

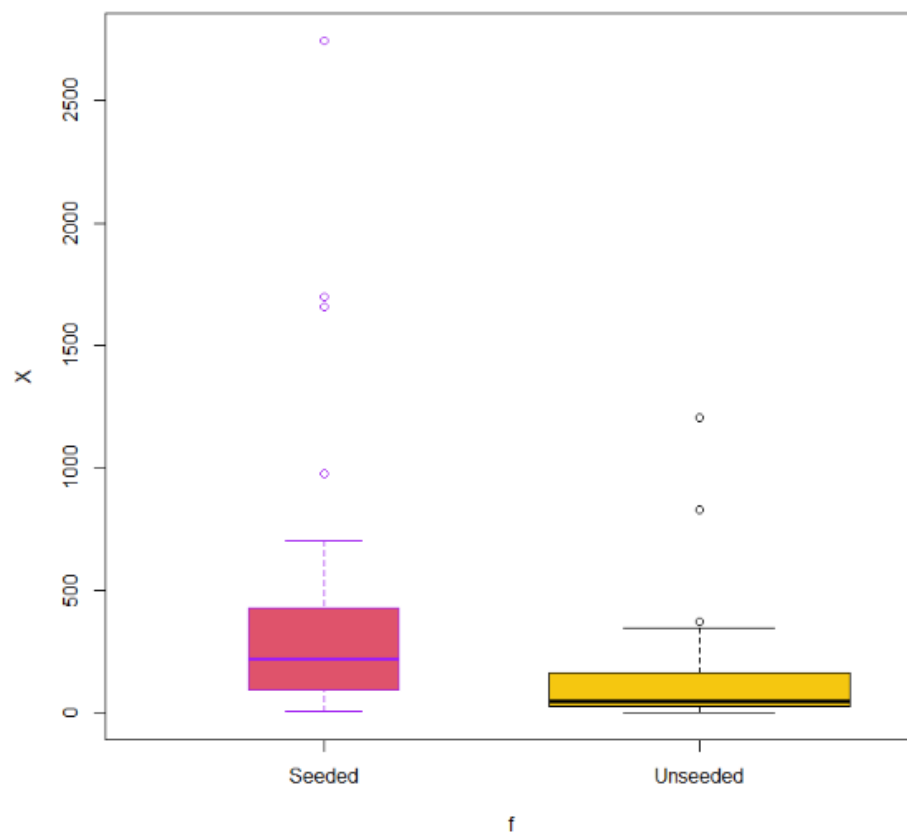
1 Cloud Seeding

在 CSDN 上看的分组画箱型图的方法；第二题根据 T 检验公式。

```
library(tidyrr)
library(dplyr)
library(ggplot2)

#1.1
X<-c(1202.6, 830.1, 372.4, 345.5, 321.2, 244.3, 163.0,
     147.8, 95.0, 87.0, 81.2, 68.5, 47.3, 41.1, 36.6,
     29.0, 28.6, 26.3, 26.0, 24.4, 21.4, 17.3, 11.5,
     4.9, 4.9, 1.0, 2745.6, 1697.1, 1656.4, 978.0, 703.4,
     489.1, 430.0, 334.1, 302.8, 274.7, 274.7, 255.0,
     242.5, 200.7, 198.6, 129.6, 119.0, 118.3, 115.3,
     92.4, 40.6, 32.7, 31.4, 17.5, 7.7, 4.1)
f<-factor(rep(c("Unseeded","Seeded"), each=26))#定义分组因子
Data<-data.frame(X,f)#生成数据框
boxplot(X~f,Data,width=c(1,2),col=c(2,7),border=c("purple","black"))

#1.2
Unseeded<-c(1202.6, 830.1, 372.4, 345.5, 321.2, 244.3, 163.0, 147.8, 95.0, 87.0, 81.2, 68.5, 47.3, 41.1, 36.6, 29.0, 28.6, 26.3, 26.0, 24.4, 21.4, 17.3, 11.5, 4.9, 4.9, 1.0)
Seeded<-c(2745.6, 1697.1, 1656.4, 978.0, 703.4, 489.1, 430.0, 334.1, 302.8, 274.7, 274.7, 255.0, 242.5, 200.7, 198.6, 129.6, 119.0, 118.3, 115.3, 92.4, 40.6, 32.7, 31.4, 17.5, 7.7, 4.1)
t.test(Unseeded,Seeded)#p值为0.05376, p值大于0.05, 表示接受原假设, 两组数据差异未达到显著水平(0.05)
#t值为-1.9983, 表示Unseeded组的样本均值低于Seeded组的均值。
```



