

# 第1篇 医学统计学内容概要

## 1 医学统计学内容分类

### 1.1 引言

人们学习统计学时，常感到统计学像一座迷宫，不知该从哪儿进哪儿出；运用统计学时，深感统计学方法十分繁杂，面对手头的实际资料，不知该选用什么统计分析方法处理最为合适。往往是仅具有十分有限的统计学知识的人，却要处理相当复杂的统计学问题。因而，在公开发表的许多医学期刊和已被批准为科研成果的许多研究报告中，误用和滥用统计学的现象相当普遍，这正是前述两方面情形的真实写照。

解决上述第一个问题的办法是：先对统计学研究的内容有一个较为全面的了解，将其划分成几个具体的模块，并抓住每一个模块中的关键性知识点，做到胸中有数，能够主动地去驾驭知识，而不是被动地去接受知识。

解决上述第二个问题的办法之一是：从问题的原形入手，去学习和运用统计学；对于本专业的研究工作可能会涉及到的各种统计学知识，都要尽可能去学习和掌握，对于难度较大的问题，一时弄不懂，至少也应有所了解。要知道：一个仅会诊治轻微感冒和腹泻的非医务工作者，是不可能成为一个称职的临床大夫的！

解决上述第二个问题的办法之二是：有时，医学研究中可能涉及到的统计学问题相当复杂，从实际工作的科研人员，由于时间和精力的限制，很难成为包括统计学在内的各种专业领域内的专家。因此，他们需要有不耻下问的学习态度和积极寻求合作的明智之举。因为有些统计学问题，对于医学研究工作者来说，可能花费了很多时间还很难解决，而对于统计学工作者来说，可能是轻而易举之事。

### 1.2 医学统计学内容分类

医学统计学的内容可分为以下六类：统计研究设计；统计资料的收集、整理、表达与描述；定量与定性资料的统计分析；相关与回归分析；生存分析；其他多元统计分析。考虑到：有些统计分析方法在具体实施时，离不开复杂的统计计算，为了有效地克服复杂统计计算给学习和运用统计学所带来的困难，还需要学习一种现成的统计分析软件的使用方法及其输出结果的解释。

## 2 统计研究设计要点

### 2.1 引言

进行科学研究探测未知事物、总结过去的工作经验以便指导今后工作等，都有两种具体的办法：其一，盲目开展工作，想到哪干到哪；其二，先制定合理的工作或研究方案，在方案的指导下，有计划、有步骤地开展工作或研究。第一种工作方法盲目性很大，容易丢三落

四、顾此失彼，结果的可靠性较差；第二种工作方法科学性强，考虑问题周密、严谨，不仅可以收到事半功倍的效果，还可以获得令人信服的结果。制定合理的、完善的工作计划或研究方案，就是统计研究设计。

## 2.2 统计研究设计分类及要点

统计研究设计包括：调查设计、实验设计和临床试验设计。

调查研究、调查设计及其要点：对客观存在的事物或现象进行被动观察，包括询问一些有关的情况和测量一些有关指标的数值，以便弄清引起某种结果的原因和已产生的影响或关于未来情况的预测，这就是为了某种目的而进行的调查研究。为了使调查研究卓有成效，需要将各方面的影响因素和各种可能的结果考虑得尽可能全面一些，以使用较少的人力、物力和时间，获得较为全面的、高质量的调查结果，这就需要制定出完善、合理的调查研究设计方案。其中，最重要的问题是要设计出内容全面、又具有可操作性的调查表格。写清调查的时间、地点、范围、对象及数量、指标、指标的测定方法和精度，另外，在调查设计方案中，还应考虑到：参与调查人员的质量和数量及其培训标准、数据的收集方法及其拟采用的统计分析方法等。

实验研究、实验设计及其要点：根据研究目的，通过具体的实验去探测未知事物或现象的本质规律，就属于实验研究。在实验研究中，研究者可以主动地去安排实验，因此，可以对各种重要的非处理因素进行严格控制，使实验因素的实验效应能更充分地显露出来。实验研究的周期是否较短、耗费的人力物力是否较少、考察的影响因素是否较多、结论是否可靠，关键取决于实验设计方案的质量和对其遵照执行的严格程度。实验设计方案的质量高低，主要体现在实验设计四个基本原则。应紧密结合专业和统计学知识，将各种重要的影响因素和观测指标全部包括在实验设计之中。应根据具体实验的特点，选择合适的实验设计类型科学地安排实验因素和区组因素，使适当的对照形式出现在所分的组中，使组间具有高度的可比性和均衡性。应采取随机的方法选取和分配受试对象，并按设计类型估计出恰当的样本含量，使各实验条件下处理的效应能真实地显露出来。

