广东工业大学华立学院

2015-2016学年第一学期

《spss数据统计与分析》（24学时）

期末考查

班级：

姓名：

学号：

得分：

考查要求：

1. 完整、准确填写封面个人信息，得分处不填。

2. 操作题目将主要的步骤截图，分析题目要按样题分析方法 作答。截取主要的图表，做到分析有理有据。作答时要注明作答题目的题号，不注或注明不清的按未做处理。

3. 按封面、考查要求、考查题目（含答案）的顺序装订并于2016年1月4日或1月5号18:30-19:00上交至图书馆四楼办公室。

**《SPSS数据统计与分析》24学时期末考查题目**

**（满分100分）**

**1、操作题（二选一）**

1.1 医疗机构为了对儿童饮食和健康进行研究，随机抽查了50名儿童进行跟踪调查，对他们

一周平均摄入的脂肪量与体重进行统计，并建立了三个变量“编号”、“平均摄入脂肪量”、“体重”，数据如表1所示：

1） 将表1的数据录入到spss中。

2） 试以体重作为加权变量对数据进行加权。

3） 对体重进行分类，20公斤以上视为超重、18～20公斤认为发育良好，18公斤以下认为发育情况一般。试对各种超重人数进行统计，并保存计数结果。

4） 试将体重数据进行分类，换算为超重、发育良好和与发育一般三种情况，分别用数字1、2、3代替。(20公斤以上视为超重、17～20公斤认为发育良好，17公斤以下认为发育情况一般。

1.2. 某科研机构对小白鼠进行分类实验，测算某一药剂对它们体内蛋白活性的影响，根据统

计数据建立“实验室”、“性别”、“蛋白活性”三个变量，观测数据如下表2所示:

