心理统计 Ultimate

第一章　绪论

测量 简答

1. 数据类型的划分：观测方法、反映的测量水平（称名、顺序、等距、比率）、是否连续（选择）17 选（按名次排列的成绩是顺序量表）16 简答（简述测验的量表水平）统 09 选（职称是顺序 薪水是等比）

第二章　统计图表

17 多：一般可以选用来表达内容的图包括：条形图、直方图、线型图、圆形图

1. 分组次数表的优点 3、缺点 1（归组效应）
2. 累加次数分布表的优点：方便了解到某一分组区间上限以下或下限以上数据总数 选择 统 13 选
3. 条形图和直方的区别：数据类型/表示数据多少的方式/坐标轴上的标尺分点意义/图形直观形状 P46
4. 圆形图=饼图适用：间断性资料 各部分在整体中所占的比重大小 统 15 选
5. 线形图的作用：两个变量间的函数关系/一种现象随另一种现象变化的情况/某种现象在时间上的发展趋势/不同人在相同现象上的联系 统 10 选
6. 散点图的作用：表示两种现象的相关关系 统 07 选

第三章　集中量数

1. 平均数的特点：离均差和为 0/线性关系 P57 16 选（研究报告中平均数附加标准差为了说明平均数的代表性）统 13 选（平均数的线性关系）
2. 计算应用平均数原则：3 个 同质性原则 平均数与个体数值相结合原则 平均数与标准差方差结合原则
3. 在研究报告中，平均数生附加了标准差是为了说明平均数的代表性 16 选择
4. 平均数优点缺点：优：反应灵敏、计算严密、计算简单、简明易解、适于进一步代数运算、较少受抽样变动的影响|缺点：易受极端值影响、模糊不清的数据出现时不能用平均数 选择 16 选（去掉一个最高分，去掉一个最低分为的是排除平均分的局限性）统 08 选（受极端值影响的是）
5. 中数计算（特别中间的数有重复的时候）统 12 选
6. 中数的优缺点：计算简单、容易理解、不受极端值影响|不够灵敏、受抽样影响大、计算麻烦先要排序、md*n≠和、不能进一步代数运算
7. 皮尔逊经验法：$M_o = 3M_d - 2M$ 了解
8. 双峰分布时有两个众数 选择 统 12 选
9. 众数优缺点：简单明了、易于理解，较少受极差值的影响|缺点和上面一样 选择 统 09 选（一组正态分布数据中去掉极端值一定不会受到影响的是众数）统 15 选（正态两端加上一个极端值一定不受影响的是众数）
10. 正态、正偏态、负偏态平均数、中数与众数的关系 16 选（高分段人多是负偏态=少量人低分）统 10 选×3
11. 三个别的平均数（加权平均数：所得数据权重不等时；几何平均数：实验数据有少数偏大偏小[心理物理学等距量表、计算平均增长率]；调和平均数：学习速度）

第四章　差异量数

1. 百分等级的计算 计算 统 07 选（百分等级的意义）统 10 选（累加分布表可以快速算出百分等级）

（1）通常先求 Z 分数，再转换成概率，就是百分等级

（2）已经排名算（要修正 0.5）统 12 选（计算）

16 选 全距是说明数据离散程度最简单的统计量

2. 方差的原始计算式 $\sigma^2=\frac{\sum_{i=1}^{n}X_i^2}{N}-\left(\frac{\sum_{i=1}^{n}X_i}{N}\right)^2$ 基本能力
3. 总方差的合成 $S_T^2=\frac{\sum N_i s_i^2+\sum N_i d_i^2}{\Sigma N_i}$ P90 例题 选择
4. 标准差的线性性质 选择 13 选（方差的线性关系）
5. 方差是最常用的统计特征数
6. 标准差的优点：反应灵敏、计算机严密、容易计算（然而不好算）、适合代数运算、受抽样变动影响小、简单明了
7. 切比雪夫定律：连接方差和概率 $P\{|X-\mu|\le k\sigma\}\ge 1-\frac{1}{k^2}$ （在 k 个标准差内至少有 $1-\frac{1}{k^2}$ 的数据） 16 选（正态分布中距离平均数上下 3 个标准差）
8. 差异系数计算/适用：同一团体不同观测值离散程度、水平相关较大的两个团体，同一观测值离散程度的比较 选择 统 08 选（计算） 统 11 选（给两组方差和平均数，问离散程度） 统 14 选（适用）
9. 标准差的性质：无实际单位，平均数为参照点，标准差为单位（但是还是没有单位）/Z 的和是 0，平均数是 0，标准差是 1/原始数据正态分布，则转换成 Z 变为标准正态分布 统 11 选（正态的 Z 的标准差是 1）
10. 标准分数评价：优点：4 点 可比、可加、明确、稳定 缺点：3 点 计算繁杂、负值零值小数（单位过大）、比较时还要满足原始数据分布形态相同 18 多（缺点）
11. 标准分数三个应用：不同观测值在团体中的相对位置/不同观测值求和/表示标准测验分数/制作常模

