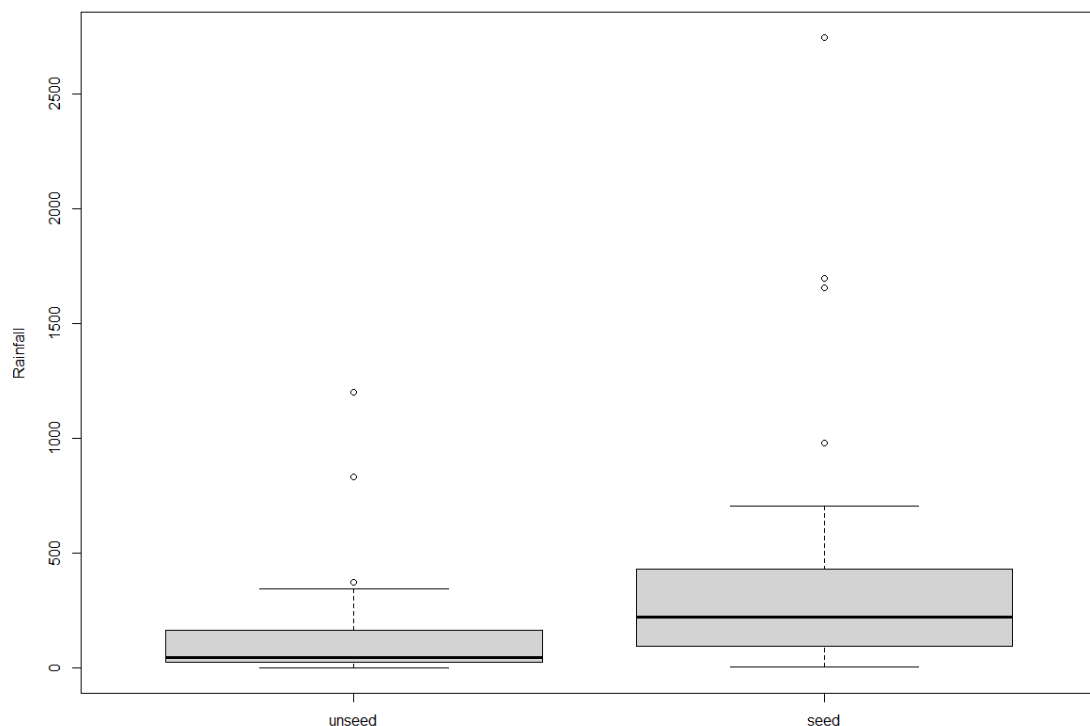


PS3 REPORTS

12032382 庄巩发

PS3_1

读取 unseed 和 seed 时候降雨量的两组数据，并将其并排画为箱型图。如图所示 Unseeded Days 中降水量整体偏小，降水量离散程度小；Seeded Days 中降水量整体有所升高，四分位间距较开，数据离散程度大。



由于需要判断 seed 对于降水量是否有影响，且每个独立样本数目小于 30，这里采用 t 检验的方式对两组数据进行均值检验。从结果上看， $p\text{-value} > 0.05$ ，表明两者均值差异不显著，即在该实验中无法确切的证明 seeding 对于降水量有影响。

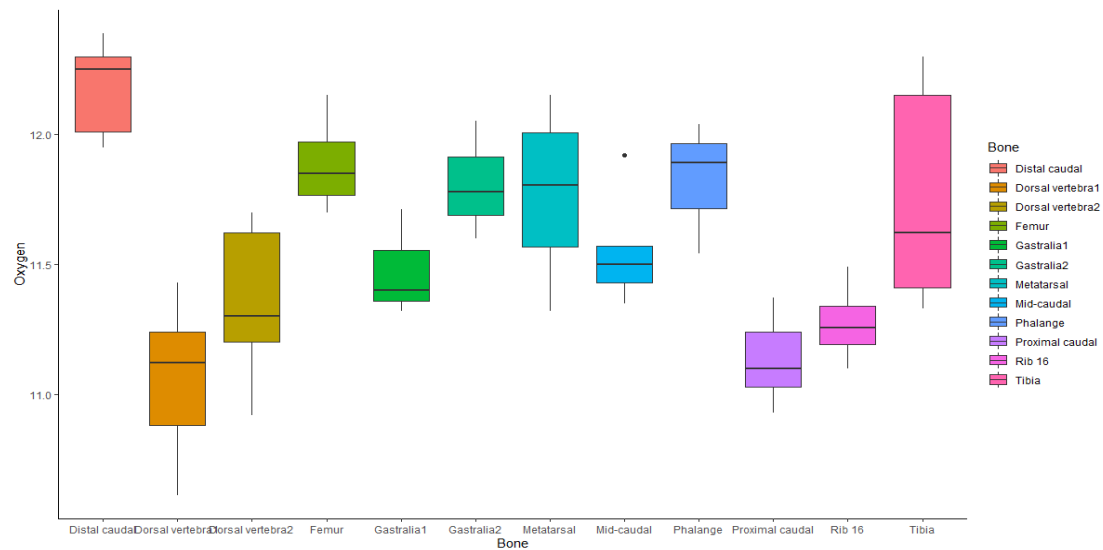
```
> t.test(Rainfall_unseed, Rainfall_seed)
```

```
Welch Two Sample t-test
```

```
data: Rainfall_unseed and Rainfall_seed
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
```

