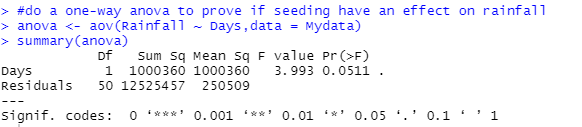
PS3\_1：

* 1. 将两组数据写入csv表格中，读取数据并作出Rainfall相对于Days的箱线图。给箱线图命名并调整颜色：
  2. 对数据进行anova分析：

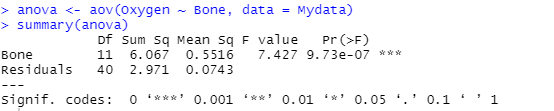


从anova的结果来看，Pr(>F) > 0.05，可以认为seeding对于rainfall没有明显影响。

PS3\_2：

2.1 采用anova分析方法对多组数据进行均值检验。首先将数据写入csv表格中，

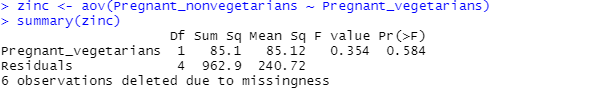
再进行anova方差分析，结果：



从anova的结果来看，Pr(>F) < 0.05，可以认为霸王龙的骨骼样品含氧量存在差异，霸王龙并非是温血动物。

PS3\_3：

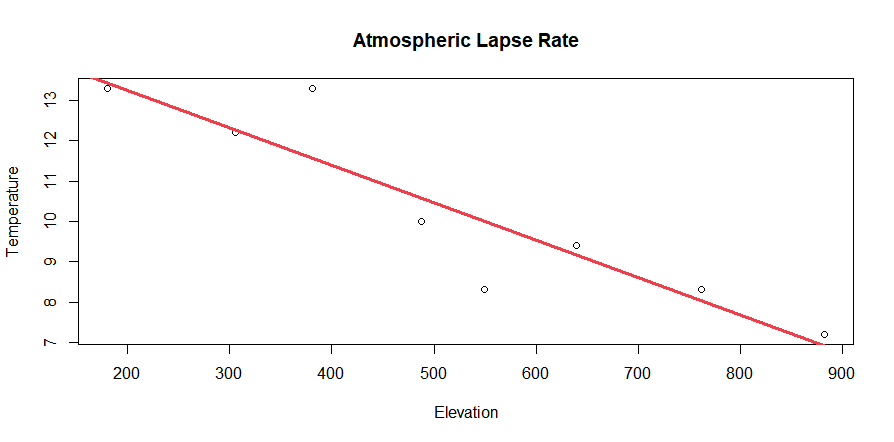
3.1 采用anova分析方法对Pregnant\_nonvegetarians，Pregnant\_vegetarians两组数据进行方差分析，结果：

