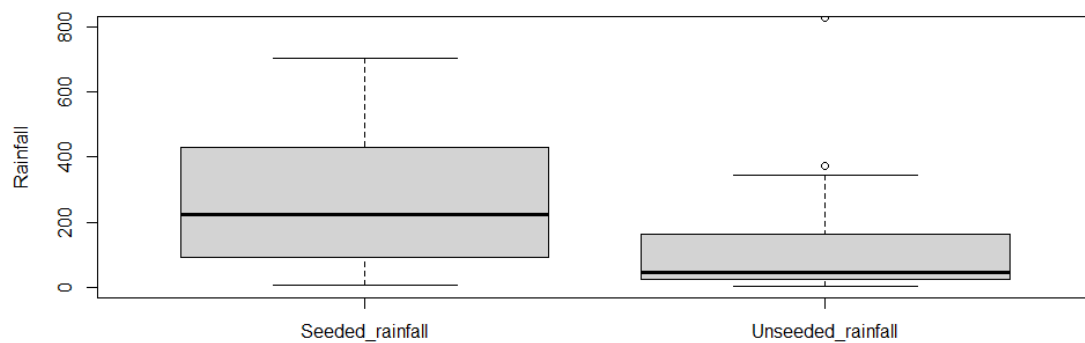


3.1

3.1.1



3.1.2

画直方图，可以发现降雨数据不是正态分布，因此进行单因素方差分析

```
> summary(anova_one_way3_1)
      Df Sum Sq Mean Sq F value Pr(>F)
f      1 1000360 1000360   3.993  0.0511 .
Residuals 50 12525457 250509
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

有 90%的概率认为这两组降水数据存在差异

3.2

进行单因素方差分析

```
> summary(anova_one_way3_2)
      Df Sum Sq Mean Sq F value    Pr(>F)
Bone1    9   5.688   0.6320    7.922 1.01e-06 ***
Residuals 42   3.351   0.0798
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

因此有 99.9%的概率认为这些骨头不是在同一温度下形成的，暴龙不属于恒温动物

3.3

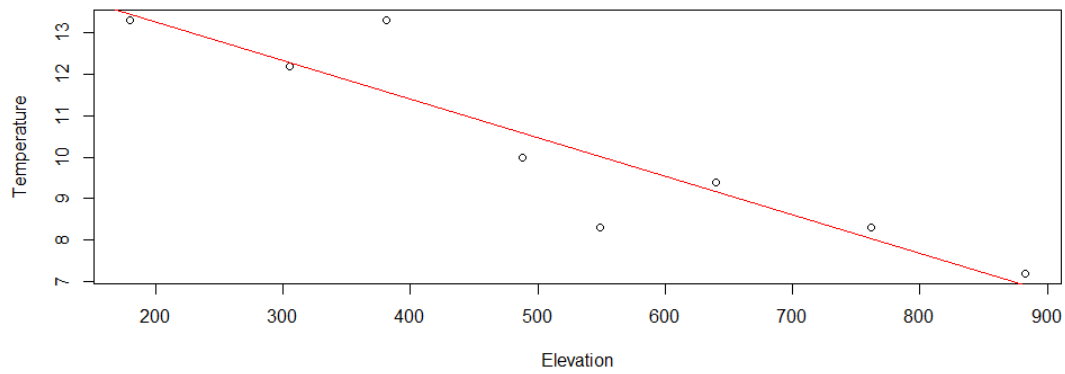
对是否怀孕、是否为素食者进行多因素方差分析

```
> anova_two_way3_3 <- aov(zinc ~ If_Pregnant + If_Vegetarians, data = data3_3)
> summary(anova_two_way3_3)
      Df Sum Sq Mean Sq F value Pr(>F)
If_Pregnant    1    13    12.8   0.029  0.866
If_Vegetarians  1     3     3.4   0.008  0.931
Residuals     20   8816   440.8
```