1） 将表2的数据录入到spss中，并用1代表雄性，2代表雌性。

2） 按性别分别输出蛋白活性均值，并将数据按照性别排序。

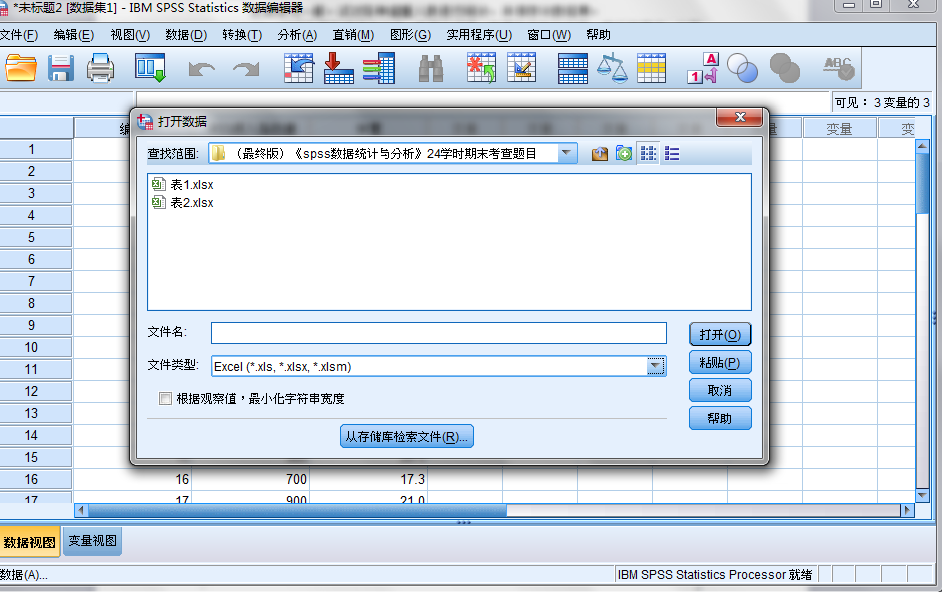
3） 按照蛋白活性的高低对数据进行升序排序。

4） 试将蛋白活性进行分类，换算为强，中，弱三种情况，分别用数字1、2、3代替。（127以上视为“强”、116-127视为“中”，小于116以下视为“弱”）

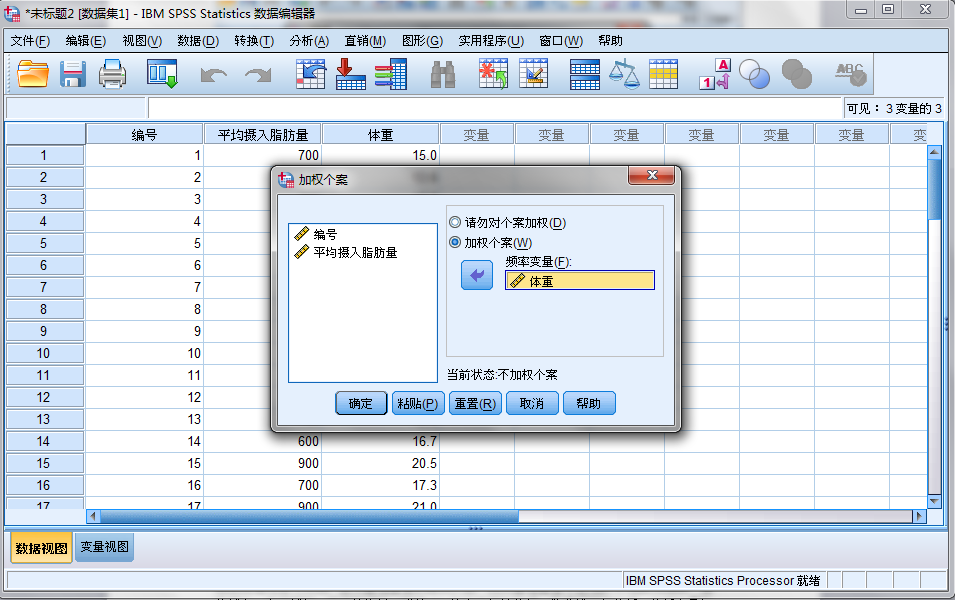
**操作题答题区域**

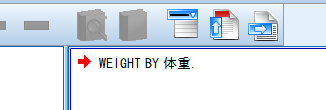
**1.1**

1. 将表1的数据录入到spss中。

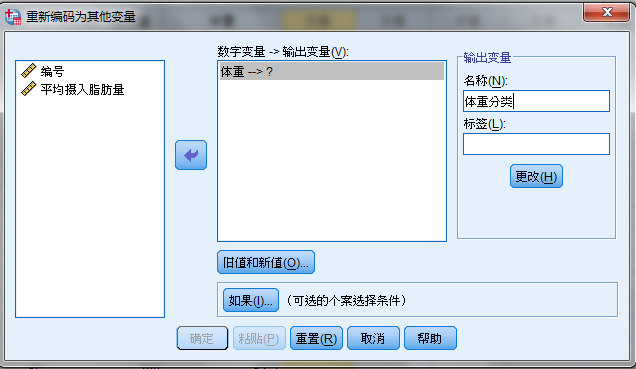


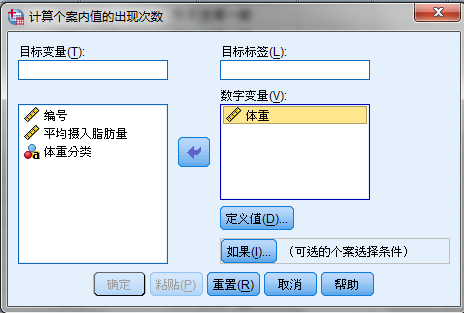
2） 试以体重作为加权变量对数据进行加权。



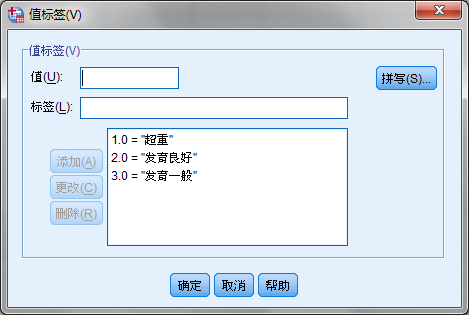
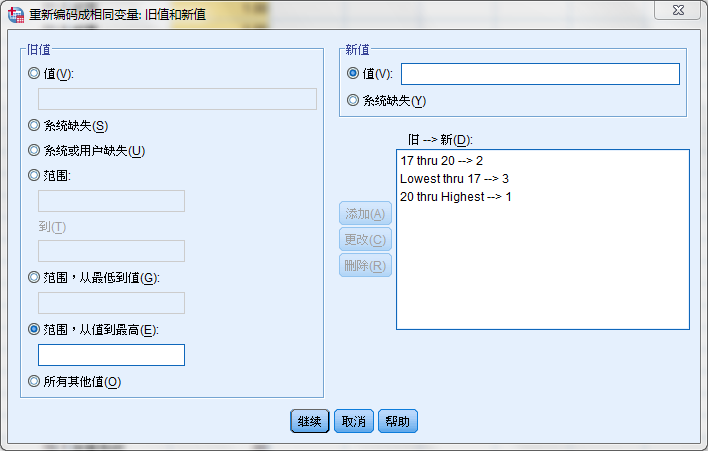
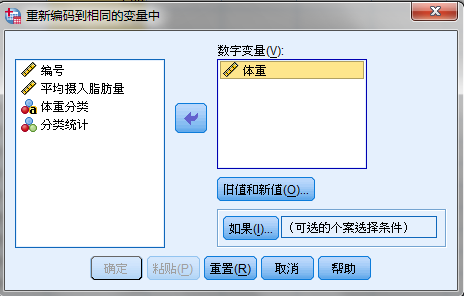


1. 对体重进行分类，20公斤以上视为超重、18～20公斤认为发育良好，18公斤以下认为发育情况一般。试对各种超重人数进行统计，并保存计数结果。