临床试验研究、临床试验设计及其要点：一个新药，经过实验研究（受试对象为动物）发现它具有较好的应用前景，需要在临床医生的大力协助下，以健康志愿者和患者作为受试对象，在人体上进行试验研究，以观察药物的疗效、毒副作用的大小，进一步确定合适的剂型、剂量和给药途径等，这就是临床试验研究。由于临床试验研究的对象是人，既要考虑到伦理道德问题，又要考虑到受试对象的依从性问题。需要控制的影响因素比动物实验要复杂得多，因此，需要制定出科学、严谨的临床试验设计方案。其中，最关键的问题是除了严格按实验设计中四个基本原则操作外，还应采用盲法，提高受试对象的依从性，尽可能减少人为因素的干扰和影响。

# 3 各种常见设计类型的特点和适用场合

## 3.1 调查研究设计

### (1) 横断面研究设计

就是通过对事物或现象现状的调查，反映出此时此刻环境中某些物质的状况或所处的水平、或受试对象的现状或体内某些指标所处的水平，从而达到了解事情的真相，找到解决问题的对策。这种研究，往往不能准确判定某些事物发生的时间先后关系。其研究结果主要用于为进一步研究提供一些重要的线索。

### (2) 队列研究设计

也称为前瞻性研究设计或由因索果的研究设计。事先并不能准确预言某种原因会引起某种结果，只是根据现有的专业知识，提出一些设想或假定，并按假设的可能危险因素的水平，将受试对象分成几个组，然后，研究者对他们进行追踪观察一段较长的时间，并记录处在不同组中的受试对象所发生的各种结果。这样获得的资料，可以清楚地反映出影响因素与结果出现的时间先后顺序。为研究危险因素与疾病之间的因果关系，提供了有力的证据。其缺点是需要的样本含量较大，追踪观察的时间较长。

### (3) 病例-对照研究设计

也称为回顾性研究设计或由果索因的研究设计。面对一组患有某病的患者，寻找另一组情况与患者接近的非该病患者或健康人为对照组，对这两组人进行回顾性调查，了解他们分别接触过哪些可能的危险因素，并通过假设检验等方法，来说明疾病与危险因素之间的联系。这种设计的优点是需要的样本含量不大，花费的时间较少；缺点是不能十分肯定所发现的可疑危险因素是否一定是该病的危险因素、危险因素是否一定先于疾病的出现。需要借助其他方法或试验予以证明。

## 3.2 实验研究设计

### (1) 单组设计

所考察的试验因素只取一个水平，通常是在一个特定的条件下，只观测一组受试对象。如果只有一个定量的观测指标，就只有一组数据；如果有  $K$  个定量的观测指标，就有  $K$  组数据。其要点是：对这种资料进行统计分析时，必须给出标准值或理论值或总体均数。

### (2) 配对设计

全部实验数据以成对的方式出现，同一对中的两个数据，要么来自同一个个体（称为自身配对设计）要么来自相同的母体（称为同源配对设计）要么来自条件相近的两个个体（称为条件相近者配对设计）。如果一个实验因素的两个水平中，有一个是空白对照，另一个是真实的处理，则可采用配对条件最严的自身配对设计；如果一个实验因素的两个水平都是真实的处理，当易于获得多对同源的受试对象时，应首选同源配对设计，反之，可考虑选用配对条件最差的条件相近者配对设计。但能否采用最后一种配对设计，要根据具体情况而定，千万不要生搬硬套。若很难找到合适的配对条件（指各种重要的非处理因素），就应改用下面的成组设计。由此可见，配对设计的要点是：所选择的配对条件确实是最重要的非处理因素，而且，受试对象确实可以严格按此配对条件进行配对。

### (3) 成组设计

如果一个实验因素的两个水平都是真实的处理，又很难找到合适的配对条件，将全部受试对象进行配对时，应按如下的方法进行：当实验因素的水平与受试对象的分组无关时，就应将全部受试对象完全随机地分成两组；当实验因素的水平与受试对象的分组有关（如考察的实验因素是性别、血型）时，就应在定性因素的各水平组内进行完全随机化选取，从而形

成两个组。上述两种分组方式所产生的结果，都称为成组设计。成组设计的要点：不应将符合配对设计的问题简化为成组设计；应根据实验因素与受试对象的关系选择合适的分配受试对象的方式；两组样本含量应力求相等，可使标准误差小。

#### (4) 单因素 $K$ 水平设计 ( $K \geq 3$ )

这个设计与成组设计完全相同，只是所考察的这个实验因素的水平数大于 2 而已。

#### (5) 配伍组设计

有时，实验中只考察一个具有  $K$  水平的实验因素，但根据专业知识得知：存在某个重要的非处理因素，若不考虑它，对全部受试对象进行完全随机化分组，有可能出现这样的现象：即各组之间，在这个重要的非处理因素上均衡性很差。为了有效地消除或减弱它的影响，将全部受试对象按这个重要的非处理因素的水平取值由小到大排队，依次将相邻的每  $K$  个受试对象分成一组。然后，将每组中的  $K$  个受试对象随机地分入  $K$  个实验组中去。这样的安排受试对象的方法，称为配伍组设计。配伍组设计的要点：与配对设计的要点相似，即所选取的区组因素的确是重要的非处理因素，而且，受试对象确实可以严格按此区组因素进行分组。