第五章　相关关系

1. 相关系数的解释：①有相关系不一定存在因果②r 是一个比值不能四则运算，不等距③受样本容量和取样的影响④纯理论研究中即使是很小的相关，若在统计中显著也能说明心理问题⑤线性回归时也是决定系统 r^2 表示两变量共变的比例 选择 统 10 多（散点图是一条直线，两个变量方差均不为 0，它们之间的相关关系可能为 1 或-1） 统 11 选（$r_1=-r_2$ 则它们俩表示的相关程度相同 ~~它们表示的意义相同~~） 统 13 选（r=0 代表不存在线性相关关系）
2. 相系系数的合并（r 不等距）：转换成等距的 Z_r 再修正系数 n-3,再转回去：$\bar{Z}_r=\frac{\sum(n_i-3)Z_{ri}}{\Sigma(n_i-3)}$ 选择

   条件：样本接近、研究事物接近、测量工具相同
3. 斯皮尔曼等级相关的计算 $r_R=1-\frac{3\Sigma D^2}{N(N^2-1)}$
4. 肯德尔 w 系数的计算 $W=\frac{\Sigma R_i^2-\frac{(\Sigma R)^2}{N}}{\frac{N^3-N}{12}K^2}$
5. 点二列相关的计算：$r_{pb}=\frac{\bar{X}_p-\bar{X}_q}{S_t}\sqrt{pq}$
6. ϕ 系数的计算 $r_\phi=\frac{ad-bc}{\sqrt{(a+b)(a+c)(b+d)(c+d)}}$ 统 08 选（给出一个 2*2 表计算）

7. 相关系数的适用 选择

占个位	积差相关	两个连续变量，成对，正态，线性
等级相关	斯皮尔曼相关	两列等级（一列是连续一列等级可以转换）
	肯德尔 W	多个评价者
	肯德尔 U 系数	对偶比较法
质与量相关	点二列相关	一列连续，一列二分
	二列相关	一列连续，一列人为二分（两列都是连续）
	多列相关	一列连续，一列人为多分
品质相关	四分相关	两列人为
	ϕ 系数	两列二分
	列联表相关	$R\times C$ 列联表

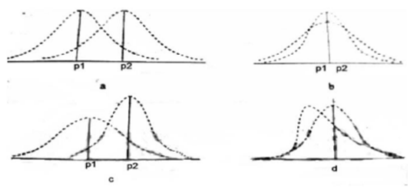
18 选：皮尔逊积差相关
假设两个变量正态分布

注：两列连续变量可以用的相关系数：积差相关（最好的）、斯皮尔曼（可以转换成等级）、二列相关（人为二分实质上是连续变量分的

第六章　概率分布

统 16 选

、关于下列四个图形的描述，错误的是（ ）



A、图 A 中两个随机变量的均值不同，方差相同
B、图 B 中两个随机变量的均值相同，方差不同
C、图 C 中两个随机变量的均值、方差均不相同
D、图 D 中两个随机变量的均值相同，方差也相同

