**数据统计与分析第7章作业**

**学号： 姓名： 班级： 总分：**

要求写出解题过程。**请将作业转换为PDF后再上传到教学平台**

第1题

今测得汽车的行驶速度speed和刹车距离dist数据如下：

speed: 4, 4, 7, 7, 8, 9

dist: 2, 10,4,22,16,10

(1) 作speed和dist的散点图，并以此判断speed与dist之间是否大致呈线性关系；

(2) 计算speed与dist的相关系数并做假设检验；

(3) 做speed对dist的OLS(普通的最小二乘线性回归，如教材所讲)回归，并给出回归直线；

(4) 对回归系数进行假设检验，并说明能否接受汽车速度对刹车距离有显著线性影响；

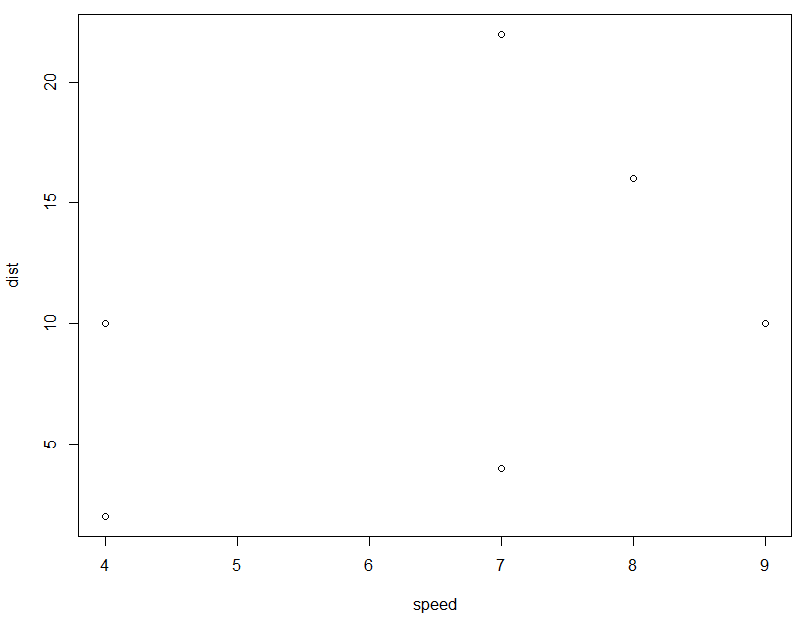
(5) 预测当speed = 30时，dist等于多少？

参考解答：

(1) 习惯上，横轴是自变量，纵轴是因变量。

speed = c(4, 4, 7, 7, 8, 9);dist = c(2, 10,4,22,16,10);

plot(speed,dist);



plot(dist,speed)也行。

从图形看，speed与dist的线性关系不明显。

(2)

cor(speed, dist);

[1] 0.4144095

注释：0.4144095是样本的一次测量值的相关系数，用r来表示样本的相关系数，那么0.4144095只是r的一次测量的结果，如果我们测量很多次，那么r会有到很多个，因此r是随机变量。我们需要根据已知的样本来检验总体中，speed和dist是否存在线性关系。总体speed和dist的相关系数记为。

。用函数 cor.test()来检验，该函数中默认。当时，统计量

注意：因为r是随机变量，所以t也是随机变量。因为随机变量r是的估计量，所以若**不应该很大**。我们假设**较大**(姑且称大于某个数k为较大)的概率为，即，等价于。即当时，我们接受**较大这个假设**，也即是拒绝**不应该很大(拒绝)这个假设。**因为随机变量t服从自由度为n-2的t分布,并且t分布是对称的，由很容易得到k1。记为自由度为m的上分位数，即。

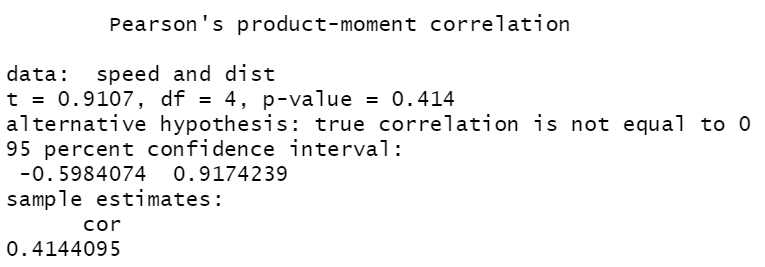
因此，有。从而，若在某一次的测量中，有测量值，即

，那么就拒绝.(注：，由对称性可得)

.