#### (6) 拉丁方设计

当实验中只考察一个  $K$  水平的实验因素，但同时又涉及两个都具有  $K$  水平的区组因素，且它们之间不存在显著性的交互作用时，应选用拉丁方设计。此设计的要点是：实验因素的水平应采用随机化法，排成一个拉丁方阵；两个区组因素的选取和实施，都有与配伍组设计中的区组因素一样的要求。

#### (7) 交叉设计

当实验中涉及到一个具有 2 水平的实验因素（一般不含空白对照组），根据专业知识的需要，希望该实验因素的两个水平要先后作用于同一个受试对象，此时，就需要采用交叉设计。此设计可考察一个具有 2 水平的实验因素，两个区组因素（通常为受试对象和实验顺序）。此设计的要点是：实验因素的某个水平作用于受试对象后，其实验效应将在不长的一段时间内消失，观测指标的取值会很快恢复到原先的水平，然后，才能施加第二种处理；对全部受试对象来说，前后两次接受处理的时间间隔相等，实验因素的两个水平在条件相近的同一对受试对象之间交叉出现。

#### (8) 析因设计

析因设计可用于涉及多个实验因素的实验场合，尤其适用于因素之间有复杂交互作用的实验研究问题之中。其实验条件数等于全部因素各水平数的乘积，在每种实验条件下，至少要做 2 次独立重复实验。各因素在专业上的地位平等，做实验时同时施加。此设计的要点是：因素水平之间全面组合；各种组合下要做重复实验；因素地位平等、同时施加。

#### (9) 正交设计

当实验中拟考察的因素较多，因素的水平不太多（一般在 2~4 之间），但因素之间的某些一级交互作用必须考虑。此时，若选用析因设计，需要的实验次数太多，研究者难以承受，选用正交设计是比较合适的。其要点是：选用合适的正交表，妥善安排交互作用列，注意：主效应不应与可能有显著性意义的交互作用列重叠；正交表中剩余的空列不应太少，否则，会因误差项的自由度过小，而使结果不可靠。

#### (10) 均匀设计

当实验中拟考察的因素很多，因素的水平也较多时，即使用正交设计，仍感到实验次数太多，此时，可考虑选用均匀设计。它以牺牲因素水平组合之间的正交性，仅强调实验点在实验空间中的均匀性为代价，来达到大幅度减少实验次数的目的。此设计最适合用于对诸因素及其交互作用的重要性一概不知的大规模（或每做一次实验，费用十分昂贵的）实验研究之中，通过此设计进行因素筛选。当因素和水平的数目缩小后，再改用正交设计或析因设计，作详细研究。其要点是：选择合适的均匀设计表以及与之相匹配的使用表，并运用多元回归分析方法处理实验结果。

#### (11) 具有重复测量的设计

最常用的重复测量设计是：考察一个具有  $m$  水平的实验因素，将受试对象随机地均分成  $m$  个组，在  $k$  个不同时间点上，对每组中的  $n$  个受试对象进行重复测量。以了解接受不同处理后的受试对象体内某些定量指标随时间推移的动态变化趋势。此设计常被医学研究工作者不自觉地运用，因为在许多具体的实验研究中，专业上要求研究者应当对特定条件下的受试对象在不同时间点上进行重复观测。其要点是：在  $K$  个不同时间点上从同一个受试对象身上重复观测到的  $K$  个数据必须写在同一行上，以便选用与该设计对应的统计分析方法进行统计处理。

### 3.3 临床试验设计

从设计类型上来看，临床试验设计并没有一套特殊的设计类型，基本上与实验研究设计相同。所不同的是：在试验分组和处理因素的施加上，要考虑到伦理道德问题和提高受试对象的依从性问题。例如：若研究的是某种非慢性病，就不应设置空白对照组或仅用安慰剂的对照组；又例如：为了避免患者怀疑自己服的药不如另一组患者服的药好，而中途改用他药，破坏了整体设计方案，试验组与对照组所用的药物除了在外观上完全相同外，还必须采取单盲法（即患者不知道自己是属于试验组还是对照组的受试对象）或双盲法（患者与医护人员都不知道试验分组的详细情况），必要时，还应采用三盲法（即患者、医护人员、统计分析人员三方面的人都不知道试验分组的详细情况，只有临床试验研究设计者知道）。在临床试验设计中，是否采用盲法试验，往往是试验成败的关键所在。

## 4 统计资料收集与整理要点

如果是小规模实验研究，每次有一个特定的设计类型，就必然有与此设计类型相对应的收集资料的表格，如拉丁方设计、析因设计、正交设计、具有重复测量的设计。按各种不同设计类型的表格要求去记录实验数据，既简便清楚，又便于进行统计整理和分析；如果实验研究的规模很大，当它可以拆成一系列小规模实验时，仍采用前述的方法记录和整理数据；反之，需要按受试对象或观测单位进行记录，其内容是与研究问题可能有关的全部影响因素和观测指标的具体取值。假定一共考察或观测了  $n$  个受试对象或观测单位，一共有  $m$  个影响因素和  $k$  个观测指标，那么，其收集资料的表格就相当于是一个  $n$  行、 $m+k$  列的数据矩阵。最好在实验研究过程中，就建立数据库，有了一些实验数据后，就及时输入