1. 区分先验概率和后验概率 选择
2. 标准正态分布在 0 处的值是 $\frac{1}{\sqrt{2\pi}}=0.3989$
3. 正态分布的拐点在 $\pm\sigma$ 处，即 $Z=\pm1$
4. 记住正态分布的几个常用的概率和正态分布的转换（统计推断必用）
   $\pm\sigma:68.26\%$
   $\pm1.96\sigma:95\%\ \pm2.58\sigma:99\%$(双侧)　17 选（±2.58 99%）11 专选（±3 99.73%）
   $\le95\%:1.645\sigma\le99\%:2.323\sigma$(单侧)
5. 正态分布的应用：把老师给的等级转换成分数、由题目通过率确定题目难度、能力分组时确定人数、测验分数正态化
6. 正态分布和标准正态分布的区别与联系 简答（统考 2008）
7. $n\to\infty, B(n,p)\to N\{np,np(1-p)\}$ 大样本的条件 $\min(p,1-p)\cdot n\ge5$ 统 09 选（100 题起码 27 题）
8. 答对多少题才能说明不是猜的？：二项趋于正态，单侧 16 选（选择题的 P 大于概率）
9. 样本平均数的分布 $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}:Z(\sigma\ known)$　$\frac{\bar{X}-\mu}{S/\sqrt{n}}:t_{n-1}(\sigma\ unknown)$　$\frac{\bar{X}-\mu}{S/\sqrt{n}}:Z(big\ sample)$ 统 10 选
10. 样本标准差、方差的分布 $s:N(\sigma,\frac{\sigma}{\sqrt{2n}})$　$\frac{s_{n-1}^2(n-1)}{\sigma^2}:\chi_{n-1}^2$
11. t 分布和标准正态分布的关系：t 分布是重尾，方差比 Z 分布大
12. t 分布、χ^2 分布、F 分布自由度越多，越接近正态分布 17 选（只有标准正态与自由度无关）
13. $E(\chi_n^2)=n\ \ Var(\chi_n^2)=2n\ \ df=n$ 统 12 选

第七章　参数估计

1. 良好点估计的标准：无偏性（$\sum\hat{g}(X_1,\ldots,X_n)=g(\theta_1,\ldots,\theta_n)$ $M_o,M_d,\bar{X}$ 都是 μ 的无偏估计）、有效性=最小方差无偏性、一致性=相合性（$N\to\infty\ \hat{g}(X_1,\ldots,X_n)=g(\theta_1,\ldots,\theta_n)$）、充分性（用到了每个数据）

2. 置信区间=置信间距=2d；置信界限；显著性水平=意义阶段=信任系数=α；置信度=置信水平=$1-\alpha$
3. 缩小置信区间的方法：3 个 减小置信度 减少样本统计标准误 增大样本容量 $\mu=\bar{X}\pm\frac{Z\cdot\sigma}{\sqrt{n}}$
4. 区间估计的原理：样本分布理论
5. 总体平均数的估计：$\bar{X}\pm d$

   ① σ 已知：$d=\frac{\sigma}{\sqrt{n}}Z(\alpha/2)$

   ② σ 未知：$d=\frac{S}{\sqrt{n}}t_{n-1}(\alpha/2)$

   ③ 大样本：$d=\frac{s}{\sqrt{n}}Z(\alpha/2)$
6. 两样本平均数差的估计：$\left(\bar{X}_1+\bar{X}_2\right)\pm d$

   ① σ_1,σ_2 已知，$d=\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\cdot Z(\alpha/2)$

   ② σ_1,σ_2 未知，但相等 $d=\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\cdot S_T\cdot t_{n_1+n_2-2}$ $S_T=\frac{1}{n_1+n_2-2}\left[(n_1-1)S_1^2+(n_2-1)S_2^2\right]$

   ③ 大样本 $d=\sqrt{\frac{S_1^2}{n_1}+\frac{S_2^2}{n_2}}\cdot Z(\alpha/2)$
7. 二项分布比率的估计：$\hat{p}\pm\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}Z(\alpha/2)$

第八章　假设检验

1. 假设检验的原理：反证法、小概率事件原理
2. 假设检验前提（甘 P65-66）:随机抽样、独立观察、原总体标准差保持恒定、样本均值正态分布
3. 假设检验两个假设：虚无假设=无差假设=零假设=原假设 H_0；对立假设=备择假设 H_1
4. 假设检验中两类错误：Ⅰ型错误=α 型错误=弃真错误 Ⅱ型错误=β 型错误=存伪错误
5. 两个错误概率的关系：① $\alpha+\beta$ 不一定等于 1 ② α 和 β 不可能同时变大或变小在其它条件不变的情况下，α 变大，β 变小
6. 检验量的值大于临界值，不一定可以拒绝 H_0，还可能犯 α 错误
7. 统计检验力=统计功效=效力（power）：$1-\beta$ （甘 P67-68，统 16 多）

