2015 南京理工大学大学生数学建模竞赛

承　诺　书

我们仔细阅读了中国大学生数学建模竞赛的竞赛规则.

我们完全明白，在竞赛开始后参赛队员不能以任何方式（包括电话、电子邮件、网上咨询等）与队外的任何人（包括指导教师）研究、讨论与赛题有关的问题。

我们知道，抄袭别人的成果是违反竞赛规则的, 如果引用别人的成果或其他公开的资料（包括网上查到的资料），必须按照规定的参考文献的表述方式在正文引用处和参考文献中明确列出。

我们郑重承诺，严格遵守竞赛规则，以保证竞赛的公正、公平性。如有违反竞赛规则的行为，我们将受到严肃处理。

我们授权全国大学生数学建模竞赛组委会，可将我们的论文以任何形式进行公开展示（包括进行网上公示，在书籍、期刊和其他媒体进行正式或非正式发表等)。

我们参赛选择的题号是（从 A/B 中选择一项填写）： A

我们的参赛报名号为（报名编号）： 36

所属学院（请填写完整的全名）： 自动化学院

参赛队员 (打印并签名) ：1. 刘鹏

2. 季蔡娟

3. 贺青

日期： 2016 年 05 月 25 日

评阅编号（由组委会评阅前进行编号）：

信息系统安全程度评价

摘要

处在一个互联网信息化时代，历年来，威胁信息系统安全的事件层出不穷，保障网络信息安全的挑战愈发严峻。如何达到对信息系统安全程度的准确度量是我们现在所面临的一个重要难题。本文针对历年来发生的各种信息系统安全事件，综合分析了不同信息系统安全事件的起因、发生频率和覆盖范围，确定了影响信息系统安全的各项因素指标。在定性建立信息系统安全等级的基础上，通过建立一个信息系统安全程度数学模型——多元线性回归模型，采用逐步回归分析法，结合 SPSS 软件，对影响信息系统安全的各项有利和不利因素进行分析，确定了一个能够定性且初步定量评判信息系统安全程度的标准，并在其中选取了一个重要的信息安全度量指标，最后用实例加以说明。

对问题一，已知影响信息系统安全程度主要从系统本身对外界网络陷阱的自身免疫能力和外界对信息系统产生干扰的不良因素破坏能力两个角度进行分析，进而得到衡量信息系统安全程度的十一个指标。选取多组样本数据，通过 SPSS 软件，建立多元线性回归模型，获得信息系统安全程度与十一个指标之间的关系。为获得这个关系，则分别从系统本身对外界网络陷阱的自身免疫能力和与之相关的四个指标之间的关系，以及外界对信息系统产生干扰的不良因素破坏能力和与之相关的七个指标之间的关系，最终获得所需的信息系统在有利和不利指标相互影响下的安全等级。

对问题二，为选取一个重要的信息系统安全度量指标，建立多元线性回归模型和一元线性回归模型，采用逐步回归的分析方法，利用 SPSS 软件从问题一所确定的影响信息系统安全的十一个指标中选出对信息系统安全程度影响最大的指标。最终得出信息系统自身的网络漏洞扫描系统的免疫能力是一个重要的信息系统安全度量指标。

对问题三，利用上述两题建立的数学模型，具体对小组成员所持有的 HTC、HUAWEI；两种手机信息系统的安全程度进行分析比较，验证两组数学模型建立的准确性。

关键词 信息系统安全程度、指标、多元线性回归模型、逐步回归分析

问题一

1. 问题分析

根据国家互联网应急中心公布的“2015年我国互联网网络安全势态综述”统计2015年我国木马和僵尸网络、个人信息泄露、移动互联网恶意程序、拒绝服务攻击、安全漏洞、网页仿冒、网页篡改等发生的频率、事件造成的破坏力、系统自身的抵抗力和系统自身的恢复力，作为衡量信息系统安全程度指标的侧面数据。我们在建立信息系统安全指标衡量体系的基础上, 构造信息系统综合安全程度评价。

1.1 信息系统安全衡量指标体系

实际的信息系统安全衡量指标主要包括系统本身对外界网络陷阱的自身免疫能力和外界对信息系统产生干扰的不良因素破坏能力两个角度。

1.1.1 从信息系统的外界对信息系统产生干扰的不良因素破坏能力的角度

(1)自然及不可抗拒因素

指地震、火灾、水灾、风暴以及社会暴力或战争等。

(2)硬件及物理因素

指系统硬件及环境的安全可靠，包括机房设施、计算机主体、存储系统、辅助设备、数据通讯设施以及信息存储介质的安全性。

(3)电磁波因素

计算机系统及其控制的信息和数据传输通道，在工作过程中都会产生电磁波辐射，在一定地理范围内用无线电接收机很容易检测并接收到，有可能造成信息通过电磁辐射而泄漏。另外，空间电磁波也可能对系统产生电磁干扰，影响系统正常运行。