计算机，等实验结束时，数据库也建立完毕了。

## 5 统计资料表达与描述要点

表达统计资料最常用的工具是统计表和统计图。统计表的标题要简单明了，重点突出，表中的纵横栏目要安排得妥当，尤其是表中的数据代表的含义要表达清楚；统计图表达资料形象生动，所选择的统计图类型要与资料的类型相匹配，尤其是坐标轴上的刻度标得要符合数学原则，当用普通算术尺度表示时，等长的间隔，应当代表相等的数量。要根据资料的分布特点和具体情况，选用合适的平均指标表达资料的平均水平，选用合适的变异指标表达资料的离散程度。千万不要用表达正态分布资料的方法表达明显呈偏态分布的资料，估计定量资料正常值范围和估计总体均数置信区间的公式很相似，不应误用。

## 6 定量资料假设检验要点

所谓定量资料，这里指观测指标是定量的，但影响因素往往是定性的、定量的和由定量变量转化而成的定性变量。当定量的观测数据满足正态性和方差齐性这两个前提条件时，应选用参数检验法。反之，需对数据进行变量变换，使之满足参数检验的前提条件；或直接采用相应的非参数检验方法。在使用参数检验法中，人们常误用  $t$  检验分析一切定量资料，这是很不妥当的！具体应选用 3 种  $t$  检验中的哪一种、10 多种方差分析中的哪一种，应根据实验设计类型和资料的实际情况来确定，千万不可盲目套用。

## 7 定性资料假设检验要点

所谓定性资料，通常指影响因素和观测指标都是定性变量，有时，即使它们中的某些是定量变量，根据专业知识的需要，将其转换成定性的变量。定性变量的具体取值是一些名称（称其为名义变量）或等级（称其为有序变量），故需按各定性变量的分组标志进行排列组合，形成二维、三维或高维的列联表，表中的正整数是数出来的，而不是对每一个受试对象直接观测到的，称其为频数。对列联表资料，人们常不管定性变量的性质和具体要求，盲目套用  $\chi^2$  检验。事实上，对常用的二维列联表来说，应该考察它是属于双向无序、单向有序、双向有序属性不同、还是双向有序属性相同的列联表，并根据资料所满足的前提条件和分析目的等，选用相应的统计分析方法。

## 8 直线相关与回归分析要点

研究两个定量变量之间的相互关系或依存关系时，需要运用直线相关与回归分析方法。其要点是：首先，应有专业知识为依据，即在专业上研究它们之间的关系是有意义的。因为全靠统计方法来证明它们之间是否有关，那是很危险的做法，可能会得出荒唐可笑的结论来！其次，要绘制反映两变量变化趋势的散布图，以便根据图形中反映出来的变化趋势，选择合适的回归模型，的确是直线趋势时，才可选用直线相关与回归分析；若呈某种明显的曲线变化趋势，就应设法拟合相应的曲线方程。尤其是当散点呈 S 形或倒 S 形变化趋势时，盲目拟合直线回归方程，假设检验的结果认为效果是满意的，其实，用直线回归方程去预测，可能会出现脱离实际的结果。第三步，才谈得上进行具体的统计计算和假设检验。

## 9 曲线回归分析要点

进行曲线回归分析时，其要点是选择合适的曲线类型。事实上，根据散点图中各散点的分布趋势准确地判断曲线类型是很困难的，往往需要采用几种最可能的曲线类型，分别拟合同一个资料，先对每种拟合结果进行拟合优度检验，然后对它们进行拟合优度比较。从而挑选出拟合得最好的曲线模型。

## 10 多元回归分析要点

多元回归分析可分为多元线性回归分析（含多项式回归分析）、多元非线性回归分析（含多元 Logistic 回归分析）。其要点是考察因变量的分布类型。当因变量是服从正态分布的定量变量时，常可选用多元线性回归分析；当因变量是二值变量或多值的有序变量时，应选用多元 Logistic 回归分析。

## 11 生存分析要点

生存资料有两个明显的特点：其一，资料一般不服从正态分布；其二，资料中常含有信息不全的数据。由于这两个特点，所以，生存资料不适合用一般的统计分析方法进行处理，而需要运用专门的统计分析方法——生存分析。生存分析方法中有三类方法：即非参数法、半参数法和参数法。非参数法（有乘积极限法和寿命表法等多种方法）只适合处理单因素的问题；半参数法（也称为 COX 模型）可处理多因素的问题，适用面很宽；参数法（也称为参数模型回归分析法）精确度高，但需要知道生存时间数据的准确分布，适用面较窄。