2 Was Tyrannosaurus Rex Warm-Blooded?

如图，问了同学关于 excel 数据格式的问题，根据讲义写了这题，最终 p 值低于 0.05，看出两者存在差异，并证明暴龙是温血动物。

```

1 library(tidyr)
2 library(dplyr)
3 library(ggplot2)
4
5 Tyrannosaurus <- read.csv("Tyrannosaurus.csv", header = T, encoding = "UTF-8")
6 #数据导入excel, 另存为逗号分隔形式文件, 并读取
7 Tyrannosaurus <- as_tibble(Tyrannosaurus)
8
9 #使用one-way ANOVA test
10 anova_one_way <- aov(Oxygen.isotopic.composition~factor(Bone), data = Tyrannosaurus)
11 summary(anova_one_way)
12 #P值低于0.05, 数据之间存在统计差异, 两者之间存在差异
13
14 ggplot(Tyrannosaurus, aes(x = Bone, y = Oxygen.isotopic.composition, fill = Bone)) +
15   geom_boxplot() +
16   theme_classic()
17 #从图形中也可看出两者存在差异
18 #以上可以证明雷克斯暴龙是温血动物

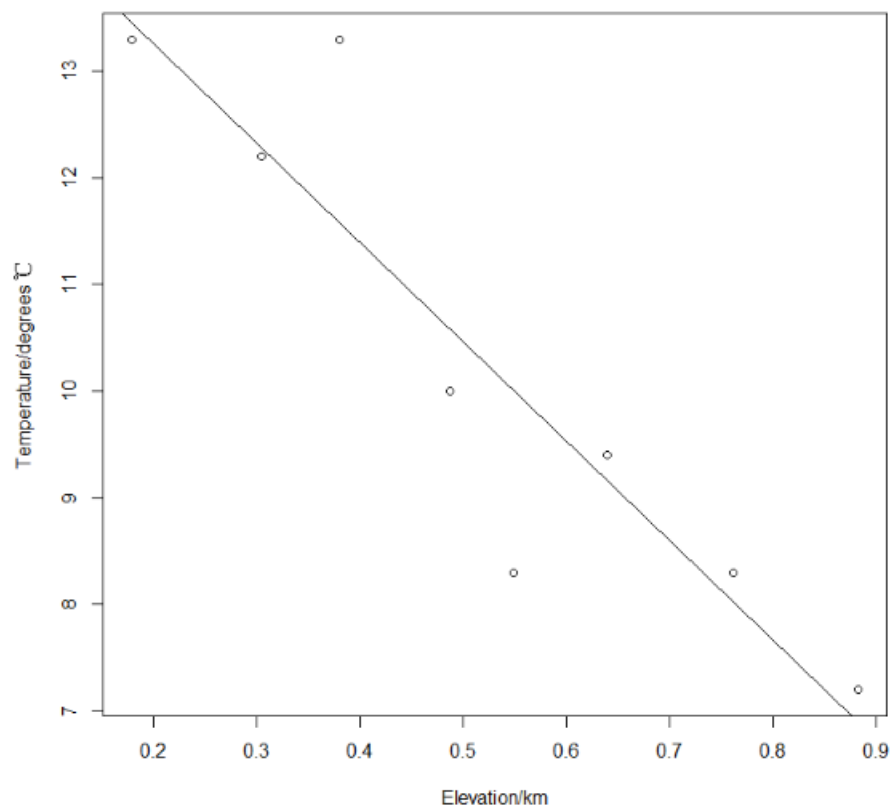
```

