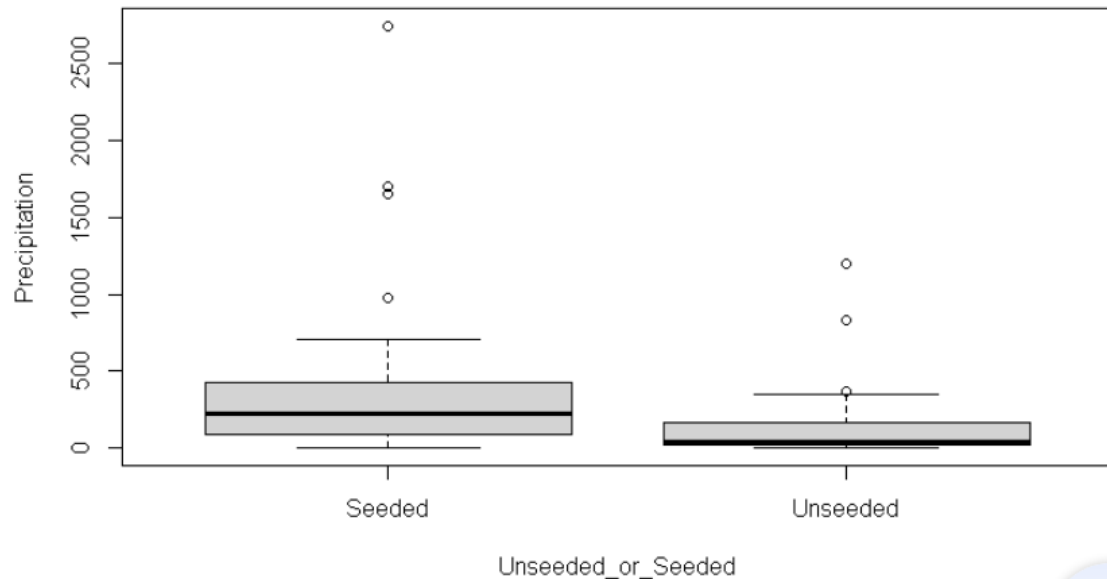


## Report\_PS3

### 1、箱线图：



从箱线图中可以看到，**Seeded** 一组的降水量整体上高于 **Unseeded** 一组的降水量，并且两组降水量都有较高的异常值。

### t-test 结果：

```
welch Two Sample t-test

data: Unseeded and Seeded
t = -1.9983, df = 33.856, p-value = 0.05376
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-559.552533  4.752533
sample estimates:
mean of x mean of y
164.5731  441.9731
```

可以看到  $p$  值为 0.0536，大于 0.05，因此在 95% 的水平上可以认为是否播种对降雨没有影响。

### 2、对数据进行方差分析，结果如图：

```
              Df Sum Sq Mean Sq F value    Pr(>F)
bone             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
```

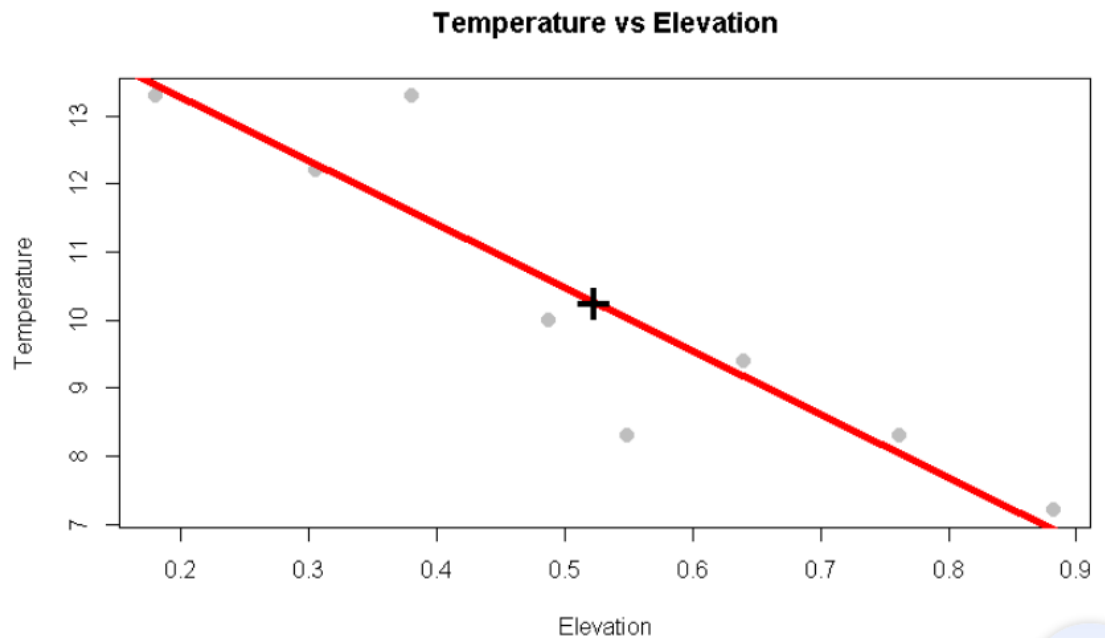
可以看到  $p$  值为  $1.01e-06$ ，远小于 0.05，因此可以认为不同骨骼的氧同位素含量均值有差异，因此此证据不支持暴龙是温血动物。