PS3_2

先对单个霸王龙骨架的 12 个骨骼样本中每个骨骼的磷酸盐中氧同位素组成的一些测量结果绘制分布图。



anova 分析多组数据进行均值检验。首先将数据写入 csv 表格中，再进行 anova 方差分析，结果：

```
> anova_one_way <- aov(Oxygen~factor(Bone), data = Tyrannosaurus_bone)
> summary(anova_one_way)
              Df Sum Sq Mean Sq F value    Pr(>F)
factor(Bone)  11   6.067   0.5516    7.427 9.73e-07 ***
Residuals     40   2.971   0.0743
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

从 anova 的结果来看， $\text{Pr}(>F) < 0.05$ ，可以认为霸王龙的骨骼样品含氧量存在差异，霸王龙不是温血动物。

PS3_3

此处认为“怀孕”这个因素是烟雾弹，忽略不计，将数据分为“素食”与“非素食”两组。进行 anova 分析。结果如下

```
> one_way_anova <- aov(vegetarian ~ non_vegetarian)
> summary(one_way_anova)
              Df Sum Sq Mean Sq F value    Pr(>F)
non_vegetarian  1   87.2    87.2    0.217  0.673
Residuals       3 1208.0   402.7
13 observations deleted due to missingness
> |
```

从 anova 的结果来看， $\text{Pr}(>F) > 0.05$ ，可以认为怀孕非素食主义者和怀孕素食主义者的女性体内 Zn 含量无明显差异。

PS3_4

导入数据，同时将高度单位转换为 km，绘制 Elevation(km)和 Temperature(degrees C)散点

图，然后使用 `lm()` 和 `abline()` 拟合回归曲线。

```
> summary(fit)

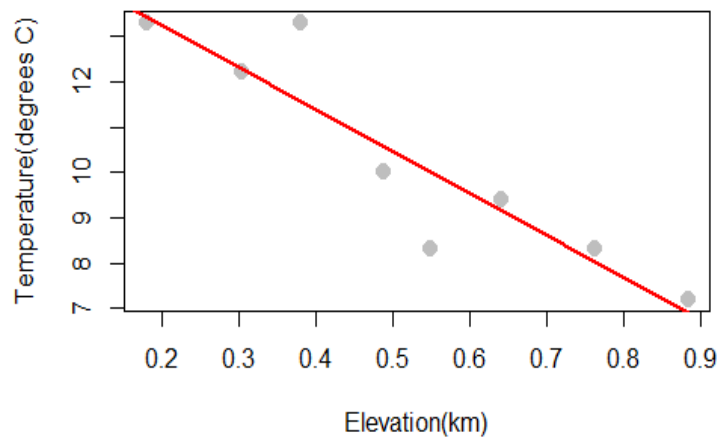
call:
lm(formula = Temp ~ Ele)

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 ***
Ele          -9.3121     1.6698   -5.577 0.00141 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

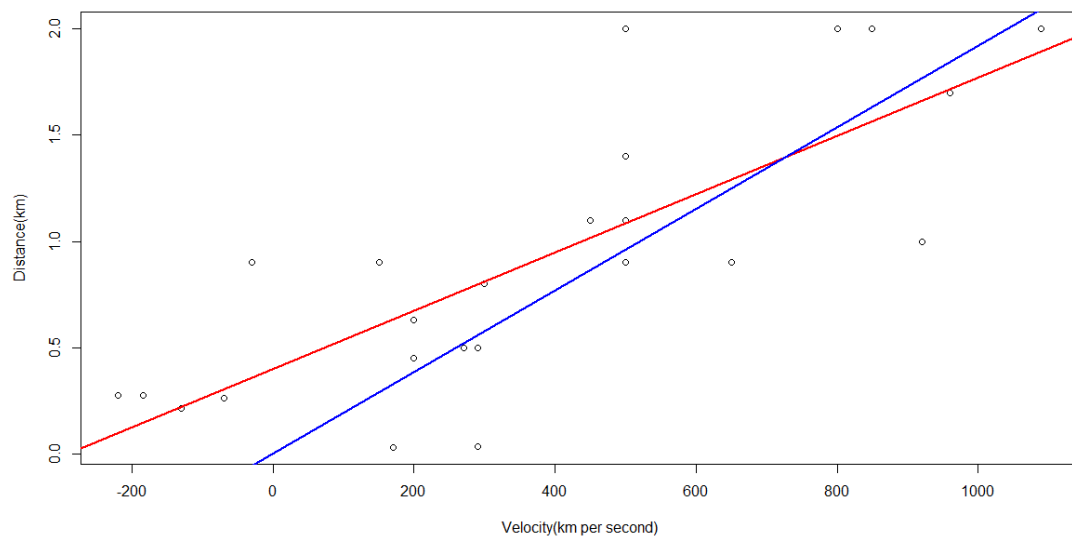
Residual standard error: 1.04 on 6 degrees of freedom
Multiple R-squared:  0.8383,    Adjusted R-squared:  0.8113
F-statistic: 31.1 on 1 and 6 DF, p-value: 0.001411

