本指从总体中抽取部分个体以供研究，又称子样。例如从河水中采水样作水质检验，从病人中采尿样作铅含量的检验，所取的水样和尿样，称为样本。研究样本为了说明整条河流的污染，研究尿样为了说明病人全部尿液的铅含量。因此，整个河流或病人全部尿液就称为总体。研究样本就是为了说明总体。总体有时是具体的，有时是假想的。有时是有限的，有时是无限的。如上述例子中的尿液就是具体的。但如果发现一种净化河流水质的方法，拟取水样检查使用这种方法前后的差别时，则不论取水样个数多少，都称为样本，这时总体指在其它条件（如温度等）相同时，用此方法处理河流的全部水质情况。这时全部水质是大量的、无法测出的，所以，是无限的总体。

### (二) 参数 (parameter) 和统计数 (statistics)

描述总体特殊性的有关指标（如总体平均数等），称为参数。通过随机化抽样，研究样本特性的有关指标，以估计总体参数，是统计推断的基本任务。反映样本特性的有关指标（如样本均数），称为统计量，亦即样本观测值的函数。统计学习惯用希腊字母代表总体参数，用拉丁字母代表样本统计数。

### (三) 随机化 (randomization)

从总体中抽取样本进行研究时，为了使样本能代表总体，并使其抽样误差大小可用统计学方法来估计，必须按随机化原则抽样，做到总体中每个个体都有同等被抽取的机会。这样，才能避免人为的因素所造成的偏差，以便应用统计方法处理各种数据，因为统计分析和推断方法是根据随机化抽样而建立的。

### (四) 概率 (probability)

概率又称机率，或然率：指用数量表示某一事件发生的可能性。如掷硬币，出现正面或反面的概率均为 $\frac{1}{2}$ ，或0.5，或50%，称为概率，用符号P代表。概率总是在0~1范围。愈接近“1”，发生的可能性愈大；愈接近“0”，发生的可能性愈小。概率的大小是相对的。数理统计在习惯上认为 $P \leq 0.05$ 为小概率。指事物发生的可能性等于或少于0.05，作为事物之间差别有显著性（意义）的界限。

### (五) 误差 (error)

误差是指测定值与真值之差，以及样本指标与总体指标之差。误差可分为系统误差 (systematic error) 和偶然误差 (accidental error)。系统误差是指在收集数据过程中，由于仪器不准，标准试剂未经校正，医生掌握标准偏高或偏低等原因，使观

察结果有倾向性、系统性或周期性的偏大或偏小。系统误差是可以通过实验设计和技术措施来清除或使之减弱。偶然误差是排除系统误差后尚存在的误差，主要有重复误差和抽样误差，重复误差是由于同一受试对象或样品用同一方法重复测定所出现的误差。抽样误差是指样本有关统计数（如样本平均数）与总体参数（如总体平均数）的差异，这是由于总体中各个个体存在着差异，这些个体差异是不可避免，但却是有规律性的。统计学的任务之一是研究误差的统计规律性。如某地在同一个环境污染下，在甲、乙两个学校学生中引起某症状的发生率，甲校为14.3%，乙校为16.5%，表面数值虽不同，但两校均在同一污染条件下，两个率的差别只是数字的表面值，并不是真正的相差，其误差是由抽样引起的。如果根据此表面数值下结论，认为乙校有症状率高于甲校，那就产生了判断错误。

任何客观现象都有偶然性和必然性，随机性和决定性这两重性质。在大量的同类客观现象中，就其个别现象来说，偶然性占重要地位。但从整个客观现象来说，却遵循一定的规律性，这种规律性称为统计规律性。统计学的任务就是透过大量表面的偶然性去研究内部隐藏的规律性，通过随机性去认识决定性，通过偶然性去认识必然性。随机性和决定性，偶然性和必然性，是对立统一的。研究统计学必须掌握统计的思维方法，以辩证唯物思想为指导，正确地收集和分析取得的资料。

