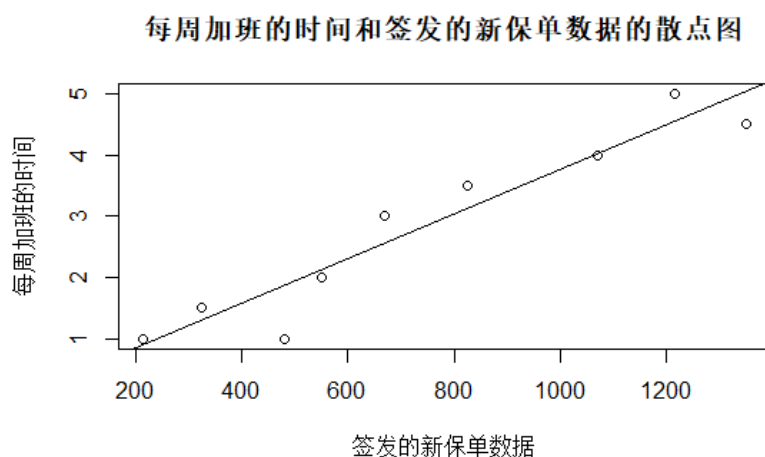


[ggplot绘图系统之qplot](#)

回归分析 (Regression Analysis) 是用来确定2个或2个以上变量间关系的一种统计分析方法。如果回归分析中, 只包括了一个自变量X和一个因变量Y时, 且它们的关系是线性的, 那么这种回归分析称为一元线性回归分析。

```
1 x<-c(825,215,1070,550,480,1350,325,670,1215)
2 y<-c(3.5,1,4,2,1,4.5,1.5,3,5)
3 library(ggplot2)
4 plot(x,y,main="每周加班的时间和签发的新保单数据的散点图",xlab="签发的新保单数据",ylab="每周
5 abline(lm(y~x))
6
7 #每周加班的时间和签发的新保单数据呈线性关系, 因而可以考虑一元线性回归模型
```



```
1 #2. 求回归方程, 并对相应的方程做检验。程序中lm()函数表示作线性模型, summary()函数提取模型的计
2 a<-lm(y~x)
3 > summary(a)
4
5 Call:
6 lm(formula = y ~ x)
7
8 Residuals:
9      Min       1Q   Median       3Q      Max
10 -0.87004 -0.12503  0.09527  0.37323  0.45258
11
12 Coefficients:
13      Estimate Std. Error t value Pr(>|t|)
```

```

14 (Intercept) 0.1215500 0.3588377 0.339 0.745
15 x          0.0036427 0.0004303 8.465 6.34e-05 ***
16 ---
17 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
18
19 Residual standard error: 0.485 on 7 degrees of freedom
20 Multiple R-squared:  0.911, Adjusted R-squared:  0.8983
21 F-statistic: 71.65 on 1 and 7 DF, p-value: 6.344e-05

```

残差的标准差 (Residual standard error) 为0.485，其自由度为 7 (计算方式: $n-2$);

相关系数的平方 (Multiple R-squared) 即判别系数为 0.911,

Adjusted R-squared (调整后的R方) : 0.8983，说明整个模型的拟合度较高，两者存在较强的线性关系

F统计量(F-statistic) 的自由度(1,7),回归方程为: **Intercept的意思是截距**

$$\hat{y} = 0.12155 + 0.0036427x$$

```

1 #3.参数的区间估计
2 > confint(a)
3           2.5 %      97.5 %
4 (Intercept) -0.726966290 0.970066254
5 x           0.002625117 0.004660271

```

常数项的置信区间为[-0.726966290,0.970066254],系数的置信区间为 [0.002625117,0.004660271]

```

1 #4.预测
2 > new<-data.frame(x=600)
3 > new
4     x
5 1 600
6
7 > new2<-predict(a,new,interval = "confidence",level = 0.95)
8 > new2
9      fit      lwr      upr
10 1 2.307166 1.897614 2.716719

```

平均值的置信区间为[1.897614,2.716719]

1 #5.残差分析

2 > e<-resid(a,digits=5) #计算回归a的残差，并保留5位小数

3 > e

	1	2	3	4	5	6
4						
5	0.37322742	0.09527080	-0.01923262	-0.12503171	-0.87004313	-0.53918695
6	7	8	9			
7	0.19457445	0.43784500	0.45257674			