

本试卷适用范围
经济管理学院
2018 级本科生

南京农业大学试题纸

2019-2020 学年 1 学期

课程类型：必修

试卷类型：A

课程 应用统计学

班级

学号

姓名

特别提醒：请各位同学严格遵守考试纪律。

题号	一	二	三	四	总分	签名
得分						

一、名词解释（共 20 分，每题 5 分）

名词解释答出概念就行了，展开不展开分析都对。

1. 假设检验

答案：事先对总体参数或分布形式作出某种假设（如果是对总体均值 μ 进行检验，则是对 μ 的值提出一个假设），然后利用样本信息来判断原假设是否成立，有参数假设检验和非参数假设检验。采用逻辑上的反证法，依据统计上的小概率原理。

2. 回归模型

答案：描述因变量 y 如何依赖于自变量 x 和误差项 ε 的方程称为回归模型。

3. 组内平方和

答案：反映组内误差大小的平方和称为组内平方和，也称为误差平方和或残差平方和，记为SSE。例如，每个样本内部的数据平方和加起来就是组内平方和，它反映了每个样本内各观测值的离散状况。

4. 平稳序列

答案：基本上不存在趋势的序列，各观察值基本上在某个固定的水平上波动；或虽有波动，但并不存在某种规律，而其波动可以看成是随机的。

二、选择题（共 10 分，每题 2 分）

1. 如果让同一批打字员分别使用不同型号的打字机打字，这时的样本是（ ）

A. 大样本 B. 小样本 C. 独立样本 D. 匹配样本

答案：D

2. 变量之间的关系可以分为两大类（ ）。

A 函数关系与相关关系 B 线性相关关系和非线性相关关系
C 正相关关系和负相关关系 D 简单相关关系和复杂相关关系

答案：A

装
订
线

装
订
线

3. 组内误差是衡量某一水平下样本数据之间的误差，它（ ）

- A. 只包括随机误差
- B. 既包括随机误差，也包括系统误差
- C. 只包括系统误差
- D. 既包括随机误差，也包括系统误差

答案：A

4. 包含趋势性、季节性或周期性的序列称为（ ）

- A、平稳序列
- B、季节序列
- C、周期序列
- D、非平稳序列

答案：D

5. 移动平均法适合于预测（ ）

- A、非平稳序列
- B、有季节成分序列
- C、有趋势成分序列
- D、平稳序列

答案：D

三、简答题（共 32 分，每题 8 分）

1. 解释假设检验中的 P 值，并阐述如何用 P 值进行检验判断及影响 P 值的三个因素

答案：（1）p 值是当原假设为真，样本观察结果或更极端结果出现的概率。是抽样分布中大于或小于样本统计量的概率。

（2）利用 P 值进行检验：若 p 值 $> \alpha$, 不拒绝 H_0 若 p 值 $< \alpha$, 拒绝 H_0

（3）影响 P 值的三个因素:一是样本数据与原假设之间的差异；二是样本量；三是被假设参数的总体分布。

注：简答题第一题的 p 值概念答成类似犯第一类错误的精确概率也可以，即拒绝原假设所犯错误的精确可能性。 检验判断说成 $p < \alpha$ 拒绝原假设也对，或 $p > \alpha$ 不能拒绝原假设。 影响 p 值的因素第三因素总体分布答成总体方差或样本方差(总体离散度或样本离散度)都可以;除此以外影响 P 值第四因素是双尾还是单尾检验，双尾的 P 值是单尾 p 值的 2 倍。如果学生答出这一点不能认为错，因为参考答案没这一点。

2. 简述相关系数的性质。

答案：

- （1）r 的取值范围是 $[-1, 1]$ 。
- （2）r 具有对称性。

- (3) r 的大小与 x 和 y 的原点及尺度无关。
- (4) r 仅仅是 x 与 y 之间线性关系的一个度量，它不能用于描述非线性关系。
- (5) r 虽然是两个变量之间线性关系的一个度量，却不意味着 x 与 y 一定有因果关系。

3. 简述方差分析的基本步骤。

答案：方差分析的基本步骤如下：

(1) 提出假设

在方差分析中，原假设所描述的是在按照自变量的取值分成的类中，因变量的均值相等。因此，检验因素的 k 个水平（总体）的均值是否相等，需要提出如下形式的假设：

$$H_0: \mu_1 = \mu_2 = \dots = \mu_i = \dots = \mu_k$$

$$H_1: \mu_i \ (i=1,2,\dots,k) \text{ 不全相等}$$

如果拒绝原假设 H_0 ，则意味着自变量对因变量有显著影响；如果不拒绝原假设 H_0 ，则没有证据表明自变量对因变量有显著影响，不能认为自变量与因变量之间有显著关系。