4） 试将体重数据进行分类，换算为超重、发育良好和与发育一般三种情况，分别用数字1、2、3代替。(20公斤以上视为超重、17～20公斤认为发育良好，17公斤以下认为发育情况一般。



**2分析题（以下四道题，2.3为必答题，其余三题三选一）**

2.1 下面的资料（附件1）给出了山东省的两所中学的初三毕业生的升学情况。试据此对两所学校学生的升学情况进行列联表分析，研究两所学校学生的升学率之间有无明显的差别。资料见附件1。

2.2 试对（附件2）中山东省某学校某班学生的高考数学成绩做独立样本t检验，研究该班不同性别的学生之间成绩有无明显的差别。

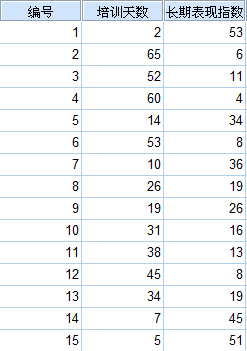
2.3 某著名总裁培训班的讲师想建立一个回归模型，对参与培训的企业高管毕业后的长期表现情况进行预测。自变量是高管的培训天数，因变量是高管毕业后的长期表现指数，指数越大，表现越好。下面的资料（附件3）给出了相关数据，做曲线回归分析，分析出最好的曲线回归模型。

2.4下面的资料（附件4）给出了某调查问卷的测量数据。该调查问卷共有10道题目，均为10分量表，高分代表同意题目代表的观点，共测量了100人。试考察此问卷的信度。

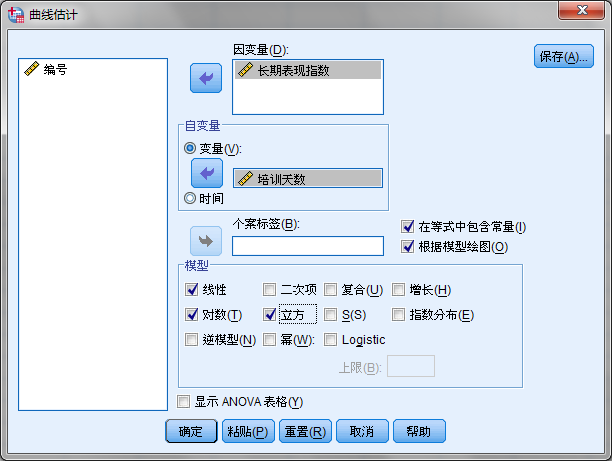
**分析题答题区域：**

**2.3**

（1) 录入数据



1. 分析过程



根据上图我们可以看出，在操作的时候我们选择生成的模型分别是线性、对数以及立方，而对应线图的模型描述自变量为培训天数，方程为线性、对数、三次，相对应1/2/3，因变量为长期变现指数；