## 12 其他多元统计分析要点

如果资料中只有多个定量的观测指标，它们之间没有自变量和因变量之分，欲研究变量之间的亲疏关系时，可选用变量聚类分析；欲根据变量之间的关系，将受试对象进行分类时，可选用样品聚类分析；欲降低变量的维数，用少数几个综合变量表达众多原变量所反映的绝大部分信息时，需选用主成分分析或因子分析；欲将变量和样品的信息同时反映在同一个直角坐标系内时，应选用对应分析。

当资料中有定性变量，还有一系列定量的观测指标，若只想比较定性变量取不同水平，多个在专业上有一定联系的定量指标的平均数之间的差别是否有显著性意义时，可以选用多元方差分析；若定性变量代表的是几个明确分类的总体，希望建立一种方法，对未知的个体进行归属的判断时，就应选用判别分析；若在作多元方差分析的时候，发现还有一个或多个定量的影响因素，希望将其影响消除后，再作多元方差分析，此时，就应选用多元协方差分析。

当资料中有两组在专业上有一定联系的定量变量，希望研究这两组定量变量之间的相关关系时，就可选用典型相关分析。

## 13 统计资料综合分析要点

事实上，人们常常面对的是多因素、多指标的复杂统计资料，不存在某一种统计方法能利用全部的数据，并回答专业上期待解决的全部问题。这就需要结合专业和统计学知识，选择其中不同的变量子集，进行各种相应的统计分析，以便回答各种不同的专业问题，这就是所谓的统计资料综合分析。其要点是：对资料中各变量（包括影响因素和观测指标）的性质和专业背景非常了解，对各种统计分析方法非常熟悉，从而根据专业上提出来的分析要求，正确选用各种相应的统计分析方法进行处理，并紧密联系专业知识和统计学知识，对计算结果作出合理的分析与解释，以便指导今后的类似研究工作。

## 第 2 篇 医学统计学考题精选

### 试卷 1

(说明：开卷考试，时间 2 小时)

1. 单选题 共 20 分，每小题 2 分)：

1.1 设某试验因素 A 有  $K(K \geq 3)$  个水平，观测数据是连续性资料，且满足各种参数检验的前提条件。用多次  $t$  检验取代方差分析和  $Q$  检验。将会( )。

- A. 明显增大犯第 II 类错误的概率      B. 使结论更加具体  
C. 明显增大犯第 I 类错误的概率      D. 使计算更加简便

1.2 反映一组性质相同的观测值的平均水平，常选用算术平均数和几何平均数。适合用几何平均数计算的资料应该是服从( )的。

- A. 正态分布      B. 负偏态分布      C. 对数正态分布      D.  $t$  分布

1.3 设有  $X$ 、 $Y$  两组数据，求得  $\hat{y} = a + bx$ ，经过统计学检验，在  $\alpha = 0.05$  水准上拒绝  $H_0: \beta = 0$ ，则至少有 95% 的把握断言  $y$  与  $x$  之间在专业上有直线关系。这一结论( )。

- A. 根据充分      B. 很有科学性      C. 脱离实际      D. 略有问题

1.4 某疗养院测得 1096 名飞行人员红细胞数 (万/mm<sup>3</sup>)，经检验该资料服从正态分布，其均值为 414.1 标准差为 42.8 求得的区间  $(414.1 - 1.96 \times 42.8, 414.1 + 1.96 \times 42.8)$ ，称为红细胞数的( )。

- A. 99% 正常值范围      B. 95% 正常值范围      C. 99% 置信区间      D. 95% 置信区间

1.5 为了直观地比较化疗后相同时间点上一组乳癌患者血清肌酐和血液尿素氮两项指标观测值的变异程度的大小，可选用的变异指标是( )。

- A. 标准差      B. 标准误      C. 相关系数      D. 变异系数

1.6 试验设计中强调必须遵守“随机、对照、重复、均衡”四大原则，其根本目的就是为( )。

- A. 减少过失误差、降低随机误差、消除系统误差；  
B. 便于收集资料、便于统计处理、便于撰写论文；  
C. 纯化“信号”、降低“噪音”、多快好省；  
D. 仅用一、两次，最多十几次试验，就可得到可靠结果。

1.7 已知 A、B、C 都是三水平因素，且根据预试验得知： $A \times B$ 、 $B \times C$  不可忽视。若希望试验次数尽可能少一些，试验设计时最好选择( )。

- A. 拉丁方设计      B. 正交设计      C. 析因设计      D. 交叉设计

1.8 在某项试验研究中，考察的指标有一百多项。它们彼此之间并无因变量和自变量之分，但从专业上看，可把这些变量分成几个部分，每一部分可用综合指标来表示，这样便于抓住问题的本质。为此，可选用( )。

- A. 聚类分析    B. 多元回归分析    C. 典型相关分析    D. 因子分析

1.9 为了研究一组因素  $X_1, X_2, \dots, X_r$  对一组肺癌患者生存时间长短的影响,对病人进行追踪随访观察,观察结果分为两档,即病人生存时间  $\leq 2$  年和  $> 2$  年,考察的危险因素有治疗方法(4种)、疗前病人的状态(6种)、病人的年龄、癌细胞的类型(3种)。问这种资料最适合选用什么统计方法分析?