可以认为两种孕妇体内锌含量无显著区别

3.4

绘制散点图并添加回归线



```
Call:
lm(formula = Temperature ~ Elevation, data = data3_4)

Residuals:
    Min       1Q   Median       3Q      Max
-1.71254 -0.25668  0.07508  0.27763  1.72303

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 15.124887   0.948282  15.950 3.86e-06 ***
Elevation   -0.009312   0.001670  -5.577 0.00141 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

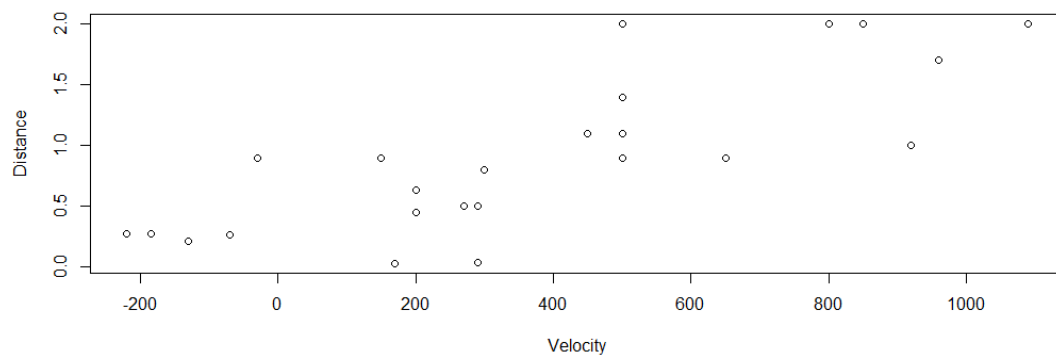
Residual standard error: 1.04 on 6 degrees of freedom
Multiple R-squared:  0.8383,    Adjusted R-squared:  0.8113
F-statistic: 31.1 on 1 and 6 DF, p-value: 0.001411
```

计算得到的气温直减率为 $-0.009312^{\circ}\text{C}/\text{m}$ ，即 $-9.312^{\circ}\text{C}/\text{Km}$

3.5

3.5.1

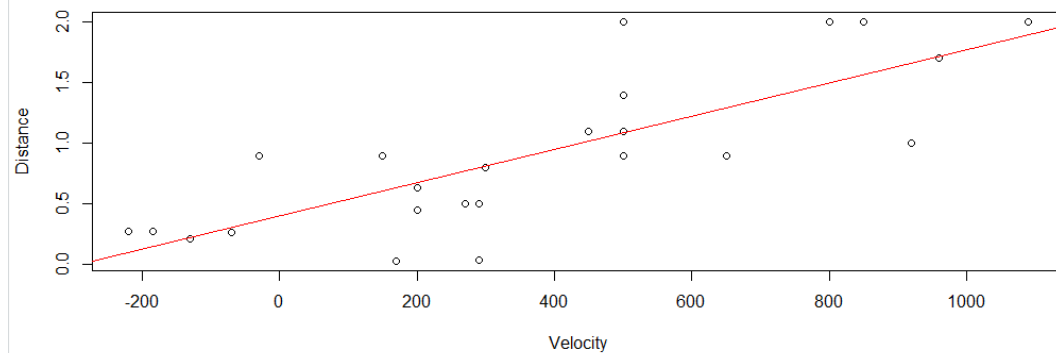
绘制散点图



可以发现 recession velocity 与 distance 近似成正相关

3.5.2

进行线性拟合并添加回归线



3.5.3

假设地球的速度为 v_0 ，宇宙诞生的时间为 k ，那么地球与宇宙大爆炸奇点之间的距离 $d_0 = k * v_0$

某星云 i 与地球之间的相对速度为 v_i ，那么该星云 i 与奇点之间的距离 $d_i = k * (v_0 + v_i)$

因此地球与星云 i 之间的距离 $d = d_i - d_0 = k * v_i$

因此，斜率 k 为宇宙诞生的时间，截距为 0

```
> #宇宙诞生的时间
> summary(model3_5_2)$coefficients[1,1]*1000000*30.9*1000000000000/60/60/24/365
[1] 1883048021
```