### 3、对数据进行方差分析（只选取了怀孕的女性），结果如下：

```
              Df Sum Sq Mean Sq F value    Pr(>F)
vegetarians_or_non 1     3     3.4   0.009  0.925
Residuals        16  5841  365.1
```

可以看到  $p$  值为 0.925，远远大于 0.05，因此没有证据支持素食怀孕者的锌水平比非素食怀孕者低。

4、如图所示：



拟合结果如下图所示：

```
Call:
lm(formula = Temperature ~ Elevation)

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

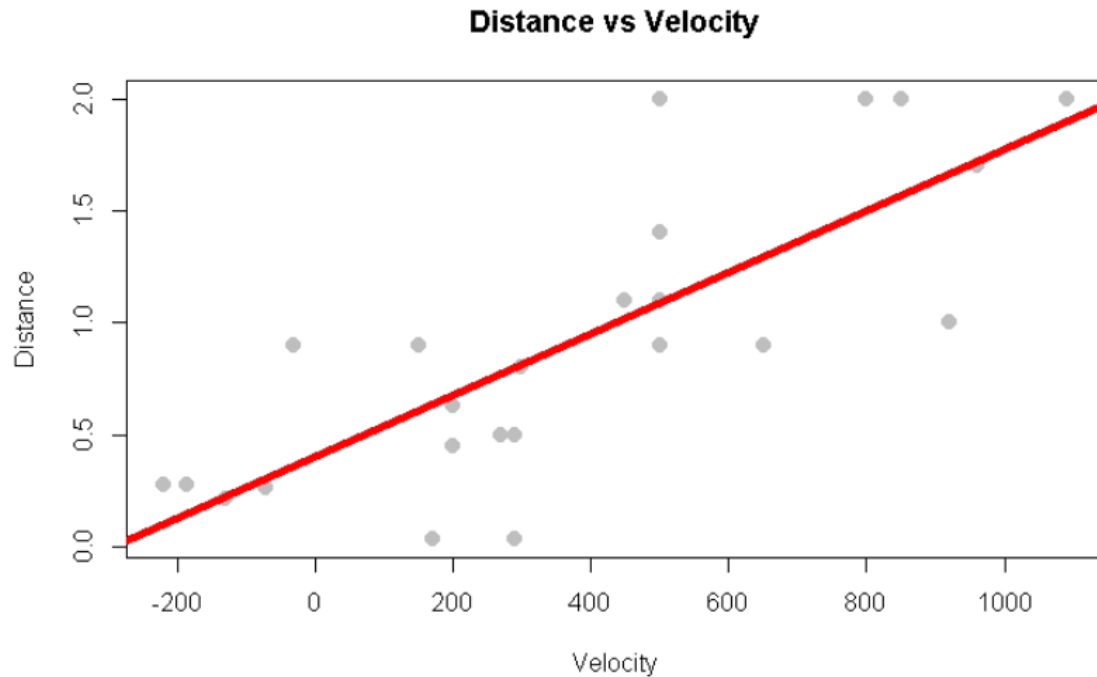
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  15.1249    0.9483   15.950 3.86e-06 ***
Elevation     -9.3121    1.6698   -5.577 0.00141 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

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
```