- A. 方差分析    B. 多元线性回归分析    C. 聚类分析    D. Logistic 回归分析

1.10 用两种不同成份的培养基(分别为 701 批与 702 批)分别培养鼠疫杆菌,重复试验单元数均为 5 个,将 48 小时内各试验单元上生长的活菌数记录如下,请识别该资料的类型。

701 批 :各有 48, 84, 90, 123, 171 个活菌数 ;

702 批 :各有 90, 116, 124, 225, 84 个活菌数 ;

- A. 计数资料    B. 计量资料    C. 等级资料    D. 品质资料

2. 试验设计及设计类型的识别(共 20 分,每小题 5 分)

2.1 资料见表 1,每一行中的三个数据测自同一只大鼠,问此资料来自什么试验设计?

表 1 大鼠分别接受某药三种不同剂量后在三个不同时间点上心率的测定结果

药物 剂量	鼠 号	心率(次/分)			
		给药后时间(min):	0	2	4
低剂量组	1		205	246	215
	2		313	338	340
	...		...	...	...
中剂量组	7		277	268	275
	8		339	296	284
	...		...	...	...
高剂量组	13		221	193	205
	14		278	245	253
	...		...	...	...

此设计称为(单选题):

- A.  $3 \times 3$  析因设计    B. 含区组因素的析因设计  
C. 组内分组设计    D. 具有一个重复测量的双因素设计

2.2 某试验需在 5% 和 15% 两种氧浓度下进行,每种氧浓度下分别使用甲、乙、丙、丁四种类似的药物对  $K_{562}$  细胞进行抑制,每个试验条件下均重复 4 次试验(独立的 4 个样品,所含细胞的质和量都相同)。观测的指标为计量资料。此设计称为(单选题):

- A. 析因设计    B. 配对设计    C. 拉丁方设计    D. 交叉设计

2.3 请用表 2 中简化的随机排列表第 15 行随机排列数字将 14 只家兔均分到对照组和试验组中去。

表 2 随机排列表 (n=20) 的简化结果

No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	$r_k$	
	15	17	1	0	16	9	12	2	4	5	18	14	15	7	19	6	8	11	3	10	13	0.1053

2.4 在某项试验研究工作中, 要研究 A、B、C、D、E、F、G 七个试验因素对观测指标 Y(计量的)的影响是否有显著性意义, 每个因素只考虑 3 个水平, 根据专业知识得知, 还需考虑交互作用 D×E 和 D×F。请选择合适的正交表, 并进行表头设计(表体内因素的水平代码全部省略)

3. 统计计算和显著性检验 共 60分, 每小题 15分)

3.1 用某种饲料养猪, 其背膘 y(mm)与屠宰重量 x(lb)的数据如下(设这两项指标在专业上有一定的联系, 且散布图中各散点呈直线趋势):

x: 206 261 279 221 216 198 277 250

y: 42 38 53 34 35 31 45 43

试求 y 倚 x 变化的直线回归方程并检验其显著性。 ( $l_{xx}=7270, l_{xy}=1225.5, l_{yy}=352.875$ )

3.2 维吾尔族与回族居民血型构成比资料见表 3, 问两民族血型分布是否相同?

表 3 维吾尔族与回族居民血型构成比调查结果

民 族	构成比(%)				调查例数
	血型: A	B	O	AB	
维吾尔族	29.21	31.92	27.50	11.37	1513
回 族	27.23	28.34	35.94	8.49	1355

3.3 为试验“736”对肉瘤的抑制作用, 将 16 只长出肉瘤的小鼠随机分为两组, 实验组注射“736”, 对照组注射同量生理盐水, 10 天后解剖称瘤重, 资料见表 4, 假定资料满足参数检验所需的各种前提条件, 问“736”是否对肉瘤有抑制作用?

表 4 对照组与实验组小鼠瘤重的测定结果

组 别	瘤重(克)							
给药组	1.6	2.2	2.0	2.0	2.5	1.0	3.7	1.5
对照组	2.1	4.9	2.7	4.3	2.5	1.7	4.5	3.4

3.4 随机抽样对 12 个人的某项生理功能在服药前后给以打分, 所得结果见表 5, 假定资料满足参数检验所需的各种前提条件, 问服药前后之间的差别是否有统计学意义?