(4)软件因素

软件的非法删改、复制与窃取将使系统的软件受到损失，并可能造成泄密。计算机网络病毒也是以软件为手段侵入系统进行破坏的。

(5)数据因素

指数据信息在存储和传递过程中的安全性，这是计算机犯罪的主攻核心，是必须加以安全和保密的重点。

(6)人为及管理因素

涉及到工作人员的素质、责任心、以及严密的行政管理制度和法律法规，以防范人为的主动因素直接对系统安全所造成的威胁。

(7)其他因素

指系统安全一旦出现问题，能将损失降到最小，把产生的影响限制在许可的范围内，保证迅速有效地恢复系统运行的一切因素。

1.1.2 从信息系统的系统本身对外界网络陷阱的自身免疫能力的角度

(1)防火墙

防止内部的信息系统遭受外部的攻击，它划定一个边界，使经认证的局域网内用户可以安全访问外部网络或者安全地接受外部网络的范围，拒绝内外网之间的非授权访问。

(2)网络入侵检测系统

侧重于基于网络的监测的一种安全技术，能发现和阻止来自于内部“授权”用户或远程用户穿透边界防护进行的攻击行为，是对防火墙技术的有力补充。

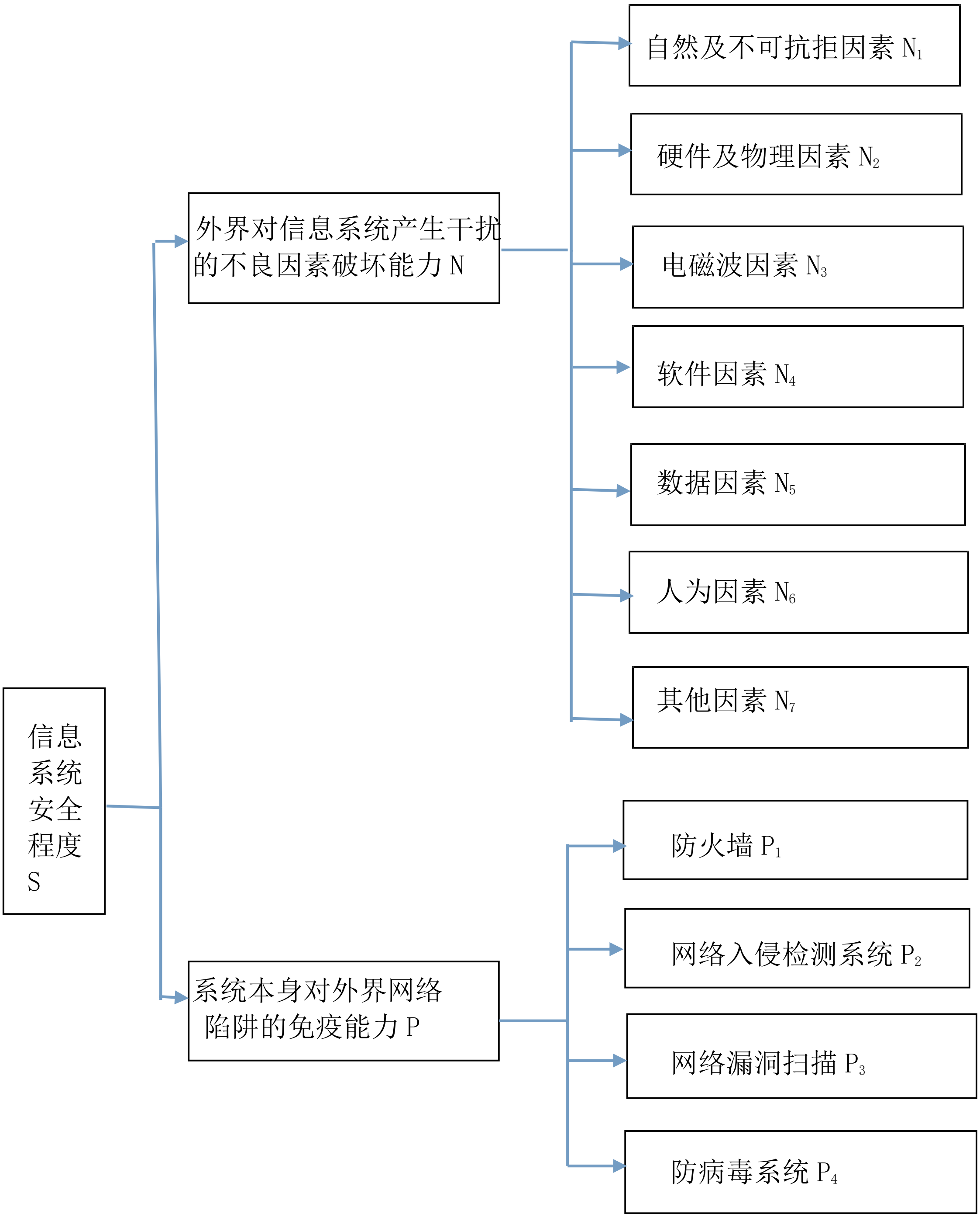
(3) 网络漏洞扫描

漏洞分析工具可以帮助自动识别网络或系统中的漏洞，以使用户赶在攻击者之前发现这些漏洞或网络中的不合理配置。它为系统加固指明了方向和目标。

(4) 防病毒系统

恶意代码和病毒能以多种方式攻击局域网计算机，修改传输的数据、插入恶意代码、攻击协议或配置的漏洞，防病毒系统能防止和检测恶意代码及其攻击。在网关处保护网络，是阻止病毒在网络中传播并影响更多工作站的有效手段。

从上述二个角度考虑的指标共有 11 个，组成评价指标体系如下图所示。



1.2 各项指标的侧面数据以及计算方式

1.2.1 衡量指标值的标准化处理

从图 1 所示的评价指标体系来看,绝大部分的指标虽然可以用为定性指标,难于用数量表示。因此,在实际评价工作中,对图 1 中所示的所有指标,首先由专家先做定性评价,然后设法予以量化。设专家对定性的指标 P_j、N_i、F_k 的评语为好、较好、一般、较差、差,与评语相对应的量化了的数值见表 1。

表 1　定性指标的量化

评语	好	较好	一般	较差	差
等级	[5,4)	[4,3)	[3,2)	[2,1)	[1,0)

1.2.2 从信息系统的外界对信息系统产生干扰的不良因素破坏能力的角度

(1)发生频率 F_1

指预计该指标所代表的事件在的未来一年内发生的频率即发生次数占总次数比例。

(2)破坏力等级 F_2

指该指标所代表的事件在过去的一年所造成的信息、数据及财产损失。

1.2.3 从信息系统的系统本身对外界网络陷阱的自身免疫能力指标的角度

(1)抵抗力等级 F_3

抵抗能力是指信息系统能够应对威胁的能力所构成的系统的安全保护能力之一(通过对破坏的抵抗与修复能力进行等级量化)。

(2)抵抗力比重 F_4

整个防御系统该防御方式所占比重(可通过投入资金、人力量化)。

由上述侧面数据计算可得:

$$P_j=F_3*F_4 \qquad (1)$$

$$N_i=F1*F2 \qquad (2)$$

$$S=N-P \qquad (3)$$

2. 模型假设

1. 在试验中，各个信息系统除自身信息结构网络不同（每个系统均带有自身的防火墙、网络入侵检测系统、网络漏洞扫描、防病毒系统等）以外，均安装用户体验软件（比如：QQ、微信、相同的浏览器等）。

2. 在实际问题中，信息系统还受系统使用周期、新旧程度、用户日常维护等的影响，但在该模型中忽略以上各种因素的影响，仅仅考虑系统本身对外界网络陷阱的自身免疫能力和外界对信息系统产生干扰的不良因素的影响

3. 忽略不用用户试验者使用用户体验软件的不同习惯。

3. 符号定义

S： 信息系统的安全程度

P： 系统本身对外界网络陷阱的自身免疫能力

N： 外界对信息系统产生干扰的不良因素的破坏能力

F_1　发生频率

F_2: 破坏力等级

F_3: 抵抗力等级

F_4: 抵抗力比重

4. 模型建立

4.1 多元线性回归模型的建立

假设某一因变量 y 受 k 个自变量 x_1、x_2……x_k 的影响，其 n 组观测值为（$y_a, x_{1a}, x_{2a}, \ldots, x_{ka}$），a=1，2……n。那么，多元线性回归模型的结构形式为：

$$y_a = \beta_0 + \beta_1 x_{1a} + \beta_2 x_{2a} + \ldots + \beta_k x_{ka} + \varepsilon_a \quad (4.1.1)$$

式中：$\beta_0, \beta_1, \ldots, \beta_k$ 为待定参数；ε_a 为随机变量。

如果 $b_0, b_1, \ldots, b_k$ 分别为 $\beta_0, \beta_1, \beta_2, \ldots, \beta_k$ 的拟合值，则回归方程为

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_k x_k \quad (4.1.2)$$

式中：b_0 为常数；$b_1, b_2, \ldots, b_k$ 称为偏回归系数。

偏回归系数 b_i（$i = 1,2,\ldots,k$）的意义是，当其他自变量 x_j（$j \neq i$）都固定时，自变量 x_i 每变化一个单位而使因变量 y 平均改变的数值。

根据最小二乘法原理，β_i（$i = 0,1,2,\ldots,k$）的估计值 b_i（$i = 0,1,2,\ldots,k$）应该使

$$Q = \sum_{a=1}^{n}\left(y_a - \hat{y}_a\right)^2 = \sum_{a=1}^{n}\left[y_a - \left(b_0 + b_1 x_{1a} + b_2 x_{2a} + \ldots + b_k x_{ka}\right)\right]^2 \to \min$$

