第 13 章

1：

a、　错，原因：logistic 回归用于描述范畴型响应变量与分类预测变量集之间的关系

logistic 回归是一种描述范畴型响应变量与预测变量之间关系的方法,考虑在给定 $X=x$ 情况下 Y 的条件均值为 $E(Y|x)$,表示在给定预测变量值的情况下,期望的响应变量值.在线性回归中,响应变量被认为是一个定义为 $Y=\beta+\beta_1 x+\varepsilon$的随机变量.当误差项$\varepsilon$均值为 0 时,利用线性回归,得到 $E(Y|x)=\beta_0+\beta_1 x$,其可能的取值包含整个实数域.将条件均值 $E(Y|x)$定义为$\pi(x)$,logistic 回归的条件均值具有与线性回归不同的形式..等式中的曲线形式被称为 sigmoidal,因为其形式为 S 型,而且是非线性的.

$$\pi(x)=\frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}}$$

b、对

c、对

d、错

逻辑回归和线性回归首先都是广义的线性回归。(2)经典线性模型的优化目标函数是最小二乘，而逻辑回归则是似然函数。(3)线性回归在整个实数域范围内进行预测，敏感度一致，而分类范围，需要在[0,1]。逻辑回归就是一种减小预测范围，将预测值限定为[0,1]间的一

种回归模型，因而对于这类问题来说，逻辑回归的鲁棒性比线性回归的要好。或者说，线性回归模型无法做到 sigmoid 的非线性形式，sigmoid 可以轻松处理 0/1 分类问题。

线性回归模型的正态性指的是模型的残差服从均值为 0 方差为 σ^2（标准化残差服从均数为 0，方差为 1）的正态分布。

当自变量为分类变量、因变量为连续变量时，也是可以采用线性回归的。只是在更多的时候，这种类型的分析我们更关注的是组间差异比较而不是线性回归预测，通常采用方差分析或者 t 检验，尤其是自变量只有 1 个对的时候。模型假定不同的组来自同一个总体中的抽样，各组（严格说应该是各个单元格）的残差服从同一个正态分布，不同组的残差均服从同一个均数为 0 标准差为 σ2 的正态分布。在实际考察的时候我们往往直接考察固定的自变量值（不同的组）对应的因变量值是否呈正态分布。

e、错

线性回归最具有吸引力的特征之一是能够获得回归系数最优值的封闭形式解集，这也是最小二乘法的优点。遗憾的是，在估计 logistic 回归时，这样的封闭形式解并不存在。为此，必须利用最大似然估计方法，通过该方法获得参数估计，以使观察到的观察数据的似然性最大化。

f、对

g、对

h、对

i、对

j、对

k、对

2:

$$g(x)=In\left[\frac{\lambda(x)}{1-\lambda(x)}\right]=In\left(e^{\beta_0+\beta_1 x}\right)$$

3:

当记录 x=1 时，响应变量发生的概率与当记录在 x=0 时响应发生的概率比的比值.

:Odd 是概率比,定义为一个事件发生的概率与该事件不发生的概率的比值.

OddRatio 是概率比比率.定义为当记录在 *x*=1 时,响应变量发生的概率比与当记录在 *x*=0 时响应变量发生的概率比的比值.相对风险定义为 *x*=1 时响应发生的概率与 *x*=0 时响应发生的概率的比值,.OddRatio 有时被用于估计相对风险,若[1-*π*(0)]/[1-*π*(1)]≈1,则 OR 能准确地估计相对风险.

优势比(oddsratio；OR)是另外一种描述概率的方式。优势比将会告诉我们某种推测的概率比其反向推测的概率大多少。换句话说，优势比是指某种推测为真的概率与某种推测为假的概率的比值。

$$\text{Odds ratio}=\text{OR}=\frac{\pi(1)/[1-\pi(1)]}{\pi(0)/[1-\pi(0)]}=\frac{e^{\beta_0+\beta_1}}{e^{\beta_0}}=e^{\beta_1}$$

7：

针对 100(1-a)%置信区间的概率比(OR)可如下获得:

$$\exp\left[b_1 \pm z \cdot \mathrm{SE}\left(b_1\right)\right]$$

8：

非显著性变量。

措施：分析人员可以注释估计，注明哪些变量可以用于估计，一般来说在估计时应该使用显著性变量。

“显著性检验”实际上是英文 significancetest 的汉语译名。在统计学中，显著性检验是“统计假设检验”（Statisticalhypothesistesting）的一种，显著性检验是用于检测科学实验中实验组与对照组之间是否有差异以及差异是否显著的办法。实际上，了解显著性检验的“宗门背景”（统计假设检验）更有助于一个科研新手理解显著性检验。“统计假设检验”这一正名实际上指出了“显著性检验”的前提条件是“统计假设”，换言之“无假设，不检验”。任何人在使用显著性检验之前必须在心里明白自己的科研假设是什么，否则显著性检验就是“水中月，镜中花”，可望而不可即。用更通俗的话来说就是要先对科研数据做一个假设，然后用检验来检查假设对不对。一般而言，把要检验的假设称之为原假设，记为 H0；把与 H0 相对应（相反）的假设称之为备择假设，记为 H1。

如果原假设为真，而检验的结论却劝你放弃原假设。此时，我们把这种错误称之为第一类错误。通常把第一类错误出现的概率记为 α

如果原假设不真，而检验的结论却劝你不放弃原假设。此时，我们把这种错误称之为第二类错误。通常把第二类错误出现的概率记为β

通常只限定犯第一类错误的最大概率 α，不考虑犯第二类错误的概率 β。我们把这样的假设检验称为显著性检验，概率 α 称为显著性水平。

9：

通过统计推理可以将海量样本信息中的显著性变量得到，降低数据处理的复杂度，提升估计显著性检验可以分为参数检验和非参数检验。参数检验要求样本来源于正态总体（服从正态分布），且这些正态总体拥有相同的方差，在这样的基本假定（正态性假定和方差齐性假定）下检验各总体均值是否相等，属于参数检验。

当数据不满足正态性和方差齐性假定时，参数检验可能会给出错误的答案，此时应采用基于秩的非参数检验。

参数检验的方法及其相应知识点的解释（这里只给出参数检验中常见的方差分析）：

方差分析主要分为'①单因素一元方差分析'；'②双因素一元方差分析'；'③多因素一元方差分析'；'④单因素多元方差分析'。下面一节对各种方差分析的实现方法进行介绍。但在介绍之前，我要首先“剧透”一下两个重要的点，理解这些点有助于区别不同类型的方差分析。的准确性。

12：

这些变量可能具有显著性，对于估计的结果的适用性有较大的影响。

显著性，又称统计显著性（Statisticalsignificance），是指零假设为真的情况下拒绝零假设所要承担的风险水平，又叫概率水平，或者显著水平。