|  |  |  |
| --- | --- | --- |
| **模型描述** | | |
| 模型名称 | | MOD\_2 |
| 因变量 | 1 | 长期表现指数 |
| 方程 | 1 | 线性 |
| 2 | 对数 |
| 3 | 三次 |
| 自变量 | | 培训天数 |
| 常数 | | 包含 |
| 其值在图中标记为观测值的变量 | | 未指定 |
| 用于在方程中输入项的容差 | | .0001 |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **模型汇总和参数估计值** | | | | | | | | | |
| 因变量: 长期表现指数 | | | | | | | | | |
| 方程 | 模型汇总 | | | | | 参数估计值 | | | |
| R 方 | F | df1 | df2 | Sig. | 常数 | b1 | b2 | b3 |
| 线性 | .883 | 98.449 | 1 | 13 | .000 | 46.211 | -.747 |  |  |
| 对数 | .960 | 311.776 | 1 | 13 | .000 | 71.754 | -15.815 |  |  |
| 三次 | .985 | 234.610 | 3 | 11 | .000 | 58.876 | -2.356 | .040 | .000 |
| 自变量为 培训天数。 | | | | | | | | | |

另外，从“模型的汇总和参数估计值”的表中可以看出模型的汇总分R方、F、df1、df2、sig.，其中三次的R方最大说明其模型的表现比较明显。

最后我们从下图的模型表现中可以看出三条曲线的拟合情况，图中的圆圈的则是模型的实际值，不难发现“三次”曲线的模型拟合实际值的效果是最好的。

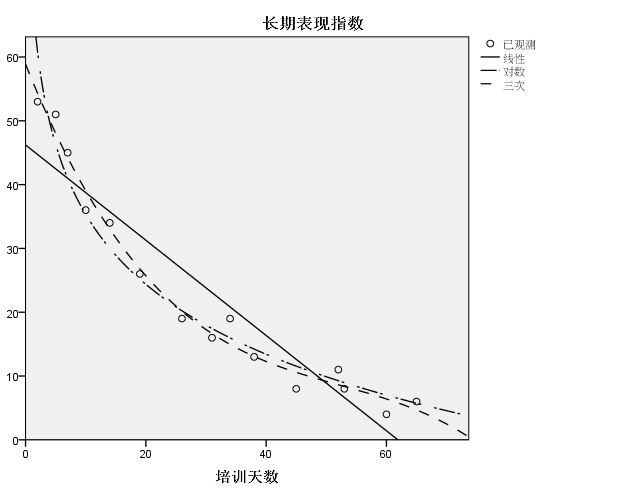


图 拟合曲线

（3）分析结果

从模型的汇总我们可以知道“三次”曲线模型的R方最高，所以“三次”模型对数据的解释能力最强，且模型也是很显著的。

从三条曲线的模型图中所示，发现“三次”曲线的拟合效果是最好的，所以，最终我们选择了“三次”曲线模型。

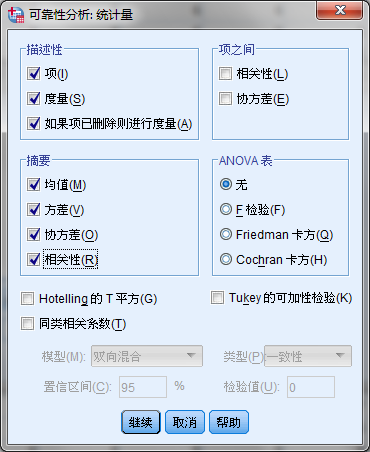
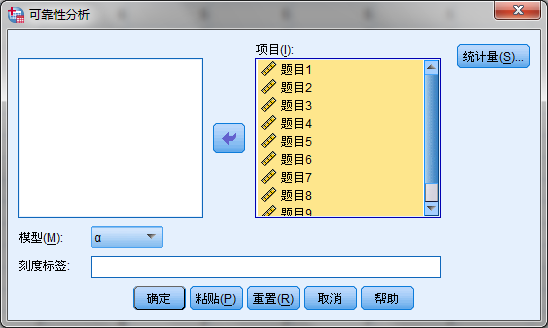
从模型的描述我们可以清楚各变量的系数值，而很明显模型的自变量的显著性是根据各个自变量系数后面的Sig值判断的，如果小于0.05可以说在95%的显著性水平下显著，小于0.01就可以说在99%的显著性水平下显著了，模型显著性小于0.01，为显著，拟合优度很多，可决系数为0.985。