(2) 构造检验的统计量

首先，计算各样本的均值；其次，计算全部观测值的总均值，它是全部观测值的总和除以观测值的总个数的结果；再次，计算各误差平方和，分别是总平方和 SST ，是全部观测值与总均值的误差平方和；组间平方和 SSA ，是各组均值与总均值的误差平方和；组内平方和 SSE ，是每个水平或组的各样本数据与其组均值的误差平方和；最后，计算统计量：例如，构造 F 统计量等于 MSA 除以 MSE ，其中 MSA 为组间平方和除以其自由度； MSE 为组内平方和除以其自由度。

(3) 作出统计决策

根据给定的显著性水平 α ，在 F 分布表中查找与分子自由度 $df_1=k-1$ 、分母自由度 $df_2=n-k$ 相应的临界值 $F_\alpha(k-1, n-k)$ 。

若 $F > F_\alpha$ ，则拒绝原假设，表明均值之间的差异是显著的；

若 $F < F_\alpha$ ，则不拒绝原假设，没有证据表明均值之间有显著差异。

4. 简述时间序列的构成要素。

答案：时间序列的构成要素可以分为 4 种：

趋势(T)在长时期内持续向上或持续下降的变动，

季节性(S)也称季节变动时间序列在一年内重复出现的周期性波动，

周期性(C) 也称循环波动围绕长期趋势的一种波浪形或振荡式变动，

随机性(I) 也称不规则波动除去趋势、周期性和季节性之后的偶然性波动。

四、计算分析题（共 38 分，第 1 题 8 分，其它 3 题各 10 分）

1. 为比较哪种排队方式使顾客等待的时间更短，银行各随机抽取了 10 名顾客，得到数据如下表：

方式 1	方式 2
6.5	4.2
6.6	5.4
6.7	5.8
6.8	6.2
7.1	6.7
7.3	7.7
7.4	7.7

7.7	8.5
7.7	9.3
7.7	10.0

假设方差相同，比较两种排队方式平均等待时间是否有显著差异？（ $t_{\alpha/2}=2.101$ ）

答案：

组统计量

种类	N	均值	标准差	均值的标准误
方式 方式1	10	7.1500	.47668	.15074
方式2	10	7.1500	1.82163	.57605

独立样本检验

	方差方程的 Levene 检验		均值方程的 t 检验						
	F	Sig.	t	df	Sig.(双侧)	均值差值	标准误差值	差分的 95% 置信区间	
								下限	上限
方 假设方	13.076	.002	.000	18	1.000	.00000	.59545	-1.25098	1.25098
式 差相等									
假设方			.000	10.227	1.000	.00000	.59545	-1.32276	1.32276
差相等									

2. 某汽车生产商欲了解广告费用(x)对销售量(y)的影响，收集了过去 12 年的有关数据。利用 excel 操作，Y 对 x 回归得到下面的有关结果：

方差分析表

变差来源	df	SS	MS	F	Significance F
回归	1				2.17E—09
残差		40158.07		—	—
总计	11	1642866.67	—	—	—

参数估计表

	Coefficients	标准误差	t Stat	P-value
Intercept	363.6891	62.45529	5.823191	0.000168
XVariable1	1.420211	0.071091	19.97749	2.17E—09

要求：

(1) 完成上面的方差分析表。

- (2) 计算判定系数，说明汽车销售量的变差中有多少是由于广告费用的变动引起的？
- (3) 销售量与广告费用之间的相关系数是多少？
- (4) 写出估计的回归方程并解释回归系数的实际意义。
- (5) 检验回归系数的显著性($\alpha=0.01$)。

答案：

(1)

变差来源	df	SS	MS	F	Significance F
回归	1	1602708.6	1602708.6	399.1000065	2.17E—09
残差	10	40158.07	4015.807	—	—
总计	11	1642866.67	—	—	—

(2) 根据方差分析表计算的判定系数 $R^2 = \frac{SSR}{SST} = \frac{1602708.6}{1642866.67} = 0.9756$ ，表明汽车销售量的变差中有 97.56% 是由于广告费用的变动引起的。

(3) 相关系数可由判定系数的平方根求得： $r = \sqrt{R^2} = \sqrt{0.9756} = 0.9877$ 。

(4) 回归方程为： $\hat{Y} = 363.6891 + 1.4202x$ 。回归系数 $\hat{\beta}_1 = 1.4202$ 表示广告费用每增加一个单位，汽车销量平均增加 1.4202 个单位。