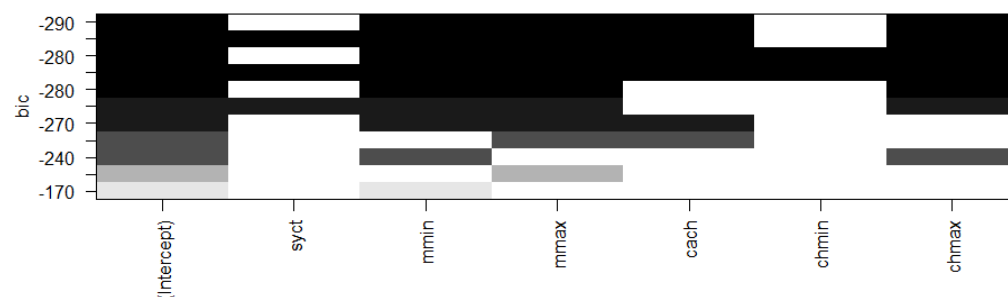
计算得到宇宙诞生的时间为 18.8 亿年，这个结果很明显不正确

3.5.4

改进对距离的测量可以得到更精确的回归方程和斜率

3.6

3.6.1



```
Step: AIC=1384.34
perf ~ syct + mmin + mmax + cach + chmax
```

	Df	Sum of Sq	RSS	AIC
<none>			618824	1384.3
- syct	1	18486	637310	1387.2
- cach	1	73555	692379	1401.1
- chmax	1	109320	728144	1409.5
- mmax	1	200804	819627	1429.3
- mmin	1	218142	836965	1432.8

```

Step: AIC=1387.25
perf ~ mmax + cach + mmin + chmax

      Df Sum of Sq  RSS   AIC
+ syct  1  18486.2 618824 1384.3
<none>          637310 1387.2
+ chmin  1    11.2 637299 1389.2

Step: AIC=1384.34
perf ~ mmax + cach + mmin + chmax + syct

      Df Sum of Sq  RSS   AIC
<none>          618824 1384.3
+ chmin  1    5.3659 618818 1386.3

Step: AIC=1384.34
perf ~ mmax + cach + mmin + chmax + syct

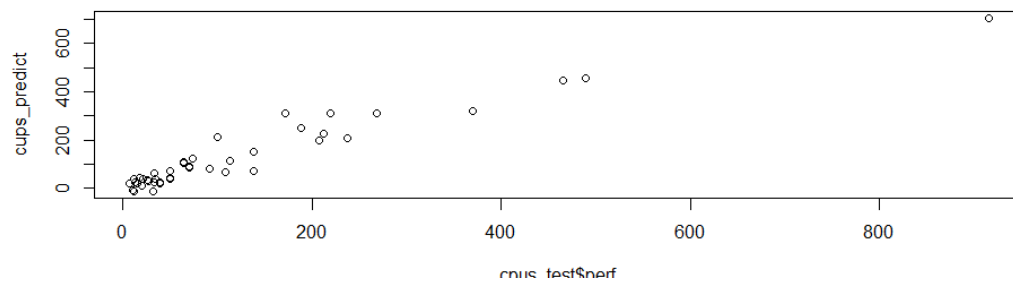
      Df Sum of Sq  RSS   AIC
<none>          618824 1384.3
+ chmin  1    5.3659 618818 1386.3
- syct  1  18486.2 637310 1387.2
- cach  1  73555.6 692379 1401.1
- chmax  1 109320.7 728144 1409.5
- mmax  1 200804.8 819627 1429.3
- mmin  1 218142.8 836965 1432.8

```

通过 bic 以及 forward/backward/stepwise regression, 得到最佳子集回归为 syct, mmin, mmax, cach 和 chmax