（4.1.3）

有求极值的必要条件得

$$\begin{cases} \dfrac{\partial Q}{\partial b_0} = -2\sum_{a=1}^{n}\left(y_a - \hat{y}_a\right) = 0 \\ \dfrac{\partial Q}{\partial b_j} = -2\sum_{a=1}^{n}\left(y_a - \hat{y}_a\right)x_{ja} = 0 (j = 1,2,\ldots,k) \end{cases} \quad (4.1.4)$$

将方程组（3.2.14）式展开整理后得：

$$\begin{cases} nb_0 + (\sum_{a=1}^{n} x_{1a})b_1 + (\sum_{a=1}^{n} x_{2a})b_2 + \ldots + (\sum_{a=1}^{n} x_{ka})b_k = \sum_{a=1}^{n} y_a \\ (\sum_{a=1}^{n} x_{1a})b_0 + (\sum_{a=1}^{n} x_{1a}^2)b_1 + (\sum_{a=1}^{n} x_{1a}x_{2a})b_2 + \ldots + (\sum_{a=1}^{n} x_{1a}x_{ka})b_k = \sum_{a=1}^{n} x_{1a}y_a \\ (\sum_{a=1}^{n} x_{2a})b_0 + (\sum_{a=1}^{n} x_{1a}x_{2a})b_1 + (\sum_{a=1}^{n} x_{2a}^2)b_2 + \ldots + (\sum_{a=1}^{n} x_{2a}x_{ka})b_k = \sum_{a=1}^{n} x_{2a}y_a \\ \ldots \\ (\sum_{a=1}^{n} x_{ka})b_0 + (\sum_{a=1}^{n} x_{1a}x_{ka})b_1 + (\sum_{a=1}^{n} x_{2a}x_{ka})b_2 + \ldots + (\sum_{a=1}^{n} x_{ka}^2)b_k = \sum_{a=1}^{n} x_{ka}y_a \end{cases} \quad (4.1.5)$$

方程组（4.1.5）式，被称为正规方程组。

如果引入一下向量和矩阵：

$$b = \begin{pmatrix} b_0 \\ b_1 \\ b_2 \\ \ldots \\ b_k \end{pmatrix}, Y = \begin{pmatrix} y_1 \\ y_2 \\ \ldots \\ y_n \end{pmatrix}, X = \begin{pmatrix} 1 & x_{11} & x_{21} & \ldots & x_{k1} \\ 1 & x_{12} & x_{22} & \ldots & x_{k2} \\ 1 & x_{13} & x_{23} & \ldots & x_{k3} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & x_{1n} & x_{2n} & \ldots & x_{kn} \end{pmatrix}$$

$$A = X^T X = \begin{pmatrix} 1 & 1 & 1 & \dots & 1 \\ x_{11} & x_{12} & x_{13} & \dots & x_{1n} \\ x_{21} & x_{22} & x_{23} & \dots & x_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ x_{k1} & x_{k2} & x_{k3} & \dots & x_{kn} \end{pmatrix} \begin{pmatrix} 1 & x_{11} & x_{21} & \dots & x_{k1} \\ 1 & x_{12} & x_{22} & \dots & x_{k2} \\ 1 & x_{13} & x_{23} & \dots & x_{k3} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & x_{1n} & x_{2n} & \dots & x_{kn} \end{pmatrix}$$

$$= \begin{pmatrix} n & \sum_{a=1}^{n} x_{1a} & \sum_{a=1}^{n} x_{2a} & \dots & \sum_{a=1}^{n} x_{ka} \\ \sum_{a=1}^{n} x_{1a} & \sum_{a=1}^{n} x_{1a}^2 & \sum_{a=1}^{n} x_{1a}x_{2a} & \dots & \sum_{a=1}^{n} x_{1a}x_{ka} \\ \sum_{a=1}^{n} x_{2a} & \sum_{a=1}^{n} x_{1a}x_{2a} & \sum_{a=1}^{n} x_{2a}^2 & \dots & \sum_{a=1}^{n} x_{2a}x_{ka} \\ \dots & \dots & \dots & \dots & \dots \\ \sum_{a=1}^{n} x_{ka} & \sum_{a=1}^{n} x_{1a}x_{ka} & \sum_{a=1}^{n} x_{2a}x_{ka} & \dots & \sum_{a=1}^{n} x_{ka}^2 \end{pmatrix}$$

$$B = X^T Y = \begin{pmatrix} 1 & 1 & 1 & \dots & 1 \\ x_{11} & x_{12} & x_{13} & \dots & x_{1n} \\ x_{21} & x_{22} & x_{23} & \dots & x_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ x_{k1} & x_{k2} & x_{k3} & \dots & x_{kn} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \dots \\ y_n \end{pmatrix} = \begin{pmatrix} \sum_{a=1}^{n} y_a \\ \sum_{a=1}^{n} x_{1a}y_a \\ \sum_{a=1}^{n} x_{2a}y_a \\ \dots \\ \sum_{a=1}^{n} x_{ka}y_a \end{pmatrix}$$

则正规方程组（4.1.5）式可以进一步写成矩阵形式

$Ab = B$ （4.1.6）

求解（4.1.5）式可得：

$b = A^{-1}B = (X^T X)^{-1} X^T Y$ （4.1.7）

如果引入记号：

$$L_{ij} = L_{ji} = \sum_{a=1}^{n} (x_{ia} - \overline{x_i})(x_{ja} - \overline{x_j})(i, j = 1,2,\dots,k)$$

$$L_{iy} = \sum_{a=1}^{n} (x_{ia} - \overline{x_i})(y_a - \overline{y})(i = 1,2,\dots,k)$$

则正规方程组也可以写成：

$$\begin{cases} L_{11}b_1 + L_{12}b_2 + \dots + L_{1k}b_k = L_{1y} \\ L_{21}b_1 + L_{22}b_2 + \dots + L_{2k}b_k = L_{2y} \\ \qquad \dots\dots\dots\dots \\ L_{k1}b_1 + L_{k2}b_2 + \dots + L_{kk}b_k = L_{ky} \\ b_0 = \overline{y} - b_1\overline{x_1} - b_2\overline{x_2} - \dots - b_k\overline{x_k} \end{cases}$$ （4.1.8）