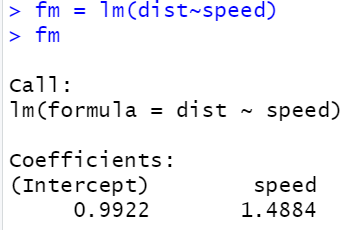
cor.test()中，默认

cor.test(speed,dist)

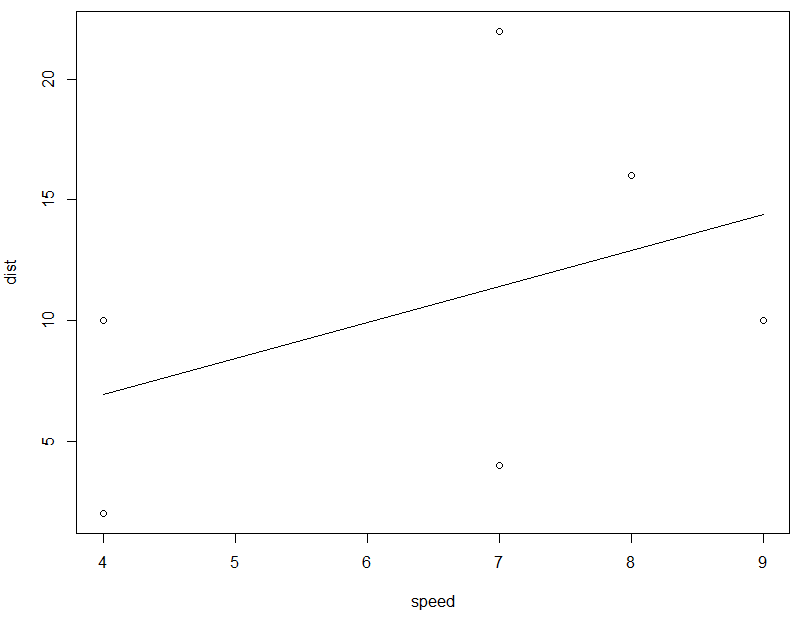


容易算的。因为测量值,所以不能拒绝，即接受(没有相关性)。上面的p-value = ，不能拒绝H0。

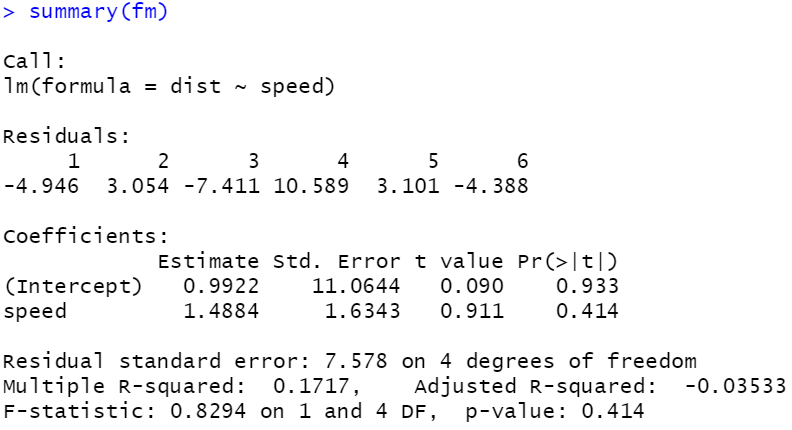
**(3)** fm = lm(dist~speed);习惯上横坐标表示自变量，纵坐标表示因变量。按题意写成fm = lm(speed~dist)也可以。