3.6.2

进行拟合, 将其应用于测试组, 得到散点图



相关性

```

> cor(cpus_test$perf, cups_predict)
[1] 0.9540719

```

以及平均偏差

```

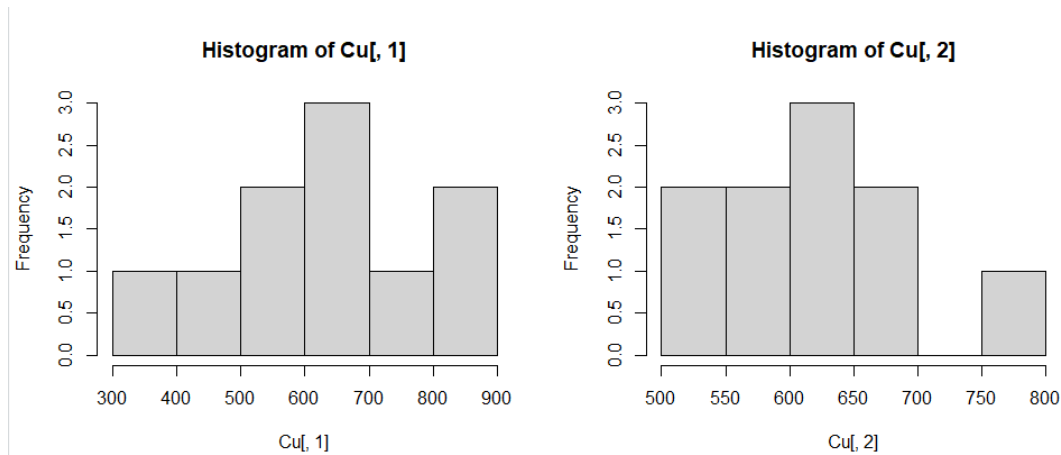
> (mean(cups_predict) - mean(cpus_test$perf))/
+ mean(cpus_test$perf)*100
[1] 2.935477
>

```

3.7

3.7.1 尾矿库 1m 深的铜矿矿渣和 0.1m 深的铜矿矿渣相比, 铜浓度有区别吗?

首先绘制直方图



可以认为数据是正态分布的（虽然每组只有 10 个值）

```
> t.test(Cu[,1] , Cu[,2])

welch Two sample t-test

data: Cu[, 1] and Cu[, 2]
t = 0.013818, df = 12.334, p-value = 0.9892
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -121.4032 122.9576
sample estimates:
mean of x mean of y
 615.4467  614.6695
```

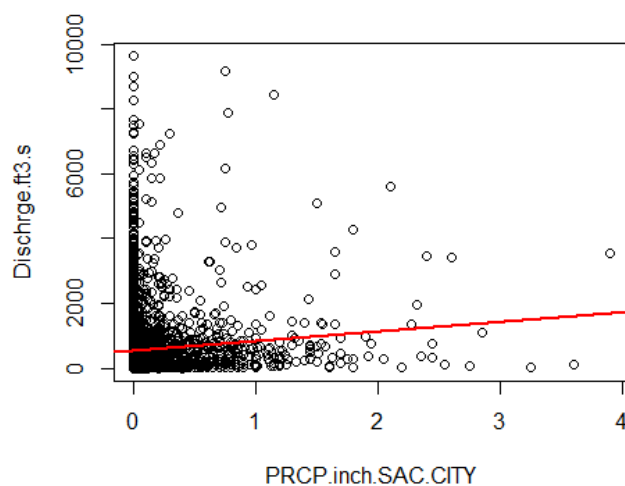
经过 t 检验，二者没有显著的浓度差异。

3.7.2 爱荷华州 sac city 和同纬度的 rockwell city 的降雨有差别吗？

```
> anova_one_way3_7 <- aov(PRCP ~ f,data = s_r)
> summary(anova_one_way3_7)
              Df Sum Sq Mean Sq F value Pr(>F)
f              1    0.0  0.01588    0.198  0.656
Residuals 8679 694.9  0.08007
47 observations deleted due to missingness
```

可以认为两者降水无显著差异

3.7.3 sac city 的降雨量与流经 sac city 的北浣熊河径流流量之间存在线性关系吗？



```

> summary(fit3_7)

Call:
lm(formula = Dischrge.ft3.s ~ PRCP.inch.SAC.CITY, data = discharge1)

Residuals:
    Min       1Q   Median       3Q      Max
-1475.6  -513.6  -320.3    95.5   9064.7

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)      565.26      15.16  37.283 < 2e-16 ***
PRCP.inch.SAC.CITY 289.81      51.63   5.613 2.11e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 954.2 on 4381 degrees of freedom
Multiple R-squared:  0.007141, Adjusted R-squared:  0.006914
F-statistic: 31.51 on 1 and 4381 DF, p-value: 2.108e-08

```

拟合效果并不理想，二者没有很显著的线性关系