   （1）处理效应大小：处理效应越明显，越容易被侦查到，假设检验效力越大；

   （2）显著性水平 α：α 越大，统计检验力越大；

   （3）检验方向性：单尾检验检验力大于双尾；

   （4）样本容量：样本容量越大，统计效力越高。
8. 总体均值的检验：检验量 ≥ 拒绝域

   （1）σ 已知：$Z=\frac{\left|\bar{X}-\mu_0\right|}{\sigma/\sqrt{n}}\ge Z(\alpha/侧)$

   （2）σ 未知：$t=\frac{\left|\bar{X}-\mu_0\right|}{s/\sqrt{n}}\ge t_{n-1}(\alpha/侧)$

   （3）大样本：$Z=\frac{\left|\bar{X}-\mu_0\right|}{s/\sqrt{n}}\ge Z(\alpha/侧)$

9. 平均数差异的检验

（1）独立样本

① σ_1,σ_2 已知：$Z=\dfrac{\left|\bar{X}_1-\bar{X}_2\right|}{\sqrt{\dfrac{\sigma_1^2}{n_1}+\dfrac{\sigma_2^2}{n_2}}}\geq Z(\alpha/\text{侧})$

② σ_1,σ_2 未知，但相等：$t=\dfrac{\left|\bar{X}_1-\bar{X}_2\right|}{\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}s_T}}\geq t_{n_1+n_2-2}$

③ 大样本：$Z=\dfrac{\left|\bar{X}_1-\bar{X}_2\right|}{\sqrt{\dfrac{s_1^2}{n_1}+\dfrac{s_2^2}{n_2}}}\geq Z(\alpha/\text{侧})$

（2）相关样本

① σ_1,σ_2 已知：$Z=\dfrac{\left|\bar{X}_1-\bar{X}_2\right|}{\sqrt{\dfrac{\sigma_1^2+\sigma_2^2-2r\sigma_1\sigma_2}{n}}}\geq Z(\alpha/\text{侧})$

② σ_1,σ_2 未知：

r 未知，把 $\bar{X}_1-\bar{X}_2$ 当成一个新的随机变量 $t=\dfrac{\left|\bar{X}_1-\bar{X}_2\right|}{s_{\bar{X}_1-\bar{X}_2}/\sqrt{n}}\geq t_{n-1}$

r 已知：$t=\dfrac{\left|\bar{X}_1-\bar{X}_2\right|}{\sqrt{\dfrac{s_1^2+s_2^2-2rs_1s_2}{n}}}\geq t_{n-1}$

③ 大样本：$Z=\dfrac{\left|\bar{X}_1-\bar{X}_2\right|}{\sqrt{\dfrac{s_1^2+s_2^2-2rs_1s_2}{n}}}\geq Z(\alpha/\text{侧})$

10. 一样本方差的检验：$\chi^2=\dfrac{(n-1)s^2}{\sigma^2}$ 接受域：$\left[\chi_{n-1}^2(1-\alpha/2),\chi_{n-1}^2(\alpha/2)\right]$

11. 两样本方差的检验：

（1） 独立样本：$F=\dfrac{s_1^2}{s_2^2}$ 接受域：$\left[\dfrac{1}{F_{n_1-1,n_2-2}(a/2)},F_{n_1-1,n_2-2}(a/2)\right]$（可以用大方差比小方差然后只比较上界）

（2） 相关样本 $t=\dfrac{s_1^2-s_2^2}{\sqrt{\dfrac{4s_1^2s_2^2\left(1-r^2\right)}{n-2}}}\geq t_{n-2}(\alpha/\text{侧})$

第九章 方差分析

1. 方差分析假定 选择 选择 17 多 统 08 多

（1）F 分布的基本假设

① 总体正态分布

② 独立性

③ 变异的同质性⇒检验方差是否性（哈特莱最大 F 比率法 $F=\dfrac{S_{\max}^2}{S_{\min}^2}$ ）

（2）实验设计模型及其假设（以单因素完全随机实验设计为例）

$Y_{ij}=\mu+a_j+e_{ij}$（Y_{ij}-第 i 个被试在第 j 个处理水平上的值 μ-总体平均数，用 $\bar{Y}_{..}$ 估价 e_{ij}-误差变异，服从平均数为 0 的正态分析用 $\bar{Y}_{ij}-\bar{Y}_{.j}$ 估价）