（梁浩材 编）

## 第二章 环境医学调查设计与数据的收集

在科学工作走向量化的今天，完整、准确、及时地收集有关科学数据，是研究问题的重要基础。在环境污染的流行病学和实验研究中，数据收集是一项复杂工作。

统计研究工作一般分为三个阶段：收集资料、整理资料和分析资料。收集资料是基础，分析资料是目的，整理资料是环节，三者密切联系，一环紧扣一环。环境医学统计研究也是如此。收集资料是否准确、完整，影响研究、监测、评价工作的全过程。做好收集资料的计划设计，必须具备专业知识和统计知识，因此，这里简单介绍环境医学研究计划的基本专业知识和统计知识。

### 第一节 环境医学调查设计的基本程序

机体、环境和污染物的关系，可见图 2-1。环境污染物要在一定环境条件下，才能作用于机体。机体对污染物作出应激性反应，可使污染物对机体健康不致产生有害的影响。环境污染物未达到一定的污染水平，不会影响机体的健康，污染物如无生物学特异性反应，也不会影响机体健康。

但是，污染物不影响机体健康，并不意味着机体没有污染负荷，因为可能有害的影响仍未显露或尚未明确，见图 2-1。

调查环境污染对机体健康的影响，是环境流行病学研究的任务。测定污染物对机体健康的生物学特异性效应，是环境毒理学研究的任务。调查环境污染物不同水平及其与产生生物效应的关系，是环境病因学研究的任务。环境毒理学和环境病因学着重实验性方面的统计研究，环境流行病学着重现场方面的统计研究，见图 2-2。