3 Vegetarians and Zinc

先进行单向 anova_one_way 分析, 说明两者存在差异, 然后分别对怀孕和素食这两个因素的组别进行 t 检验, 得出不同组别的平均值及 P 值。

4 Atmospheric Lapse Rate

直接输入数据, 数据很少, plot 画散点图, 分别改下 X、Y 的坐标名称, 在根据讲义用 model()、coef() 等, 再用 abline() 把直线画出来。



```

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   15.1249    0.9483   15.950 3.86e-06 ***
Elevation     -9.3121    1.6698   -5.577 0.00141 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.04 on 6 degrees of freedom
Multiple R-squared:  0.8383,    Adjusted R-squared:  0.8113
F-statistic: 31.1 on 1 and 6 DF,  p-value: 0.001411

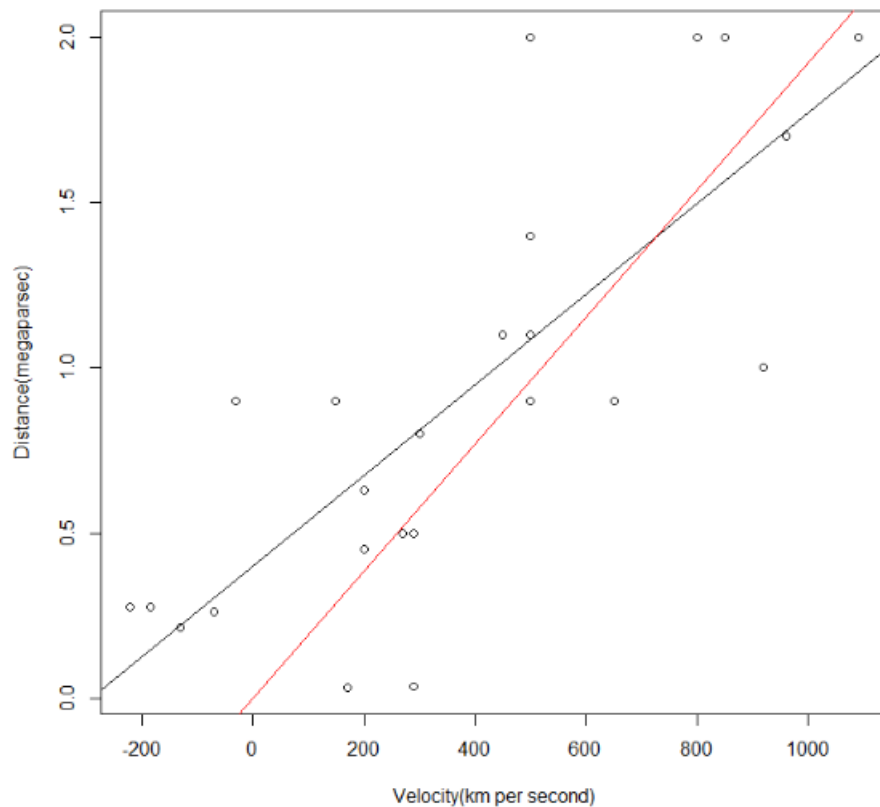
```

斜率为-9.3121，与-9.8 非常相近了。

5 The Big Bang Theory

#5.1 读数据，直接画图即可。散点分布在某一条线两边，可以明显看出这一趋势；

#5.2 用 `abline()`画一条线；



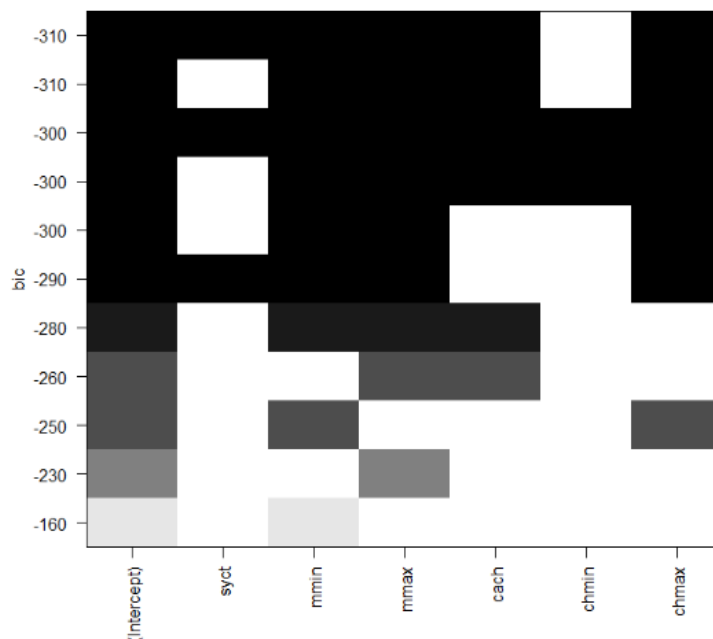
#5.3 如果理论正确，刚开始衰退速度小，没有经过时间的积累，所以距离为零，那么，截距为零；距离等于时间乘以速度，那么宇宙的年龄即为斜率，则假设正确。

第三小题询问了同学关于如何改进的思路，由图知，画出来的图宇宙伊始距离不为零肯定不对，但是宇宙的年龄是等于斜率且不变的，所以通过使截距（初始距离）变为零以后看斜率。通过在 `lm()` 函数里把速度根据图上的截距误差设为-1，使截距为 0，得到斜率如图即为年龄。

#5.4 更正误差后，结果更精确。

6 CPU Performance

#6.1 一步一步按照讲义做， `regsubsets` 求最大子集回归，并通过 `bic ()` 画图；



#6.2 由 1 知只有 `chmin` 没有关，所以用其余五个变量拟合用 `predict ()` 函数预测，求偏差，结果为 6.493262。

7 Analysis of Data Sets from Your Group

#7.1 某种男婴手长均数为 9.3cm， 12 名该种男婴双顶径数据如下：9.95 9.33 9.49 9.00 10.09 9.15 9.52 9.33 9.16 9.37 9.11 9.27（数据来源于网络），试问该种男婴的双顶径是否大于普通男婴手长？

直接 `t.test()` 两类数据即可；

#7.2 用课件 section7 中中波士顿住房数据做单向方差分析；

#7.3 同用 7.2 中数据。