① 模型反映了影响实验中观测值 Y_{ij} 的所有变异源；

② 实验中包含了研究者感兴趣的处理水平 a_j；

③ 误差变异在每个处理总体内是以均数为 0，方差为 σ_e^2 的正态分布。每一个被试的误差变异都独立于其他被试的误差变异

2. 不能用 t 检验对多个平均数的差异进行比较的原因：各组平均数两两成对多进行几次 t 检验，α 错误的概率增加，本来达不到显著差异很容易被说成是显著的 选择

3. $X_1,\ldots,X_n$ 的和方 $SS=\sum_{i=1}^{n}(X_i-\bar{X})^2=\sum_{i=1}^{n}X_i^2-\dfrac{\left(\sum_{i=1}^{n}X_i\right)^2}{n}$ 方差 $s^2=\dfrac{SS}{n-1}$

4. 协方差分析 统 14 选

（1）定义：调节协变量对因变量的影响效应，结合了方差分析与回归分析

（2）意义：研究中有些变量会影响因变量但是不能控制，研究者也知道其存在，这些就是协变量，在实验处理前给予观测，在统计时用协方差分析的方法进行处理，将协变量对因变量的影响分离出去，进一步提高实验精确度和统计检验灵敏度（控制无关变异）

5. 不同的平方和的求法及自由度的确定（符号用的甘老师的，见甘 P130）

（1） $SST=\sum_{i=1}^{n}X_i^2-\dfrac{G^2}{N}\quad df=N-1$

（2） $SSE=\sum_{i=1}^{k}SS_i\quad df=N-k$

（3） $SSA=\sum_{i=1}^{k}\dfrac{T_i^2}{n_i}-\dfrac{G^2}{N}=\sum_{i=1}^{k}\left[n_i\left(\bar{X}_i-\bar{G}\right)\right]\quad df=k-1$

6. 随机区组设计 k 个处理水平、l 个区组 思路让 SSE 更小，这样更容易显著（参考的甘的表述 见甘 P138） 习题 12

（1） $SST=\sum_{i=1}^{n}X_i^2-\dfrac{G^2}{N}\quad df=N-1$

（2） $SSA=\sum_{i=1}^{k}\dfrac{T_i^2}{n_i}-\dfrac{G^2}{N}\quad df=k-1$

（3） $SSW=\sum_{i=1}^{n}SS_i$ 进一步分解为

① $SSB=\sum_{i=1}^{n}\dfrac{P_i^2}{n_i}-\dfrac{G^2}{N}\quad df=l-1$

② $SSE=SSW-SSB\quad df=(k-1)(l-1)$

7. 重复测量方差分析的假设：①不同水平下的总体方差相等 ②每个处理条件内的观察都是独立的 ③不同处理水平下的总体服从正态分布 ④因变量的方差-协方差矩阵符合球形假设 统 15 多

8. 事后检验：N-K 检验法=q 检验法，Turkey 的可靠显著差异法 HSD（要求相等样本容量，费舍的最小显著差异法 LSD，Scheffé 检验法（保守，适用样本容量不相等）