Y（长期表现指数）=58.876—2.356（培训天数）—0.040

**2.4**

（1）导入数据

（2）分析过程



从下表中我们可以Cronbach's Alpha系数作为判断问卷可靠性的重要指标，系数如果在0.5以下说明信度不高，而在0.8-0.9代表信度不错，我们可以看出系数为0.879，信度不错。

|  |  |  |
| --- | --- | --- |
| **可靠性统计量** | | |
| Cronbach's Alpha | 基于标准化项的 Cronbachs Alpha | 项数 |
| .879 | .917 | 10 |

|  |  |  |  |
| --- | --- | --- | --- |
| **项统计量** | | | |
|  | 均值 | 标准偏差 | N |
| 题目1 | 5.41 | 1.724 | 100 |
| 题目2 | 1.19 | .615 | 100 |
| 题目3 | 6.27 | .983 | 100 |
| 题目4 | 6.29 | .967 | 100 |
| 题目5 | 6.28 | .975 | 100 |
| 题目6 | 5.62 | 1.656 | 100 |
| 题目7 | 6.28 | .975 | 100 |
| 题目8 | 6.28 | .975 | 100 |
| 题目9 | 6.27 | .983 | 100 |
| 题目10 | 6.26 | 1.001 | 100 |

从“摘要项统计量”表中我们可以看到问卷数据的个题目之间的均数、极小值、极大值、项数、范围以及方差等统计量。经过分析后我们可以获悉：问卷的极大值为6.290，极小值为1.190，跨度比较大超过了5，而且，项方差范围为2.594，大于2，这说明题目项之间差距比较大。

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **摘要项统计量** | | | | | | | |
|  | 均值 | 极小值 | 极大值 | 范围 | 极大值/极小值 | 方差 | 项数 |
| 项的均值 | 5.615 | 1.190 | 6.290 | 5.100 | 5.286 | 2.520 | 10 |
| 项方差 | 1.282 | .378 | 2.972 | 2.594 | 7.868 | .726 | 10 |
| 项之间的协方差 | .538 | -.291 | 2.602 | 2.893 | -8.950 | .285 | 10 |
| 项之间的相关性 | .525 | -.286 | 1.000 | 1.286 | -3.501 | .204 | 10 |

从下表中给出了如果将相应的变量即题目删除，着则问卷的信服将会发生改变的统计量。一次为总分的均值改变、方差改变、该题与总分的相关系数和Alpha系数的改变情况，而最重要的是后两项，如果相关系数太低，则说明该题的应答分值与总分的高低之间的关系不强，可以删除，如题目中的1、2、6，将这三题删除信度将会提高。

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **项总计统计量** | | | | | |
|  | 项已删除的刻度均值 | 项已删除的刻度方差 | 校正的项总计相关性 | 多相关性的平方 | 项已删除的 Cronbach's Alpha 值 |
| 题目1 | 50.74 | 50.740 | .307 | . | .907 |
| 题目2 | 54.96 | 59.372 | .159 | . | .889 |
| 题目3 | 49.88 | 48.329 | .875 | . | .849 |
| 题目4 | 49.86 | 48.687 | .862 | . | .850 |
| 题目5 | 49.87 | 48.296 | .886 | . | .849 |
| 题目6 | 50.53 | 54.070 | .183 | . | .915 |
| 题目7 | 49.87 | 48.296 | .886 | . | .849 |
| 题目8 | 49.87 | 48.296 | .886 | . | .849 |
| 题目9 | 49.88 | 48.329 | .875 | . | .849 |
| 题目10 | 49.89 | 48.483 | .845 | . | .851 |

（3）分析结果

1. 我们的Cronbach's Alpha系数为0.879，说明问卷的信度比较好，大部分变量都是比较可靠、科学。
2. 通过对项之间的极大值、极小值以及方差范围的比较我们可以知道我们各题目间得分差距比较大。
3. 题目1、2、6应答分值与总分的高低相关性不强，如果我们将之删除，问卷的信度将会提高。