(5) 由于 $p\text{-value} = 2.17E-09 < \alpha = 0.01$ ，拒绝原假设，回归系数显著。

3. 某企业践行国家节能减排政策，分别从A、B、C三国引进一套新型节排设备，准备采用这三套设备生产一种新型产品，为了确定哪种设备每小时生产的产品数量最多且效果最好，随机抽取30名工人，并指定每个人使用其中一种设备，通过对每个工人生产的产品数进行方差分析得到如下结果：

差异源	SS	df	MS	F	P-value	F
组间			210		0.245946	3.354131
组内	3836			—	—	—
总计		29	—	—	—	—

要求：

(1) 完成上面的方差分析表（写出具体解题过程）。

(2) 若显著性水平 $\alpha=0.05$ ，检验采用三种设备生产的产品数量之间是否有显著差异。

答案：(1) 方差分析表中所缺的数值如下表：

差异源	SS	Df	MS	F	P-value	F crit
-----	----	----	----	---	---------	--------

组间	420	2	210	1.47810219	0.245946	3.354131
组内	3836	27	142.0740741	—	—	—
总计	4256	29	—	—	—	—

注：红色加粗为答案，允许四舍五入，允许较小偏差存在。

本题具体解题过程：

据题知： $df_1=k-1=3-1=2$ ； $df_2=29-2=27$ ；

$MSE=3836/27=142.0740741$ ； $SSA=210*2=420$ ；

$SST=SSA+SSE=420+3836=4256$ ；

$F=MSA/MSE=210/142.0740741=1.47810219$

(2) 由方差分析表可知：

方法一：P-value=0.24594> $\alpha=0.05$ （或 $F=1.478<F_{\alpha}=3.354131$ ），不能拒绝原假设。没有证据表明三种方法组装的产品数量之间有显著差异。

方法二： $H_0: u_1=u_2=u_3$

$H_1: u_i (i=1,2,3)$ 不全相等

已知， $SSA=420$ ， $SSE=3836$ ， $MSA=210$ ， $MSE=142.07$

故 $F=MSA/MSE=210/142.07=1.4781$

根据 $\alpha=0.05$ ， $df_1=2$ ， $df_2=27$ ， $F_{0.05}(2, 27)=3.35$

因为 $F<F_{\alpha}$ ，则不能拒绝原假设 H_0 ，表明三种设备没有显著差异。

4. 下面是一家旅馆过去 18 个月的营业额数据：

月份	营业额（万元）	月份	营业额（万元）
1	295	10	473
2	283	11	470
3	322	12	481
4	355	13	449
5	286	14	544
6	379	15	601
7	381	16	587
8	431	17	644
9	424	18	660

要求根据数据进行一下分析：

(1) 采用指数平滑法，分别用平滑系数 $\alpha=0.3$ 和 $\alpha=0.5$ 预测各月的营业额，分析预测误差，说明哪一个平滑系数预测更合适，并预测第 19 个月的营业额。

(2) 建立一个趋势方程并预测第 19 个月的营业额。

答案： $\alpha=0.3$ 和 $\alpha=0.5$ 预测的个月营业额如下

月份	营业额（万元）	0.3	误差平方	0.5	误差平方
1	295				
2	283	295.0	144.0	295.0	144.0
3	322	291.4	936.4	289.0	1089.0
4	355	300.6	2961.5	305.5	2450.3
5	286	316.9	955.2	330.3	1958.1
6	379	307.6	5093.1	308.1	5023.3
7	381	329.0	2699.4	343.6	1401.6
8	431	344.6	7459.6	362.3	4722.3
9	424	370.5	2857.8	396.6	748.5
10	473	386.6	7468.6	410.3	3928.7

11	470	412.5	3305.6	441.7	803.1
12	481	429.8	2626.2	455.8	633.5
13	449	445.1	15.0	468.4	376.9
14	544	446.3	9547.4	458.7	7274.8
15	601	475.6	15724.5	501.4	9929.4
16	587	513.2	5443.2	551.2	1283.3
17	644	535.4	11803.7	569.1	5611.7
18	660	567.9	8473.4	606.5	2857.5
合计			87514.7		50236.0

通过误差比较说明 $\alpha=0.5$ 更为合适。

第 19 月的营业额为 $F_{19} = 0.5 * 660 + (1 - 0.5) * 606.5 = 633.3$

根据最小二乘法，建立趋势方程如下

$$Y = 239.73 + 21.93 * t$$

预测第 19 月的营业额为 $Y = 239.73 + 21.93 * 19 = 656.38$

系主任：_____

出卷人：_____