4.2 多元线性回归模型的显著性检验

与一元线性回归模型一样，当多元线性回归模型建立以后，也需要进行显著性检验。与前面的一元线性回归分析一样，因变量 y 的观测值 $y_1, y_2, \ldots, y_n$ 之间的波动或差异，是由两个因素引起的，一是由于自变量 $x_1, x_2, \ldots, x_k$ 的取之不同，另一是受其他随机因素的影响而引起的。为了从 y 的离差平方和中把它们区分开来，就需要对回归模型进行方差分析，也就是将 y 的离差平方和 S_T 或（L_{yy}）分解成两个部分，即回归平方和 U 与剩余平方和 Q：

$$S_T = L_{yy} = U + Q$$

在多元线性回归分析中，回归平方和表示的是所有 k 个自变量对 y 的变差的总影响，它可以按公式

$$U = \sum_{a=1}^{n} (\hat{y}_a - \bar{y})^2 = \sum_{i=1}^{k} b_i L_{iy}$$

计算，而剩余平方和为

$$Q = \sum_{a=1}^{n} (y_a - \hat{y}_a)^2 = L_{yy} - U$$

以上几个公式与一元线性回归分析中的有关公式完全相似。它们所代表的意义也相似，即回归平方和越大，则剩余平方和 Q 就越小，回归模型的效果就越好。不过，在多元线性回归分析中，各平方和的自由度略有不同，回归平方和 U 的自由度等于自变量的个数 k，而剩余平方和的自由度等于 $n-k-1$，所以 F 统计量为：

$$F = \frac{U/k}{Q/(n-k-1)}$$

当统计量 F 计算出来之后，就可以查 F 分布表对模型进行显著性检验。

5. 模型求解

5.1 实验数据处理

我们获取并采用了我国某银行系统司 2002—2015 年信息系统的安全统计与评估，并根据统计数据以及公式（1）（2）可得 S、N_i、P_j（i=1、2……7;j=1、2、3、4）的数据如下表二：

表二 S、Ni、Pj 等级数据

	2002 年	2003 年	2004 年	2005 年	2006 年	2007 年	2008 年
N_1	0.2	0.9	0.4	1.15	0.24	0.15	3
N_2	1.15	1.2	0.75	1.72	1	0.66	1
N_3	1.04	1.04	0.63	0.52	0.36	0.28	0
N_4	1.84	1.36	1.36	1.12	3.3	1.36	1.52
N_5	1.36	1.9	1.15	1.2	1.24	2.25	1.5
N_6	0.4	0.3	0.32	0.5	0.7	0.9	0.3
N_7	0.18	0.03	0.06	0.05	0.01	0.01	0.01
N	4.3	3.98	3.78	3	4.23	3.67	3.8
P_1	0.9378	1.0784	0.9174	0.875	0.8396	0.7336	1.0176
P_2	0.176	0.1922	0.3186	0.2346	0.3938	0.3504	0.1922

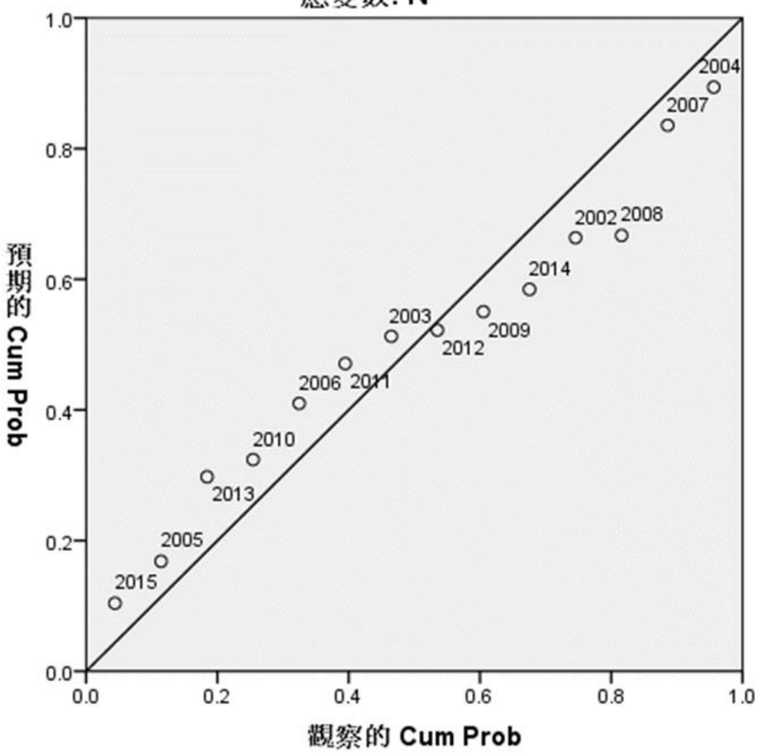
P_3	0.0894	0.0836	0.0656	0.1384	0.3672	0.1292	0.1836
P_4	1.0611	0.8433	0.6328	0.9204	0.687	0.9864	1.1433
P	1.73	1.96	2.11	2.33	2.98	2.96	2.84
S	0.18	0.54	1.46	1.86	0．6	0.78	1．6
	2009 年	2010 年	2011 年	2012 年	2013 年	2014 年	2015 年
N_1	1.5	0	0.6	0	0.6	0.15	0.6
N_2	1	0.75	0.54	0.36	0.75	0.48	0.1
N_3	0.4	0.52	0.66	0	0.4	0	0.4
N_4	1.4	1.12	1.48	1.12	1.65	1.6	1.4
N_5	2.65	1.6	1.44	2.15	2.25	1.68	1.8
N_6	0.55	0.92	1.35	2.15	0.36	0.3	0.39
N_7	0.05	0.01	0.16	0.15	0.01	0.08	0.1
N	4.1	2.89	3.4	2.8	3.76	3.6	3.33
P_1	1.2402	1.0176	1.4368	2.196	1.442	1.7815	1.6825
P_2	0.1734	0.3188	0.6741	0.3078	0.2454	0.4704	0.7392
P_3	0.4808	0.3672	0.3148	0.5052	0.6102	0.4875	0.5862
P_4	0.7785	1.2712	1.0348	0.954	1.234	1.2976	1.1332
P	3.04	3.35	3.34	3.45	3.87	3.99	4.12
S	2.46	3．83	3.11	3.10	3．37	3.75	4.11