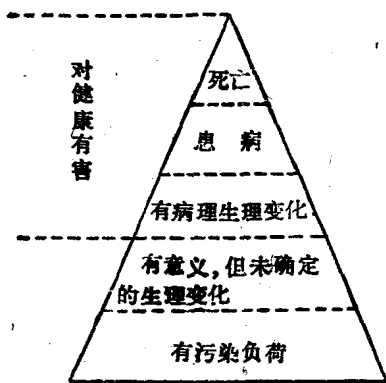


图 2-1 污染对健康影响的级别

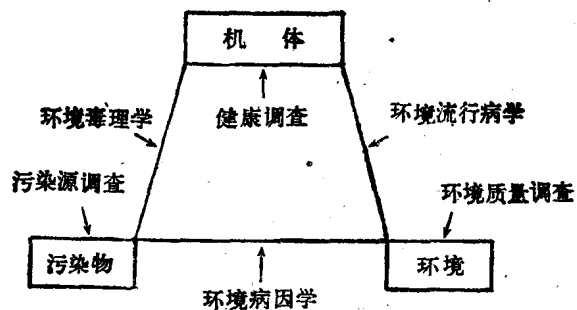


图 2-2 机体、环境和污染物的相互关系

环境医学调查，大致分为对产生污染物的污染源调查、环境质量调查和人群健康状况调查等。但不论那种型式的调查，都应按下图基本程序进行，见图 2-3。

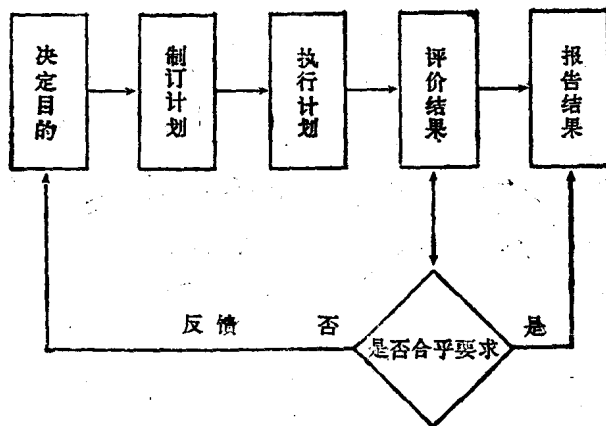


图 2-3 统计工作程序图

### 一、决定目的

决定目的一要有针对性，二要具体化，切勿笼统含糊。例如，水质调查，有时为了摸清水系的基本情况而作现况调查；有时为了掌握污染物的分布和消长而作动态调查；有时为了深入研究某一种物质污染而作专题调查；有时因为发生水质严重污染事故而作临时应急调查等。不能把上述多种调查的目的笼统地归为“了解水质污染情况”一种提法。目的大体分两大类：一类是存在污染问题时进行的基本环境监测；一类是为了特殊需要进行的专门环境监测。如空气质量调查其监测的目的可包括下列具体项目。

(一) 检验污染物的危险区内是否符合空气环境质量标准。

(二) 评定受污染影响的区内何时、何地发生最恶劣的污染物浓度；如果已作出污染的数学模式，则进一步研究的目的（特殊需要）如下。

(三) 验证或修正中范围模式，以便准确地预报未来的背景污染浓度。

(四) 验证或修正小范围模式，以便准确地评价计划中装置或现有装置排放的变化对空气质量的影响。

(注：小范围浓度指近污染源直接提供的污染浓度；中范围浓度指地区范围内污染源造成的局部污染浓度（例如城市）；大范围浓度指污染造成的区域污染浓度）。

由于大范围浓度通常很低，可略去不算，地面观测只考虑中、小范围浓度。

如果研究大气污染对居民生活卫生条件和健康的影响，进行一时性的调查，决定目的后还要确定题目。比如“1980年某市某区大气污染对居民生活卫生条件和健康影响的调查”等。在题目中时间、地点、具体目的要写清楚。因为调查目的决定调查研究的计划设计。如果目的不明确、不具体，收集资料往往多而杂，抓不住要点，耗费人力、物力，浪费时间和资金，调查质量也不好。

制订环境质量调查计划，人们往往想通过一次的努力，达到几个不同目的，这是常犯的毛病，必须防止。如建立某一类型的监测网，用于监测大气污染程度的长期趋势，就不能设想用它去调查某一特殊问题。计划设计随目的不同而异。以空气质量监测为例：

第一类目的包括：

1. 观察长期趋势：为了防止在居住区域发展工业而造成空气质量恶化。为此，用手工操作仪器、24小时间歇采样、在城市选择3~4个测定点即可，但强调这些监测点应长期（至少五年）固定在同一地点。统计分析可计算年平均值，日平均频数分布，工作日和休息日的污染趋势和检验气象因素之间的相互关系。

2. 评价控制方案：监测网应在采用新措施前建立，以便对照。要强调详细说明监测步骤，包括采样器位置，平均采样时间，有无局部污染源和异常影响大气扩散的条件。

第二类目的包括：

1. 控制事故发生：需事先研究出大气污染浓度与气象变量之间关系的数学模式，并在市区范围内得到常规观测核实验证。此时可设计每日每小时采样一次，有提前发出警报的设备。

2. 评价对人民健康的影响：分急性和慢性影响，急性作用可观察健康指标逐日变化，并短期采样测定污染物，如一天中较短时间的平均浓度。慢性作用指污染物的影响多年后才能表现出来。需测定污染物的年平均浓度，考虑到季节性变化和峰值，有时可用24小时的监测数据。

3. 居民申诉的调查：采样与前述不同。要有意识地采取局部污染源的样本，用大量采样设备和连续测定仪器，以便监测瞬时峰值，有时还可设流动采样站。

4. 初步评价的调查：出现新型污染问题的地方，可在一定时间内，用手工采样器以流动采样法进行探索性调查。了解该地使用燃料种类和工业生产种类，作为调查的线索。调查时注意预计采暖时间和特殊气象因素引起的重大变化。此外，为土地功能分类规划而作的专门调查，评价环境污染对植物危害，对建筑物损害的专题调查，以及验证污染扩散模式的调查等，其计划设计也有特点。