表 5 某项生理指标在服药前后的打分结果

时 间	评分值												
	人员编号:	1	2	3	4	5	6	7	8	9	10	11	12
服药前		9	7	6	6	5	7	6	7	5	4	8	6
服药后		9	8	8	5	8	5	9	9	7	8	9	8

试卷 2

(说明：开卷考试，时间 2 小时)

1. 填空题(12 分，每空 1 分)

- 1.1 根据研究目的确定的 ( ) 研究对象的全体称为总体。
- 1.2 概率与频率都是描述某事件发生的可能性大小的一个数量； ( ) 是对样本而言， ( ) 则是总体的属性。
- 1.3 由于个体差异的存在，经过抽样研究后，发现样本指标与总体指标之间总存在偏差，此偏差称之为 ( )。
- 1.4 Poisson 分布可看成是 ( ) 分布的特例，因为当该分布的总体均值  $\lambda \geq$  ( ) 时，可以认为 Poisson 分布近似服从 ( ) 分布。
- 1.5 处理因素、受试对象和实验效应被称为医学实验设计的 ( )。
- 1.6 对照、重复、( ) 和 ( ) 是实验设计中的四个基本原则。
- 1.7 在实验设计中严格遵循四个基本原则，其共同目的是为了尽量减少和抵消 ( ) 的干扰和影响。
- 1.8 对于四格表资料，当  $n < 40$  或理论值  $T < 1$  时，应该选用 ( ) 方法分析。

2. 单选题(28 分，每题 2 分)

2.1 某偏僻农村 144 名妇女生育情况如下：

生育胎次	0	1	2	3	4
妇女人数	5	25	70	30	14

该资料的类型是：

- A. 有序资料 B. 计数资料 C. 名义资料 D. 计量资料

2.2 用某种新疗法治疗某病患者 41 人，治疗结果如下：

治疗结果	治愈	显效	好转	恶化	死亡
治疗人数	8	23	6	3	1

该资料的类型是：

- A. 有序资料 B. 名义资料 C. 计数资料 D. 计量资料

2.3 某血库提供 6094 例 ABO 血型分布资料如下：

血型	O	A	B	AB
人数	1823	1598	2032	641

该资料的类型是：

- A. 计量资料 B. 有序资料 C. 计数资料 D. 名义资料

2.4 某医院一年内收治 202 例腰椎间盘突出后突病人，其年龄的频数分布如下：

年龄(岁)	10~	20~	30~	40~	50~	60~
人数	6	40	50	85	20	1

为了形象表达该资料，适合选用：

- A. 线图 B. 条图 C. 直方图 D. 圆图

2.5 某地 1968~1976 年两种传染病的发病率(%) 如下：

年份	1968	1970	1972	1974	1976
甲病发病率(%)	18.1	16.1	16.7	19.1	17.6
乙病发病率(%)	1.1	1.3	0.9	0.7	0.5

为了形象表达该资料，适合选用：

- A. 直方图      B. 普通线图      C. 百分条图      D. 半对数线图

2.6 学习和使用统计学，必须弄清“标准误”的概念，它的正确解释是：

- A. 样本均值的标准差    B. 样本率的标准差    C. 标准差的标准差    D. 统计量的标准差

2.7 在对两个变量  $x$  与  $y$  进行直线相关分析后发现：相关系数  $r$  的值近似等于零，经检验，得  $P>0.05$ 。下专业结论时，正确的表述应该是：

- A.  $x$  与  $y$  之间呈直线关系      B.  $x$  与  $y$  之间呈曲线关系  
C.  $x$  与  $y$  之间没有直线关系      D.  $x$  与  $y$  之间毫无关系

2.8  $x_1$ 、 $x_2$  分别代表给大鼠注射一定剂量的内毒素后，测得大鼠的红细胞与血红蛋白含量，求得直线回归方程为： $\hat{x}_2 = 6.188441 + 0.199953x_1$ ,  $n=19$ ,  $r=0.82031$  检验该直线回归方程是否有显著性意义的统计结论应该是：查  $r$  临界表

- A.  $P>0.05$     B.  $0.01 < P < 0.05$     C.  $0.005 < P < 0.01$     D.  $P < 0.001$

2.9 设某人群的身高  $x$  服从  $N(155.4, 5.3^2)$  分布，现从该总体中随机抽出一个  $n=10$  的样本，得均值为  $\bar{x}=158.36$ 、 $s=3.83$ ，求得  $\mu$  的 95% 置信区间为： $(155.62, 161.10)$ ，发现该区间竟然没有包括真正的总体均值 155.4。若随机从该总体抽含量  $n=10$  的样本 200 个，每次都求 95% 置信区间，问大约有多少个类似上面的（即不包括 155.4 在内）置信区间？

- A. 5 个      B. 20 个      C. 10 个      D. 1 个

2.10 有人根据某种沙门菌食物中毒患者 164 例的潜伏期资料，用百分位数法求得潜伏期的单侧 95% 上限为 57.8 小时，其含义是：

- A. 大约有 95 人的潜伏期小于 57.8 小时；    B. 大约有 69 人的潜伏期大于 57.8 小时；  
C. 大约有 5 人的潜伏期小于 57.8 小时；    D. 大约有 8 人的潜伏期大于 57.8 小时。