5.2 通过 SPSS 软件建立散点图考察与每一个自变量之间的相关关系

通过 SPSS 软件建立散点图考察 N 与每一个自变量 Ni 之间的相关关系，建立多元线性回归模型，并计算回归系数和统计量。

迴歸 標準化殘差 的常態 P-P 圖

應變數: N

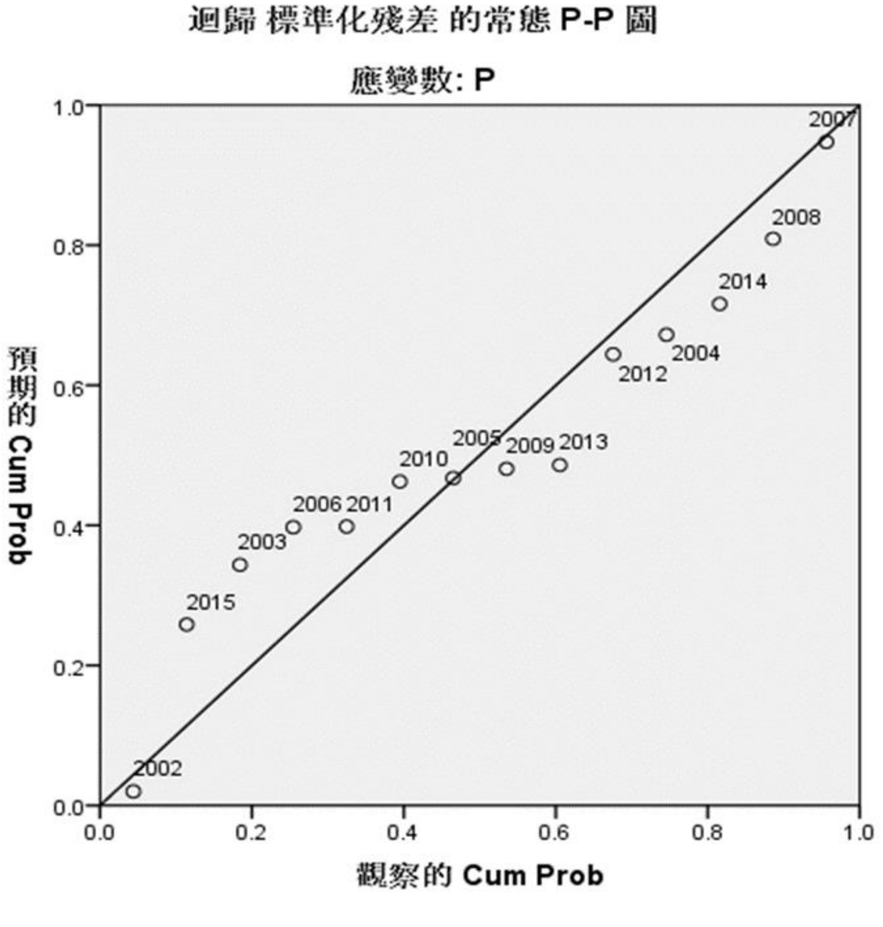


综合上述分析，可以建立如下回归模型：

$$N=\beta_0+\beta_1N_1+\beta_2N_2+\beta_3N_3+\beta_4N_4+\beta_5N_5+\beta_6N_6+\beta_7N_7$$

参数	参数估计值
β_0	1.707
β_1	0.136
β_2	0.028
β_3	0.440
β_4	0.631
β_5	0.452
β_6	-0.419
β_7	1.915
$R^2=0.833$ F=4.277 P=0.048	

5.3 通过 SPSS 软件建立散点图考察 P 与每一个自变量 Pj 之间的相关关系,建立多元线性回归模型，并计算回归系数和统计量

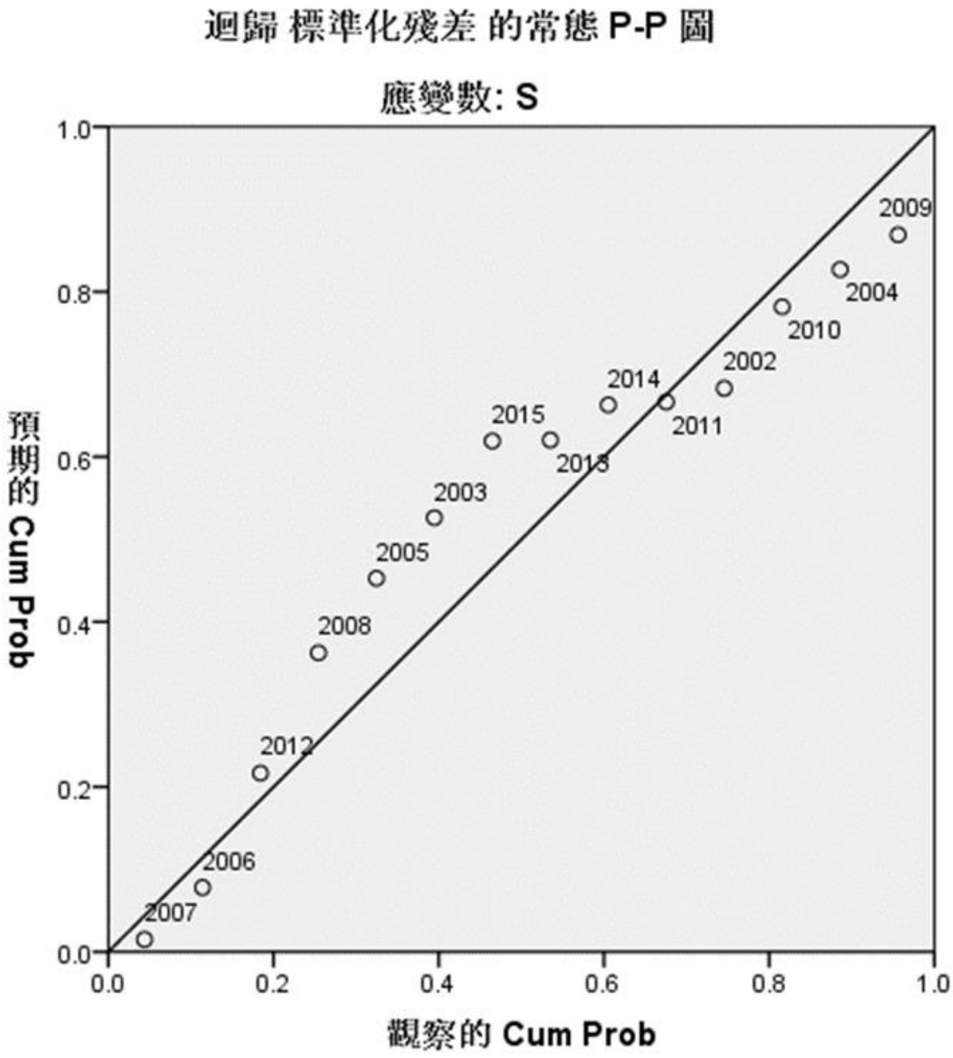


综合上述分析，可以建立如下回归模型：

$$P=\alpha_0+\alpha_1P_1+\alpha_2P_2+\alpha_3P_3+\alpha_4P_4$$

参数	参数估计值
α_0	0.951
α_1	-0.086
α_2	1.188
α_3	2.722
α_4	0.892
R^2=0.916 F=24.660 P=0.000	

5.4 通过 SPSS 软件建立散点图考察 S 与每一个自变量 P、N 之间的相关关系,建立多元线性回归模型，并计算回归系数和统计量



综合上述分析，可以建立如下回归模型：

$$S= r_0+r_1P+r_2N$$

参数	参数估计值
r_0	1.838
r_1	-0.973
r_2	1.293
$R^2=0.833$ F=27.452 P=0.000	

则 S 与 N、P 的线性回归模型为：

$S=1.838+1.293P-0.973N=1.838+1.293*(0.951-0.086P_1+1.188P_2+2.722P_3+1.707P_4)-0.973*(1.707+0.136N_1-0.028N_2+0.440N_3-0.631N_4+0.452N_5-0.419N_6+1.915N_7)$

问题二

1. 问题分析

信息系统的安全程度是指在外界的不利因素和系统自身的免疫能力综合影响下的信息系统的安全状态。每个因素对于同一信息系统安全的影响所占权重不同，即加入某一因素后对信息系统的安全程度的变化显著性是有区别的。本题可逐个引入不同的因素，每引入一个因素后，对信息系统的安全程度进行检验，并对之前引入的每个因素逐个进行检验，当原来引入的因素因为后面因素的引入变得不再显著时，则将其剔除，以确保每次引入新的因素之前，影响信息系统安全的因素只包含显著性变化的。这是一个反复的过程，直到最后得出一个影响信息系统安全的最具显著性的因素。

2. 模型假设

(1)假设所有观测数据均准确可靠，可以根据系统受入侵事件影响范围和发生频率确定各因素对信息系统安全程度的影响。

(2)假设信息系统安全程度只与所设定的七种不利因素和四种有利因素有关，不考虑其他因素对信息系统安全的影响。

(3)假设问题中所有信息系统自身防护功能都是完善的，总是按照对自身最有利的方式运行。对各种因素和信息系统安全程度模拟多元线性回归分析，剔除掉相关性较小的影响因素。