9. 二因素方差分析（18 论述 给原始数据进行方差分析）

（1）交互作用：实验考查的效应①A 的主效应 ②B 的主效应 ③$A\times B$ 交互效应

（2）和方计算：A 因素 k 个水平，B 因素 l 个水平

① $SST=\sum_{i=1}^{N}X_i^2-\frac{G^2}{N}\quad df=N-1$

② $SSE=\sum_{i=1}^{k+l}SS_i\quad df=N-kl$

③ $SS(A+B)=\sum_{i=1}^{n}\frac{T_i^2}{n}-\frac{G^2}{N}$ 进一步分解为

$$SSA=\sum_{i=1}^{k}\frac{A_i^2}{忽略B分类的A_i组的人数}-\frac{G^2}{N}\quad df=k-1$$

$$SSB=\sum_{i=1}^{l}\frac{B_i^2}{忽略A分类的B_i组的人数}-\frac{G^2}{N}\quad df=l-1$$

$$SSA\times B=SS(A+B)-SSA-SSB\quad df=(k-1)(l-1)$$

10. 简单主效应：甘 P156~161 18 简答（事后检验与简单效应检验的区别）
11. 效果量：张 P221 P140 （18 论述 方差分析后算效果量）

    单因素实验设计 $\eta^2=\frac{SSA}{SST}$ 重复测量实验设计 $\eta^2=\frac{SS_{重复}}{SST}$ 多因素实验设计 $\eta^2=\frac{SSA}{SST}$ $\eta^2=\frac{SSB}{SST}$ $\eta^2=\frac{SS_{AB}}{SST}$

    测量自变量效果的量数，反映自变量和因变量的关联程度，提供差异大小的信息，是实验处理的效应大小（统 16 选）
12. 算自由度、人数、组数（列表）：统 11 选（k=3,df_E=27，则 n=30） 统 16 选（k=3,n_i=12，则 df_E=33） 统 17 选（k=3 n_i 相同,df_T=30，则 n_i=10）
13. 考察性别（二分）是否是一项认知测验中的反应时（连续）的影响因素，可以用的统计方法是：3 个 t 检验 方差分析 点二列相关 多选

- 完全随机设计的方差分析

一个因素 A，k 个水平（k 个种子品种/分成了 k 组），一共 n 个实验对象（n 个种子、k 个被试）

变异来源	*SS*	自由度	*MS*	*F* 比	*F*
因素 *A*	SS_A	*k*-1	$MS_A=\frac{SS_A}{k-1}$	MS_A/MS_E	$F_{k-1,n-k}(\alpha)$
误差	SS_E	*n*-*k*	$MS_E=\frac{SS_E}{n-k}$		
总和	SS_T	*n*-1			

- 随机区组设计方差分析

变异来源	*SS*	自由度	*MS*	*F* 比	*F*
因素 *A*	SS_A	*k*-1	$MS_A=\frac{SS_A}{k-1}$	MS_A/MS_E	$F_{k-1,(k-1)(l-1)}(\alpha)$
因素 *B*	SS_B	*l*-1	$MS_B=\frac{SS_B}{l-1}$	MS_B/MS_E	$F_{l-1,(k-1)(l-1)}(\alpha)$
误差	SS_E	(*k*-1)(*l*-1)	$MS_E=\frac{SS_E}{(k-1)(l-1)}$		
总和	SS_T	*kl*-1			

- 多因素方差分析

两个因素 A、B，A 有 k 个因素，B 有 l 个因素（A 为品种，有 k 个；B 为播种量，有 l 个）

变异来源	***SS***	自由度	***MS***	***F*** 比
因素 A	SS_A	k-1	$MS_A=\dfrac{SS_A}{k-1}$	MS_A/MS_E
因素 B	SS_B	l-1	$MS_B=\dfrac{SS_B}{l-1}$	MS_B/MS_E
$A\times B$	SS_{AB}	(k-1)(l-1)	$MS_{AB}=\dfrac{SS_E}{(k-1)(l-1)}$	MS_{AB}/MS_E
误差	SS_E	N-kl	$MS_E=\dfrac{SS_E}{N-kl}$	
总和	SS_T	N-1		

第十章　Chi square 检验

1. 适用：处理一个因素两项或多项分类的实际观察频数与理论频数分布是否一致的问题 选择
2. χ^2 检验的假设：分类相互排斥互不包容、观测值相互独立、期望次数至少 5 多选
3. 小期望次数的性校正方法：①单元格合并法 ②增加样本数 ③去除样本法 ④使用校正公式 多选
4. χ^2 检验可以检验的类别：①配合度检验（拟合优度检验）②独立性检验 ③同质性检验
5. χ^2 的构造：$\chi^2=\sum_{i=1}^{n}\dfrac{(O_i-E_i)^2}{E_i}$ （Observation Expectation）
6. 应用：①拟合优度检验（自由度=组数-1）②正态分布匹配度检验（自由度=组数-3）③独立性检验（$R\times C$ 列联表 χ^2 自由度为 $(R-1)(C-1)$ ）
7. 四表格独立性检验：