2.11 对 15 名 4~6 月贫血儿童用枸橼酸铁铵 + 维生素 C 治疗前后测得血色素 (g%) 含量见表 1。

表 1 15 名儿童贫血治疗前后血色素的测定结果

患者编号	血色素含量(g%)			
	疗前	疗后 1 月	疗后 2 月	疗后 3 月
1	11.5	10.4	12.0	12.0
2	10.7	11.0	12.0	13.0
...	...	...	...	...
14	9.5	10.1	12.1	13.5
15	10.0	10.5	11.3	12.5

假定资料满足各种参数检验所要求的前提条件，那么，适合分析该资料的方法是：

- A.  $\chi^2$  检验      B. 单因素方差分析      C. t 检验      D. 配伍组设计方差分析

2.12 20 名糖尿病人的血糖水平 (mg/100ml) 与胰岛素水平 ( $\mu$ u/100ml) 的测定值列于表 2 中。

表 2 20 名糖尿病人的血糖水平与胰岛素水平的测定结果

病例号	胰岛素水平	血糖水平
1	15.20	220
2	16.70	262
...	...	...
19	13.70	225
20	24.40	166

适合分析该资料的方法是：

- A. 配对比较的 t 检验      B. 成组比较的 t 检验  
C. 直线相关与回归分析      D.  $\chi^2$  检验或方差分析

2.13 将钩端螺旋体病人的血清分别用标准株和水生株作凝溶试验，测得稀释倍数如下，问两组的平均效价之间的差别有无显著性意义？

标准株 (6 人): 100, 200, 400, 800, 600, 3200

水生株 (5 人): 100, 100, 200, 200, 400

为回答上述问题，最有可能被选用的假设检验方法是：

- A. 直线相关与回归分析      B. 分别计算几何均值  
C. 两几何均值比较的 t 检验      D. 两算术均值比较的 t 检验

2.14 在针刺麻醉下对三组患者进行肺切除术，效果分四级，资料见表 3。试分析三组病人针麻效果之间的差别有无显著性意义。

表 3 三组患者肺切除术的针麻效果的比较

针 麻 效 果	例 数			合 计
	患者组别: 肺癌组	肺化脓症组	肺结核组	
I	10	24	48	82
II	17	41	65	123
III	19	33	36	88
IV	4	7	8	19
合 计	50	105	157	312

最适合分析该资料的方法是：

- A.  $\chi^2$  检验      B. Q 检验      C. 秩和检验      D. F 检验

3. 简单计算题 (36分)

3.1 选用合适的变异指标，初略地评价下列两组数据的变异度何者较大？6 只中年大鼠谷丙转氨酶含量 (u/L,  $x_1$ ) 和白蛋白含量 (g/L,  $x_2$ ) 的测定结果如下：

鼠号	1	2	3	4	5	6
$x_1$	38.5	36.0	41.1	39.1	43.2	64.8
$x_2$	28.0	26.0	30.0	24.0	28.0	30.0

3.2 设有 12 只鼠按体重由轻到重编为 1~12 号，请用下面给出的随机排列数字 (见表 4)，将它们均分到甲、乙、丙三个处理组中去。

表 4 随机排列简表 (n=20)

No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
	3	18	1	10	13	17	2	0	3	8	15	7	4	19	12	5	14	9	11	6	16

注：这是从 n=20 的随机排列表中抽出的表头和第 3 行随机数字。

3.3 抽样调查某单位 2839 名成年人某病的发病情况，得原始资料见表 5，请填补表内空白栏，并求出总发病率的 95% 置信区间。

表 5 某单位 2839 名成年人某病的发病及年龄分布情况

年龄组	受检人数	病例数	患者构成比 (%)	发病率 (%)
20~	1045	9		
30~	443	13		
40~	702	91		
50~	637	102		
60~	12	12		
合计	2829	227		

4. 假设检验题 (24分)

4.1 用免疫酶法观察鼻咽癌患者 (甲组) 头颈部其他恶性肿瘤患者 (乙组) 及正常人 (丙组) 血清中的 EB 病毒壳抗原的免疫球蛋白 A (VCA-IgA) 抗体的反应情况，资料见表 6。问三组阳性率之间的差别有无显著性意义？

表 6 三组受试者中 EB 病毒 VCA-IgA 抗体阳性结果

组 别	阳性例数	合计例数
甲 组	188	204
乙 组	10	33
丙 组	49	382
合 计	247	619

4.1 试就表 7 资料说明大白鼠感染脊髓灰质炎病毒后，再作伤寒或百日咳预防接种是否影响生存日数 (假定该资料满足各种参数检验所要求的前提条件，不要求作两两比较)。

表 7 感染脊髓灰质炎病毒后的三组大鼠接受不同处理后的生存情况

编 号	生存日数		
	处理：接种伤寒疫苗	接种百日咳疫苗	空白对照
1	5	6	8
2	7	6	9
3	8	7	10
4	11	10	12
5	11	16	14
6	12	11	16