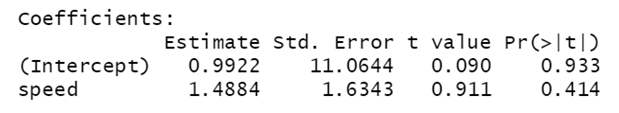


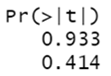
用函数summary(fm)可以对回归系数做检验

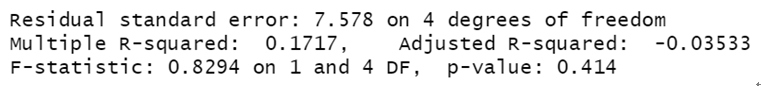


注释：

回归模型。回归直线：，，，以及都是统计量，分别用来对进行检验的统计量。在summary()函数中，默认对总体参数的检验采用t检验，且对

中给出了。

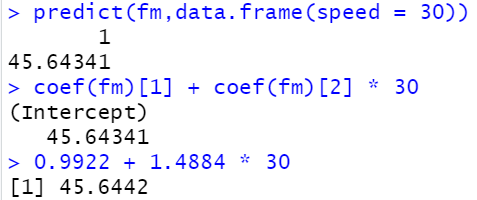
从来看，**不能拒绝.即可以认为speed和dist没有线性关系。**



也给出了几个统计量。不用考虑。

因此，调用函数summary()可以得到常用的统计量的测量值，以及完成对回归系数的检验。

(4) 两种方法进行预测，结果一致。



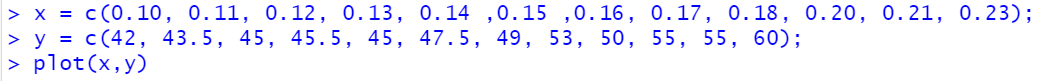
第2题

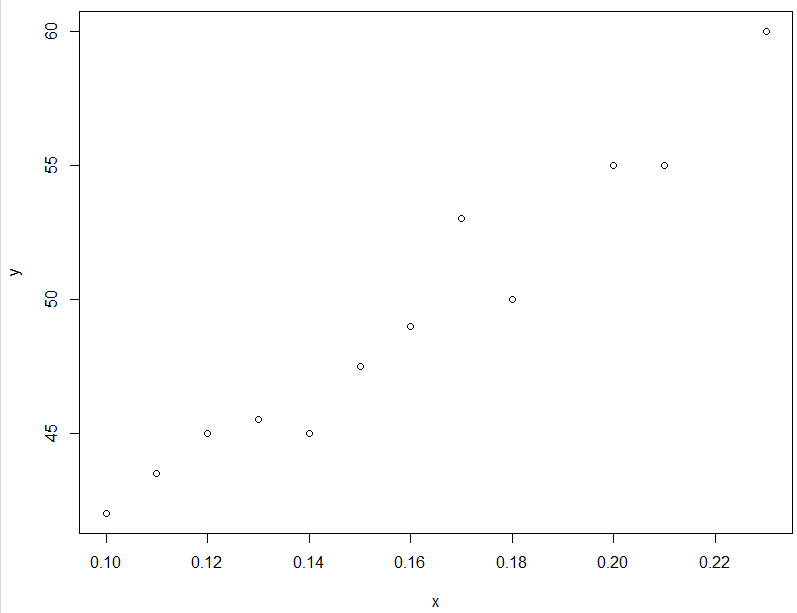
由专业知识可知，合金的强度y(107Pa)与合金中的碳含量x(%)有关。为了生产出强度满足顾客需要的合金，在冶炼时应该如何控制碳的含量？如果在冶炼过程中通过化验得知了碳的含量，能否预测这炉合金的强度？

x: 0.10, 0.11, 0.12, 0.13, 0.14 ,0.15 ,0.16, 0.17, 0.18, 0.20, 0.21, 0.23

y: 42, 43.5, 45, 45.5, 45, 47.5, 49, 53, 50, 55, 55, 60

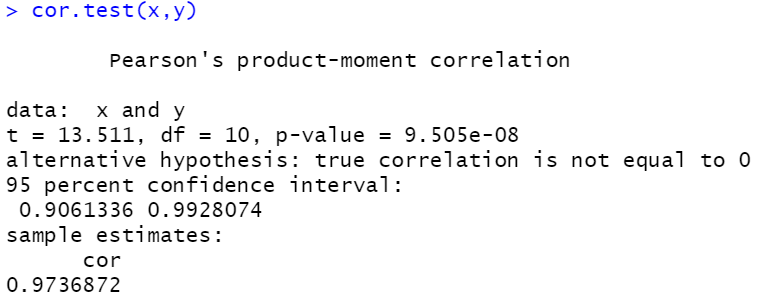
(1) 作x与y的散点图，并以此判断x与y之间是否呈大致线性关系；





经观察，大致呈线性关系。

(2) 计算x与y的相关系数并做假设检验；



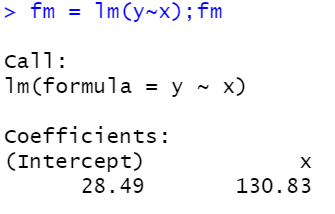
从上面看出，样本相关系数r=0.9736872。cor.test()对总体的相关系数做检验，用的是t检验,检验中默认.

从结果看出，统计量t的测量值,自由度df=10,,拒绝域

.,因此拒绝,即认为x,y有线性关系。

当然，从p-value = 9.505e-08 < 0.05,也可以得出拒绝。的95%置信区间为[0.9061336,0.9928074]

(3) 做y对x的最小二乘回归，并给出回归直线；

,回归直线

(4) 预测当x=0.22时,y等于多少？预测当x=0.25时，y等于多少？