（1）独立样本 $\chi^2=\dfrac{N(AD-BC)^2}{(A+B)(A+C)(B+D)(C+D)}$ 　$df=1$

（2）相关样本（同一批人重复测）$\chi^2=\dfrac{(A-D)^2}{A+D}$ 　$df=1$

第十一章 非参数检验

1. 适用：总体分布不明确、资料的计量信息较弱弱（顺序、等级） 选择
2. 非参数统计的特点：①没有严格的前提假设 ②特别适合于顺序资料 ③适用于小样本 ④不能利用资料全部的信息（把数据转换成等级）⑤不能处理交互作用 多选
3. 非参数检验方法：选择

（1）独立样本：秩和检验法=曼-惠特尼检验=曼-惠特尼 U 检验=维尔克松两样本检验法（总体非正态时，代替 t 检验）、中数检验法（资料不完整，偏态，平均数不适用时）

（2）相关样本：符号检验法、维尔克松符号等级检验法

（3）等级方差分析（代替方差分析）：克-瓦氏单向方差分析、弗里德曼两因素等级方差分析

第十二章 线性回归

1. 线性回归与相关分析的关系（简答题先破题，解释一下关键词是什么意思）：回归分析是以数学的方式表示变

量间的关系，相关分析是检验或度量这些关系的密切程度，两都相辅相成（统 14 简答）

2. 线性回归的假设：①XY 关系线性 ②同一个 X 对应的 Y 正态 ③独立（不同 X 对应的 Y 独立/对应的误差独立）④误差等分 多选 17 多
3. 回归方程系统的求法 $Y=\beta_0+\beta_1\left(X-\bar{X}\right)+e \quad \hat{\beta}_0=\bar{Y} \quad \hat{\beta}_1=\dfrac{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)Y_i}{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2}$
4. 回归系数与相关系数的关系 $\beta_{1X\cdot Y}=r\cdot\dfrac{S_X}{S_Y} \quad \beta_{1Y\cdot X}=r\cdot\dfrac{S_Y}{S_X} \quad r=\sqrt{\beta_{1X\cdot Y}\cdot\beta_{1Y\cdot X}}$
5. 回归模型的检验：①有效性：方差分析（残差度 $df_E=N-2$ 回归自由度 $df=1$ 统 12 选）②显著性：t 检验

17 选：回归分析中，关于确定系数，下列说法正确的是 A

A.确定系统等于相关系数的平方　B.确定系数等于回归系数的平方

C.确定系数等于相关系数　　　　D.确定系数等于回归平方和除以残差平方

6. 对线性回归方程有效性进行显著性检验的方法有：4 个 方差分析 回归系数检验 决定系数和相关系数拟合度的测定 估计标准误差的计算 多选
7. 回归效果（XY 的相关程度）：决定系数 $r^2=\dfrac{\sum_{i=1}^{n}(\hat{Y}_i-\bar{Y})^2}{\sum_{i=1}^{n}(Y_i-\bar{Y})^2}=\dfrac{SSR}{SST}$ 意义：说明 Y 中的变异有 r^2 是由变量 X 引起的
8. 多重线性回归自变量的选择方法：4 个 最优方程选择法 同时多重回归法 逐步多重回归法 层次多重回归法 17 多
9. 因子分析的目的（降维）因子的确定：特征根 $\lambda\geq1$ 可被视为一个因子

第十三章 抽样原理及方法

1. 抽样的基本原则：随机化 统 08 选
2. 抽样方法：①简单随机抽样 ②等距抽样（不适用于总体周期性变化 统 15 选）③分层随机抽样（每层抽样的人数据的确定 统 17 选）④两阶段随机抽样（总体容量很大时） 选择 统 11 选（分层随机抽样的特点是层间异质，层内同质）
3. 样本容量的计算：由 $\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}}=Z_{\alpha/2} \quad n=\left(\dfrac{Z_{\alpha/2}\cdot\sigma}{d}\right)^2 \quad n=\left(\dfrac{t_{\alpha/2}\cdot\sigma}{d}\right)^2 \quad n=\left(\dfrac{Z_{\alpha/2}\cdot\hat{p}\left(1-\hat{p}\right)}{d}\right)^2$ 张 P119 选择