可以看到斜率为-9.3121，不确定度 1.6698，因此可以近似认为温度递减率为 9.8 degrees  $C km^{-1}$ 。

5、散点图及拟合直线图：

从图中可以看到，星系退行速度越快，离我们的距离也越远，二者近似成正相关。



拟合结果如下图所示：

```
Call:
lm(formula = Distance2 ~ Velocity2)

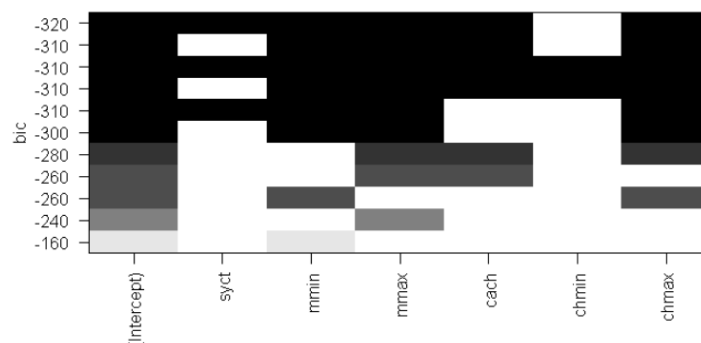
Residuals:
    Min       1Q   Median       3Q      Max
-2.358e+19 -7.268e+18 -2.719e+17  6.403e+18  2.826e+19

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.233e+19  3.661e+18   3.369  0.00277 **
Velocity2    1.347e+09  2.231e+08   6.036  4.48e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

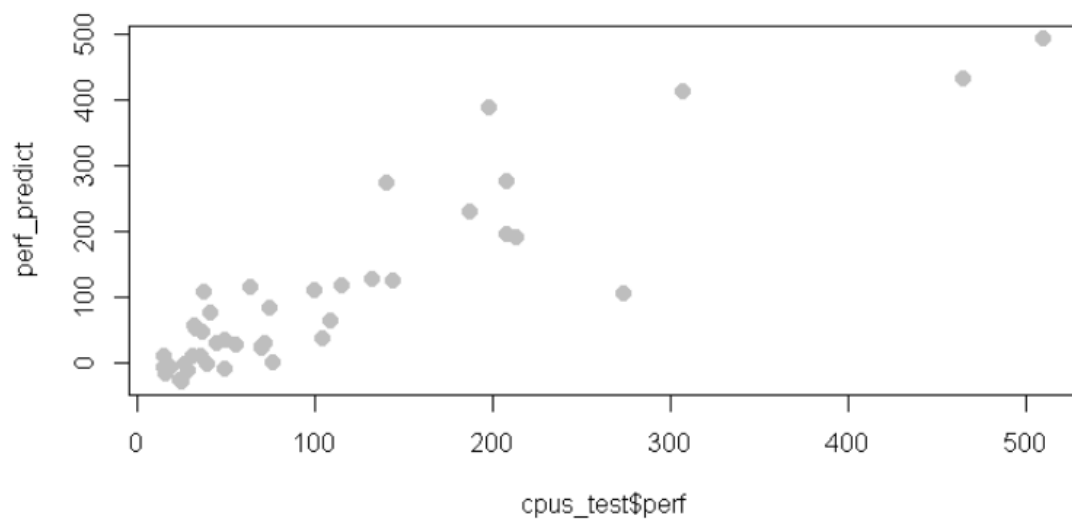
Residual standard error: 1.251e+19 on 22 degrees of freedom
Multiple R-squared:  0.6235,    Adjusted R-squared:  0.6064
F-statistic: 36.44 on 1 and 22 DF,  p-value: 4.477e-06
```

可以看到截距为  $1.233e+19$ ，并不为 0，并且斜率代表宇宙年龄，为  $1.347e+09$ ，即 13.47 亿年，与正常的认知（130 亿年）有较大的差距，因此提高距离测量的精确度对回归系数的估计还是很有必要的。

6、最佳子集回归如下图所示：



估计值与实际值的比较：



均值偏差：

```
> print(mean_bias)
[1] -6.808481
```

7、1: 通过 t-test 探究水面叶绿素浓度在 2000 年之前与 2000 年之后是否有显著差异

```
Welch Two Sample t-test

data: before and after
t = -2.782, df = 345.12, p-value = 0.0057
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -11.960350 -2.052914
sample estimates:
mean of x mean of y
 7.373103 14.379736
```

可以看到 p 值为 0.0057，小于 0.05，因此可以认为 2000 年前后水面叶绿素浓度有差异。

2: 通过 ANOVA 探究不同季节(选取 1, 4, 7, 10 月份)水面叶绿素浓度是否有显著差异

```
          Df Sum Sq Mean Sq F value Pr(>F)
date4      3  20757    6919   4.828 0.0033 **
Residuals 119 170532    1433
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
5 observations deleted due to missingness
```

可以看到 p 值为 0.0033,因此可以认为不同季节水面叶绿素浓度有差异。

### 3、线性拟合三十年来水面叶绿素浓度的值，观察水面叶绿素浓度的变化

```
Call:
lm(formula = Chlorophyll ~ date2, data = Data)

Residuals:
    Min       1Q   Median       3Q      Max
-13.841  -9.540  -7.374  -2.470  247.822

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -482.8904   305.9665  -1.578   0.115
date2         0.2469     0.1528   1.616   0.107

Residual standard error: 27.34 on 372 degrees of freedom
(12 observations deleted due to missingness)
Multiple R-squared:  0.006973, Adjusted R-squared:  0.004304
F-statistic: 2.612 on 1 and 372 DF,  p-value: 0.1069
```

拟合效果并不是非常好，但根据斜率为  $0.2469 > 0$ ，可以认为近三十年水面叶绿素浓度有增加的趋势。