3. 符号定义

S——信息系统的安全程度

N——信息系统安全的不利影响因素破坏能力

N_1——自然及不可抗拒因素破坏能力

N_2——硬件及物理因素破坏能力

N_3——电磁波因素破坏能力

N_4——软件因素破坏能力

N_5——数据因素破坏能力

N_6——人为及管理因素破坏能力

N_7——其他因素破坏能力

P——信息系统安全的有利影响因素免疫能力

P_1——防火墙免疫能力

P_2——网络入侵检测系统免疫能力

P_3——网络漏洞扫描系统免疫能力

P_4——防病毒系统免疫能力

4.模型建立与求解

根据问题一的多元回归方程定义，设安全程度待评价的信息系统有 m 个，记为集合 $S=\{S_1, S_2, S_3, \cdots\cdots, S_m\}$，拟定的不利影响因素集合为 $U=\{N_1, N_2, \cdots\cdots, N_7\}$，拟定的有利影响因素集合为 $V=\{P_1, P_2, P_3, P_4\}$。

4.1 设定变量

设信息系统的安全程度 s=y 为因变量，x_1，x_2，x_3，...，x_{11}是所有自变量，即影响信息系统安全的各项因素指标。y_i，x_{i1}，x_{i2}，...，x_{i11}（$i=1,2,3,\cdots,n$）是独立抽取的 n 组样本。设自变量被选进模型的显著性水平为$\alpha_1=0.05$，自变量被剔除出模型的显著性水平α_2，且$0<\alpha_1 \leq \alpha_2<1$。应变量 1 个，自变量 m 个，变量共 m+1 个；样本含量 n 个。最优指标的选择即最优回归方程的选择。最优回归方程的定义为：方程中包含对所有因变量影响显著的变量；方程中所包含的自变量要尽可能地少。故，建立模型中首先在影响信息系统安全度量的两大指标（有利因素 P 和不利因素 N）构成的多元回归方程中确定影响较显著的指标，再在已确定的指标中选择一个影响指标因素最显著的指标因素。

4.2 利用 SPSS 对因素指标值 N 和 P 逐步回归

选取 n 组样本数据，利用 SPSS 软件对两个主要因素指标值进行逐步回归。

4.2.1SPSS 数据处理结果

已输入/除去变量 [a]

模型	已输入变量	已除去变量	方法
1	P	.	步进（准则：F-to-enter 的概率 <= .050，F-to-remove 的概率 >= .100）。
2	N	.	步进（准则：F-to-enter 的概率 <= .050，F-to-remove 的概率 >= .100）。

a. 因变量：S

模型摘要 [c]

模型	R	R 平方	调整后的 R 平方	标准估算的错误
1	.858[a]	.736	.714	.7294547
2	.913[b]	.833	.803	.6061876

a. 预测变量：（常量），P

b. 预测变量：（常量），P, N

c. 因变量：S

ANOVA[a]（方差分析表）

模型		平方和	自由度	均方	F	显著性
1	回归	17.832	1	17.832	33.513	.000[b]
	残差	6.385	12	.532		
	总计	24.218	13			
2	回归	20.175	2	10.088	27.452	.000[c]
	残差	4.042	11	.367		
	总计	24.218	13			

a. 因变量：S

b. 预测变量：（常量），P

c. 预测变量：（常量），P, N

系数 [a]

模型		非标准化系数		标准系数	t	显著性
		B	标准错误	贝塔		
1	（常量）	-2.456	.827		-2.970	.012
	P	1.551	.268	.858	5.789	.000
2	（常量）	1.838	1.834		1.002	.338
	P	1.293	.245	.715	5.278	.000
	N	-.973	.385	-.342	-2.525	.028

a. 因变量：S

排除的变量 [a]

模型		输入贝塔	t	显著性	偏相关	共线性统计
						容许
1	N	-.342[b]	-2.525	.028	-.606	.826

a. 因变量：S

b. 模型中的预测变量：（常量），P

残差统计数据 [a]

	最小值	最大值(X)	平均值	标准偏差	数字
预测值	-.108571	3.926133	2.196429	1.2457751	14
残差	-1.3151188	.6797706	.0000000	.5576115	14
标准预测值	-1.850	1.388	.000	1.000	14
标准残差	-2.169	1.121	.000	.920	14

a. 因变量：S

4.2.2 确定决定系数 R^2

P：R^2=SS回/SS总=1-SS残/SS总=0.714

N：R^2=SS回/SS总=1-SS残/SS总=0.803-0.714=0.089

4.2.3 确定偏回归平方和

回归方程中某一自变量 x_j 的偏回归平方和表示模型中含有其他 m-1 个自变量的条件下，该自变量对因变量 Y 的贡献，相当于从回归方程中剔除 x_j 后所引起的回归平方和的减少量，或在 m-1 个自变量的基础上新增加 x_j 引起的回归平方和的增加量。

$$F_j=\frac{SS_{回}(x_j)/1}{SS_{残}/(n-m-1)}$$

SS回（x_j）表示偏回归平方和，其值越大说明自变量越重要。

其中 SS回（P）=17.832>SS回（N）。

可以得到“最优”回归方程如下：

S=1.293P-0.973N+1.838

故综上得，问题一所得多元回归方程中剔除因素指标 N（不利因素），保留因素指标 P（有利因素），即 P 为重要的信息安全度量指标。

4.3 利用 SPSS 对影响因素指标值 P 的四项因素指标值进行逐步回归

选取 n 组样本数据，利用 SPSS 软件对四个因素指标值 P_1，P_2，P_3，P_4 进行逐步回归。

剔除 P_1 得到的多元线性回归方程为：$S_1=1.163P_2+2.602P_3+0.879P_4+0.904$

剔除 P_2 得到的多元线性回归方程为：$S_2=0.025P_1+2.956P_3+0.937P_4+1.101$

剔除 P_3 得到的多元线性回归方程为：$S_3=0.705P_1+1.545P_2+1.287P_4+0.320$

剔除 P_4 得到的多元线性回归方程为：$S_4=-0.027P_1+1.237P_2+2.998P_3+1.664$

4.3.1SPSS 数据处理结果

已输入/除去变量 a

模型	已输入变量	已除去变量	方法
1	P3	.	步进（准则：F-to-enter 的概率 <= .050，F-to-remove 的概率 >= .100）。
2	P2	.	步进（准则：F-to-enter 的概率 <= .050，F-to-remove 的概率 >= .100）。
3	P4	.	步进（准则：F-to-enter 的概率 <= .050，F-to-remove 的概率 >= .100）。

a. 因变量：P

模型摘要 d

模型	R	R 平方	调整后的 R 平方	标准估算的错误
1	.892[a]	.795	.778	.3555911
2	.930[b]	.866	.841	.3009157
3	.957[c]	.915	.890	.2504563

a. 预测变量：（常量），P3

b. 预测变量：（常量），P3, P2

c. 预测变量：（常量），P3, P2, P4

d. 因变量：P

ANOVA[a]