> abline(fit, lwd = 2, col = "red")
> |
```



PS3_5

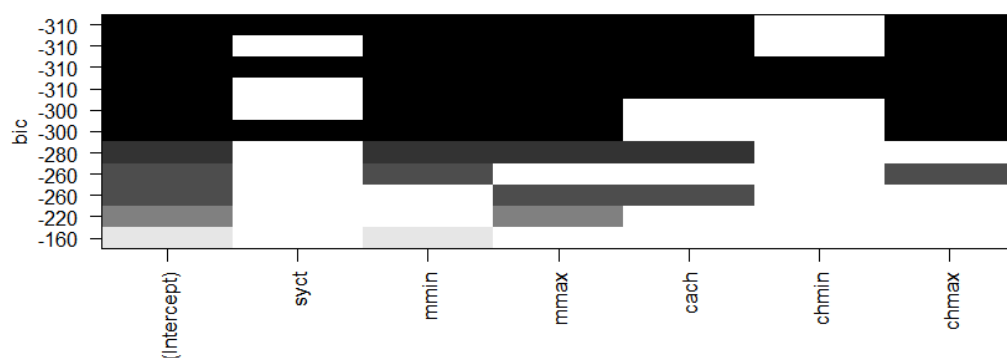
将数据写入 csv 表格中，读取 distance 和 velocity 两项数据，制作一个散点图，以距离为 Y 轴，以衰退速度为 X 轴。两者大致成正相关。在散点图中添加一条简单的线性回归线。



按照宇宙大爆炸理论，所有的星云都是从单一爆炸点向外逃逸不存在初始距离，截距应该 0；星云逃逸的时间即为宇宙的年龄，即星云逃逸距离（Distance）/速率（Velocity）。数据的获取过程会产生误差，好的测量方法能够获得更精确的数据，通过这些数据建立更接近真实值的模型估算宇宙的年龄。

PS3_6

写入 cpus 数据, 使用 sample 函数将其划分为 80% 的训练集和 20% 的测试集, 使用 regsubsets 函数进行全子集回归, 绘制出 bic 图, $sycl + mmin + mmax + cach + chmax$ 作为影响因子时, bic 结果最小, 将他们作为影响因子用于回归拟合



没有变量 chmin 时 bic 值最小, 可以得到最佳回归模型。将最佳回归模型应用于测试集, 计算出测试值和真实值之间的偏差:

```
> mean(perf_predict) - mean(cpus_test$mean_perf)
[1] -14.87073
```

PS3_7

福布斯关于阿尔卑斯山沸点的数据, 一个数据框架, 包含 17 个水的沸点和以英寸汞柱为单

位的大气压观测值。对阿尔卑斯山大气压和沸点相关关系做 t.test

```
> t.test(forbes$pres, forbes$bp)

welch Two Sample t-test

data: forbes$pres and forbes$bp
t = -112.78, df = 24.18, p-value < 2.2e-16
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -181.1483 -174.6400
sample estimates:
mean of x mean of y
 25.05882 202.95294
```

用 ANOVA 和线性拟合分析的船舶事故的发生和船舶的种类之间的关系

```
> anova_one_way <- aov(incidents ~ type, data = SH)
> summary(anova_one_way)

            Df Sum Sq Mean Sq F value    Pr(>F)    
type          4    5235   1308.8    13.11 1.26e-06 ***
Residuals    35    3494     99.8              
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> |
> summary(model2)

call:
lm(formula = incidents ~ type + year + period + type:year, data = ships)

Residuals:
    Min       1Q   Median       3Q      Max
-19.950  -2.562  -0.075   1.712   23.200

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -46.0500    39.4232  -1.168  0.252283    
typeB        198.5000    52.9098   3.752  0.000782 ***
typeC         27.3000    52.9098   0.516  0.609786    
typeD         13.7500    52.9098   0.260  0.796796    
typeE         28.4500    52.9098   0.538  0.594882    
year           0.5000     0.5524   0.905  0.372827    
period        0.2600     0.1841   1.412  0.168566    
typeB:year    -2.5500     0.7812  -3.264  0.002814 **
typeC:year    -0.4600     0.7812  -0.589  0.560517    
typeD:year    -0.2500     0.7812  -0.320  0.751239    
typeE:year    -0.4400     0.7812  -0.563  0.577591    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 8.734 on 29 degrees of freedom
Multiple R-squared:  0.7466,    Adjusted R-squared:  0.6592 
F-statistic: 8.544 on 10 and 29 DF,  p-value: 2.655e-06
```

结果证明船舶事故的发生和船舶的种类存在关系