## 二、制订计划

### (一) 确定调查对象

调查对象指被调查研究的总体，如某水系。其中组成调查对象的每个被调查成份称为调查单位。环境监测对象包括人和环境两大方面，调查单位可以是一个人，一个家庭，也可以是一个采样点、一个工厂。例如，某厂废气污染对居民生活卫生条件和健康影响的调查，可选择一个污染区和一个对照区，按随机化原则选择常年回家的、近几年没有职业接触被研究毒物的年满18岁以上的一定数量居民，作为调查对象。不在此区，不是常住人口、不足18岁者和有职业接触的人，均为非调查对象。对象的含意应明确，如不吸烟者指从未吸过烟的人，不包括过去吸过烟现在戒烟者，吸烟者则包括现在已戒烟但过去吸过烟者。又如以空气样品为调查对象，则这些样品应按事先设计的地点和时间取得。选地要注意有代表性，避免靠近污染源等。用水系作研究对象，除找到合适的采样地点外，还应按河流距岸边的不同距离、与污水出口的不同距离，以及离水面的不同深度决定采样时间和地点等。

在工业毒理和环境毒理研究中，研究对象常常用动物。如豚鼠、大鼠、小鼠，有时也用猴、狗、猫等动物。动物的选择首先要满足实验要求。其次应当考虑容易取得、经济、易饲养管理等条件。选择动物时还应根据毒物的化学性质，查阅同系动物或类似毒

物的毒理资料，选用最敏感、最合适的实验动物进行实验。

## (二) 规定总体

统计学上，总体指被研究对象的总称。根据目的要求，总体应是性质相同的。例如，上述某地区 18 岁以上成年人性质是相同的，它是指这一地区处于同一空气污染环境中的 18 岁以上的一群人，而不是别区的 18 岁以上的成年人，也不是 18 岁以下的青少年，因为他们回答问题可能有偏差。又如，用同一采样方法、同一时间、同一地点采样，故性质相同。规定总体本来是容易的，但有时考虑不周，不易选好总体。例如，调查化工厂化学物质污染环境对工人健康的影响，则除了调查接触化学物质的全部工人外，不应遗漏过去曾接触过化学物质的退休工人，后者往往因不能坚持工作退休。所以，选择总体一定要有代表性，如一日、一月、一年中不同污染水平都应当采样。选择“志愿者”来研究不能代表全体，因为这些人可能担心自己的健康而当“志愿者”，其身心可能已有不健康因素，使调查产生偏性。同时，统计学上的总体概念，有时是具体的，有时是假想的。例如，改进某种防尘装置效果的研究，其总体是指在同样条件下，用此方法防尘前后环境灰尘量的改变，这时总体环境不能具体确定多大范围，故为假想的总体。

## (三) 确定研究对象的数量

实验结果或疾病的防治效果受许多因素的影响。我们研究的对象是生物，生物本身有个体差异。由于生物的遗传和进化，造成了解剖、生理、生化一系列形态和功能的差别，也产生了毒物在体内变化过程的差异，机体或组织细胞对毒物反应的差异，对处理敏感性就存在着差异。因此，只观察几个动物或几个病例就很难确切地肯定实验结果或防治效果，在安排处理时，应遵循随机化的原则，确定研究对象时必须要有足够的数量。一般地说，在预备试验中个体差异较小时，实验对象数目可以少一些，而个体差异较大时，实验对象要多一些，对于计数资料，研究对象要略多一些，而计量资料，研究对象可相对少一些。此外，实验对象数目还决定于分组的多少与动物的品种。

在同样条件下，研究对象越多，所获得的结论越可靠。但也不是越多越好，对象多，工作量大，增加人力、物力，研究工作不易做细，反而影响质量。故不能盲目追求大样本，而应该用增强实验的敏感性和精密度，以减少个体差异等办法来减少实验单位的数目。