模型		平方和	自由度	均方	F	显著性
1	回归	5.891	1	5.891	46.590	.000[b]
	残差	1.517	12	.126		
	总计	7.408	13			
2	回归	6.412	2	3.206	35.408	.000[c]
	残差	.996	11	.091		
	总计	7.408	13			
3	回归	6.781	3	2.260	36.034	.000[d]
	残差	.627	10	.063		
	总计	7.408	13			

a. 因变量：P

b. 预测变量：（常量），P3

c. 预测变量：（常量），P3, P2

d. 预测变量：（常量），P3, P2, P4

系数 a

模型		非标准化系数		标准系数	t	显著性
		B	标准错误	贝塔		
1	（常量）	1.927	.183		10.505	.000
	P3	3.401	.498	.892	6.826	.000
2	（常量）	1.647	.194		8.469	.000
	P3	2.959	.460	.776	6.429	.000
	P2	1.229	.512	.290	2.399	.035
3	（常量）	.904	.346		2.611	.026
	P3	2.602	.410	.682	6.339	.000
	P2	1.163	.427	.274	2.723	.021
	P4	.879	.362	.245	2.425	.036

a. 因变量：P

排除的变量 [a]

模型		输入贝塔	t	显著性	偏相关	共线性统计
						容许
1	P1	.052[b]	.262	.798	.079	.471
	P2	.290[b]	2.399	.035	.586	.839
	P4	.262[b]	2.069	.063	.529	.834
2	P1	-.015[c]	-.088	.932	-.028	.457
	P4	.245[c]	2.425	.036	.608	.830
3	P1	-.048[d]	-.338	.743	-.112	.452

a. 因变量：P

b. 模型中的预测变量：（常量），P3

c. 模型中的预测变量：（常量），P3, P2

d. 模型中的预测变量：（常量），P3, P2, P4

残差统计数据 [a]

	最小值	最大值(X)	平均值	标准偏差	数字
预测值	2.001471	4.284769	2.998571	.7222341	14
残差	-.5439714	.4453420	.0000000	.2196647	14
标准预测值	-1.381	1.781	.000	1.000	14
标准残差	-2.172	1.778	.000	.877	14

a. 因变量：P

4.3.2 确定决定系数

$R^2(P_3) > R^2(P_2) > R^2(P_4)$

4.3.3 确定偏回归平方和

$SS_{回}(P_3) > SS_{回}(P_2) > SS_{回}(P_4)$

故综上得，得到“最优”回归方程如下：

$P=2.602P_3+1.163P_2+0.879P_4+0.904$

选取自变量 P_3 为影响因变量 P 最显著的因素。

4.4 问题总结

根据对影响信息系统安全度量的因素的数学模型建立，对其进行逐层与逐步回归，可以得出，在有利因素和不利因素的共同影响下，对信息系统安全影响最显著的是有利因素，在有利因素中影响显著性最强的为网络漏洞扫描系统的免疫能力。故选取有利因素中信息系统自身的网络漏洞扫描系统的免疫能力作为一个重要的信息系统安全的度量指标。

问题三

1. 问题分析

随着智能时代的到来以及近几年来以及智能手机的迅速发展，智能手机的安全性已成为不可忽视的重要问题，如今智能手机已经涉及到用户生活的方方面面，存储了用户的大量隐私和信息，智能手机系统的安全性显得尤为重要。而近两年关于智能手机信息泄露事件不断发生，引起了用户的广泛担忧。作为智能手机的核心，系统的安全性成为系统开发人员工作的中心。当今智能手机系统主要为两大阵营：谷歌公司开发的 Android 和苹果公司开发的 iOS。iOS 因其系统封闭性较强，所以安全性略胜一筹，而近两年 Android 则通过增强权限管理和改变加密方式是的安卓系统的安全性有了明显提高。手机系统的安全性如此重要，因而对于手机系统的安全评估则显得尤为重要。

我们选用小组成员持有的 HTC One m8 手机和华为荣耀 7 进行评估，其中 HTC m8 搭载 Android 6.0 操作系统，荣耀 7 打搭载 Android 5.0.2 系统。

2. 模型假设

(1)假设所有观测数据均准确可靠，可以根据系统受入侵事件影响范围和发生频率确定各因素对信息系统安全程度的影响。

(2)假设信息系统安全程度只与所设定的七种不利因素和四种有利因素有关，不考虑其他因素对信息系统安全的影响。

(3)假设问题中所有信息系统自身防护功能都是完善的，总是按照对自身最有利的方式运行。对各种因素和信息系统安全程度模拟多元线性回归分析，剔除掉相关性较小的影响因素。

3. 符号定义

S——信息系统的安全程度

N——信息系统安全的不利影响因素破坏能力

N_1——自然及不可抗拒因素破坏能力

N_2——硬件及物理因素破坏能力

N_3——电磁波因素破坏能力

N_4——软件因素破坏能力

N_5——数据因素破坏能力

N_6——人为及管理因素破坏能力

N_7——其他因素破坏能力

P——信息系统安全的有利影响因素免疫能力

P_1——防火墙免疫能力

P_2——网络入侵检测系统免疫能力

P_3——网络漏洞扫描系统免疫能力

P_4——防病毒系统免疫能力

4. 数据

HTC m8				荣耀 7			
破坏(N)				破坏(N)			
	值(n)	发生概率(α)			值(n)	发生概率(α)	
自然及不可抗拒因素	5	14%	0.7	自然及不可抗拒因素	5	18%	0.9
硬件及物理因素	5	16%	0.78	硬件及物理因素	4	21%	0.54
电磁波因素	3	23%	0.69	电磁波因素	4	25%	1
软件因素	4	26%	1.04	软件因素	4	34%	1.36
数据因素	4	52%	2.08	数据因素	5	38%	1.9
人为及管理因素	5	4%	0.2	人为及管理因素	5	2%	0.1
其他因素	3	9%	0.27	其他因素	3	5%	0.15
抵抗(P)				抵抗(P)			
	值(p)	比重(β)			值(p)	比重(β)	
防火墙	5	41.87%	2.0935	防火墙	4	43.92%	1.7568
网络入侵检测系统	3	22.42%	0.6726	网络入侵检测系统	3	15.12%	0.4536
网络漏洞扫描系统	3	16.15%	0.4845	网络漏洞扫描系统	2	12.85%	0.257
防病毒系统	4	19.56%	0.7824	防病毒系统	4	28.11%	1.1244

（数据来自 Zealer 中国）

5. 结果与分析

5.1 将数据代入问题一所建数学模型的多元线性回归方程 $S=1.838+1.293P-0.973N=1.838+1.293*(0.951-0.086P_1+1.188P_2+2.722P_3+1.707P_4)-0.973*(1.707+0.136N_1-0.028N_2+0.440N_3-0.631N_4+0.452N_5-0.419N_6+1.915N_7)$

得到:

HTC m8 的信息系统安全程度 S=4.33　安全等级为 5，评语为好；

华为荣耀 7 的信息系统安全程度 S=4.12，安全等级为 5，评语为好。

5.2 将数据带入问题二所建数学模型验证

剔除 P_1 得到的多元线性回归方程为：$S_1=1.163P_2+2.602P_3+0.879P_4+0.904$

剔除 P_2 得到的多元线性回归方程为：$S_2=0.025P_1+2.956P_3+0.937P_4+1.101$

剔除 P_3 得到的多元线性回归方程为：$S_3=0.705P_1+1.545P_2+1.287P_4+0.320$

剔除 P_4 得到的多元线性回归方程为：$S_4=-0.027P_1+1.237P_2+2.998P_3+1.664$

5.2.1 验证 HTC m8 信息系统安全最重要的度量指标

剔除 P_1 得到的多元线性回归方程为：$S_1=3.63$

剔除 P_2 得到的多元线性回归方程为：$S_2=3.32$

剔除 P_3 得到的多元线性回归方程为：$S_3=2.84$

剔除 P_4 得到的多元线性回归方程为：$S_4=3.89$

数据验证可得，剔除 P_3 时，因变量 S 变化量最大，即 P_3(网络漏洞扫描系统

的免疫能力)对信息系统的安全程度影响最大。

5.2.2 验证华为荣耀 7 信息系统安全最重要的度量指标

剔除 P_1 得到的多元线性回归方程为：S_1=3.09

剔除 P_2 得到的多元线性回归方程为：S_2=2.96

剔除 P_3 得到的多元线性回归方程为：S_3=2.79

剔除 P_4 得到的多元线性回归方程为：S_4=2.94

数据验证可得，剔除 P_3 时，因变量 S 变化量最大，即 P_3(网络漏洞扫描系统的免疫能力)对信息系统的安全程度影响最大。

模型的评价与改进

我们建立的模型变量很多，而且手机到数据很庞大，采用多元线性回归模型，更加简单方便，而且运用回归模型，只要采用的模型和数据相同，通过标准的统计方法可以计算出唯一的结果，但在图和表的形式中，数据之间关系的解释往往因人而异，不同分析者画出的拟合曲线很可能也是不一样的；且回归分析可以准确地计量各个因素之间的相关程度与回归拟合程度的高低，提高预测方程式的效果。

1.模型的优点:

(1)模型是在充分统计了 2002 年到 2015 年数据信息后建立的，通过不断地分析、检验和完善使得模型具有较高的精确性，同时也确保了思维的科学性，和整体的模型结构的严谨性。

(2)对于模型得到的结果，并能联系全文不同模型所得结果，合理地分析，反复推测，最后验证模型的可行性。

(3)对衡量模型的指标进行标准化处理，使指标表示的含义更直观易懂。

(4)数据处理及模型求解时充分运用了 MATLAB 等数学软件，较好地解决了问题，得到了较理想的结果。充分用了题目中的各种信息，并且较好地结合了对模型的检验。

(5)问题二引入了逐步回归分析的方法，它的基本思想是将因子一个个引入，引入因子的条件是，该因子的偏回归平方和经检验时显著的。同时，每引入一个新因子后，要对老因子逐个检验，将偏回归平方和变为不显著的因子一一剔除。这种方法不需要计算偏相关系数，计算较简便，并且由于每步都作检验，因而保证了最后所得的方程中所有因子都是显著的。

2.模型的缺点

(1)数据所跨时间周期很长，所涉及的数据很庞大，可能会由于数据的丢失、数据失真、数据的不全面等问题使本来以为较理想的模型在实际中存在超出误差范围以为的错误。

(2)在对问题一、问题二的模型检验中使用的是我们自己的手机，所试验的样本个数少，随机误差大，不能完全说我们建模的模型正确。

3. 模型的改进

(1)提升对破坏能力与抵抗能力的量化精度，原模型中精确到各位，为提高精确度，可以将量化精度精确到小数点后一至两位，以提高模型的精度、减小偏差；同时，可以增加量化等级，本模型中均化为五个等级，可以提升为十个等级，细化安全等级，提高精确性与严谨性；

(2)建模中将破坏因素与抵抗能力视为较为独立的两部分，在现实中两者有一定的关联，对于要求较高的信息系统，可以考虑两者之间的关系，进而建立相应的的数学模型。

(3) 本次建模采用的数据来自同一来源的不同时间，时间的延伸性较好，但缺乏广泛性，可以应用更广泛的数据来提高数学模型的稳定性。

参考文献

【1】周凯，宋军全，邹学军. 数学建模竞赛入门与提高. 杭州：浙江大学出版社，2011.

【2】司守奎，孙玺菁. 数学建模算法与应用. 北京：国防工业出版社，2011.

【3】姜启源. 数学模型. 高等教育出版社第三版,2003年.

【4】刘来福，曾文艺. 问题解决的数学模型方法. 北京师范大学出版社,1999.

【5】王适文. 国家安全检测数据. http://www.itsec.gov.cn. 国家计算机网络应急技术处理协调中心,北京 ,2016年5月21日.

【6】Rik Ferguson. http://www.ctocio.com/security/6639.html. 主流移动操作系统安全性横向评测,2016年5月22日.

【7】吴功宜. 你的计算机安全吗. 天津科技,2001年02期,P9-12.

【8】岳树民，杜立群,齐海. 1999-2002年计算机病毒的状况、趋势、流行病毒的分析. 网络安全技术与应用, 2003年03期,P31-34.

附录

1．利用 SPSS 软件对因变量 P 进行多元线性回归方程建立的源程序如下：

```
GET
  FILE='C:\Users\Thinkpad\Desktop\数据\N最终值.sav'.
DATASET NAME 数据集1 WINDOW=FRONT.
REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT N
  /METHOD=ENTER N1 N2 N3 N4 N5 N6 N7
  /RESIDUALS NORMPROB(ZRESID) ID(年份)
  /SAVE PRED.
```

2．利用SPSS软件对因变量P进行多元线性回归方程建立的源程序如下：

```
GET
  FILE='C:\Users\Thinkpad\Desktop\数据\P初始值.sav'.
DATASET NAME 数据集3 WINDOW=FRONT.
REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT P
  /METHOD=ENTER P1 P2 P3 P4
  /RESIDUALS NORMPROB(ZRESID) ID(YEAR)
  /SAVE PRED.
```

3．利用SPSS软件对因变量S进行多元线性回归方程建立的源程序如下：

```
GET
  FILE='C:\Users\Thinkpad\Desktop\数据\S数据.sav'.
DATASET NAME 数据集4 WINDOW=FRONT.
REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT S
  /METHOD=ENTER N P
  /RESIDUALS NORMPROB(ZRESID) ID(年份)
  /SAVE PRED.
```

4．利用SPSS软件对自变量N，P进行逐步回归分析的源程序如下：

```
DATASET ACTIVATE 数据集1.
DATASET CLOSE 数据集3.
DATASET ACTIVATE 数据集4.
REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT S
  /METHOD=STEPWISE N P
  /RESIDUALS NORMPROB(ZRESID) ID(年份)
  /SAVE PRED.
```

5．利用SPSS软件对自变量P_1，P_2，P_3，P_4进行逐步回归分析的源程序如下：

```
DATASET ACTIVATE 数据集3.
REGRESSION
  /MISSING LISTWISE
  /STATISTICS COEFF OUTS R ANOVA
  /CRITERIA=PIN(.05) POUT(.10)
  /NOORIGIN
  /DEPENDENT P
  /METHOD=STEPWISE P1 P2 P3 P4
  /RESIDUALS NORMPROB(ZRESID) ID(YEAR)
  /SAVE PRED.
```