## (四) 设立对照

有比较才有鉴别，设立对照组是为了突出观察研究组所处的特定条件。如为了研究污染区，则同时应研究对照区（清洁区）。这无论在居民调查时，还是抽取样本时均适用。设立对照特别适用于环境流行病学调查研究，无论前瞻性或回顾性调查均需要。严格的对照可用配对方法，它是使观察者与配对者处于基本上相同（齐性）条件下进行比较。配对又可分为个人配对和人群配对两种。

在实验的研究中，对照组除了处理因素（如药物处理外），其它条件（如动物的种属、性别、年龄、窝别、健康状况、饲养情况等），应与实验组齐同，以便突出处理因素的效应，并使对照组与实验组观察例数基本一致。根据实验要求不同，对照的方法有多种形式：如空白对照（不加任何处理因素）、标准对照（实验结果与正常值比较）、阳性对照（用阳性效应的处理因素）、相互对照（实验结果相互比较）及自身对照（实验前后比较）等。



### (五) 选择抽样方式

环境医学统计调查主要用抽样方式。抽样是指在总体中抽取一定数量的调查单位进行研究,用以估计总体的有关情况。如上述调查某区空气污染对居民健康的影响,没有必要全面调查该区每个18岁以上的成人,只需从其中抽样几十分之一,甚至百分之一的人数来研究即可。它的优点是节省人力、物力和时间,由于调查细致,准确性较高。抽样要求有代表性,就应随机化。随机化指在被研究的总体中,每一个体有同等机会被抽取。抽样在空气、水质、土壤采样中很易理解,因“采样”二字是抽样的意思。抽样方式有:

1. 单纯随机抽样:可用抽签或查随机数表方法。
2. 机械抽样:先把调查单位编号,然后从中按一定规则抽取 $\frac{1}{10} \sim \frac{1}{100}$ 等。
3. 整群抽样:如抽一个家庭、一个班组、一个时期(地点)的污染样本,用以说明全体。
4. 分层抽样:把总体先分若干层次,然后根据各层次比例抽样,以防止总体内部构成悬殊不同。
5. 多级抽样:总体中抽取第一批子样,然后在子样中再作第二级抽样,甚至第三级抽样。

例如,上述空气污染对居民健康影响的调查,可用机械抽样或整群抽样。假定抽 $\frac{1}{10}$ 。机械抽样把全区户口本编号,每十户抽一户调查,因空气污染对一户来说影响基本相同,故以户为单位即可。每户必须调查全部18岁以上的人口,不能遗漏,这是整群抽样。较大的河流分为几个断面,污水排出口作一断面,然后若干距离后又一个断面,观察稀释自净状况。交通点空气污染测定也是按与马路的距离分层抽样。

### (六) 规定调查方法

一般分直接观察法,访问调查法和通讯调查法。

直接观察法:如空气、水、土壤的采样,通过检验分析,得出的结果较客观、较准确,但工作量大,耗费多。

访问法:按一定的要求询问被调查者。提问题应客观,切忌向被调查者“启发”答案。例如,提出“您认为工厂排出的空气不好吧?”群众知其来意,往往答“不好,”但未必客观。故访问法要特别注意提问方式。

通讯法:耗费人力、物力较少,但资料往往不完整,“无答复”,回答不准也难以纠正。在条件容许,群众协助时可用之。

不论那种方法,均应注意多用数量(定量)式、选择式答案,少用叙述式、质量(定性)式答案。而且对答案的标准力求使人易理解。如提出:污染是否使室内难以保持清洁?虽然要求答“是”或“否”为选择式,但“清洁”定义不明确,更不用说“保持”多久,各人理解不一,标准不一,失去分析价值。

### (七) 选定观察指标

观察指标分三种:一般标志,如性别、年龄、种族、文化、职业等人的一般特征;时间标志,如出生日期、诊断日期、居住年限、污染影响时间,季节等时间概念;特异标志,如空气污染引起的眼部疾病、呼吸道疾病、或对金属制品的腐蚀性等为调查对象所特有的表现。指标的定义要清楚,如年龄为实足年龄,职业是指调查时点前工作最长