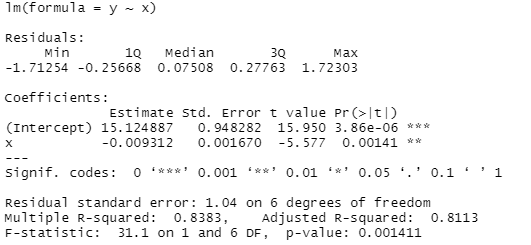


从anova的结果来看，Pr(>F) > 0.05，可以认为怀孕非素食主义者和怀孕素食主义者的女性体内Zn含量无明显差异。

PS3\_4：

4.1 将数据写入csv表格中，读取数据并plot出Temperature相对于Elevation的图：



4.2 对这两者进行线性回归。回归结果：

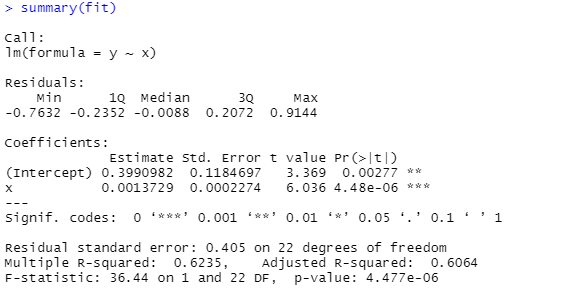
可得到回归方程：y = 15.12-0.0093x

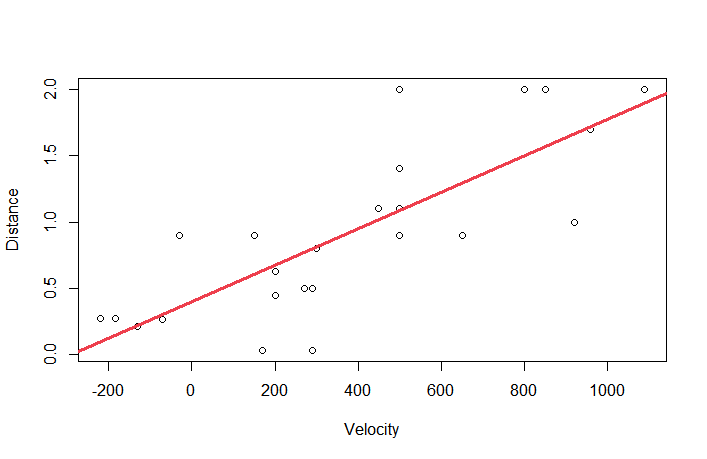
可看出lapse rate约为9.3m/℃，并非为9.8 m/℃.

PS3\_5：

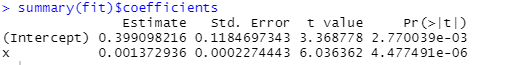
5.1 将数据写入csv表格中，读取数据并plot出distance相对于velocity的函数。发现distace和vecocity大致成正相关；

5.2对distace和vecocity进行线性回归并做出线性回归图：





5.3 假设一：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，故不存在初始距离，截距应该为0；假设二：按照宇宙大爆炸理论，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）。

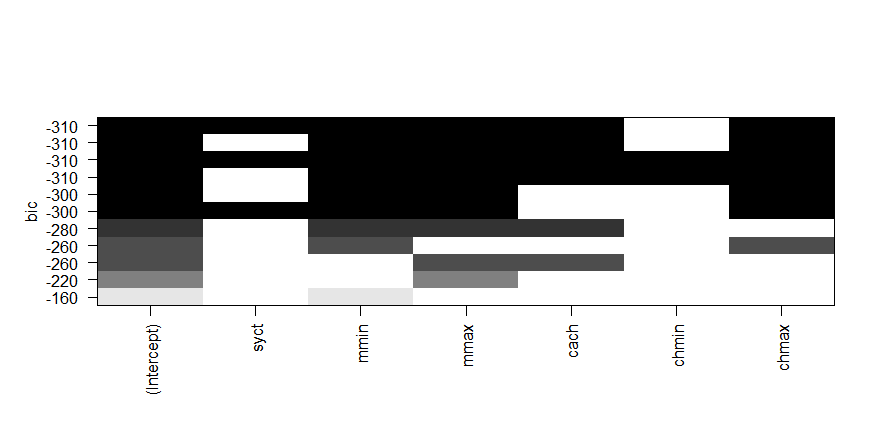


宇宙的年龄：根据以上假设，可计算宇宙的年龄：(30.9\*10^6\*10^12\*0.0013729)/(60\*60\*24\*365) = ~1.344\*10^9年。

5.4 更好的测量方法能够测量更精确更多的数据，可以通过这些数据建立更接近真实值的模型，从而能更好地估算回归系数，更好地估算宇宙的年龄。

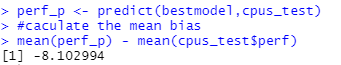
PS3\_6：

6.1 依题意，使用regsubsets函数进行全子集回归，绘制出bic图：



由此图可以看出，没有chmin这个变量时，bic值最小，故得到最佳回归模型

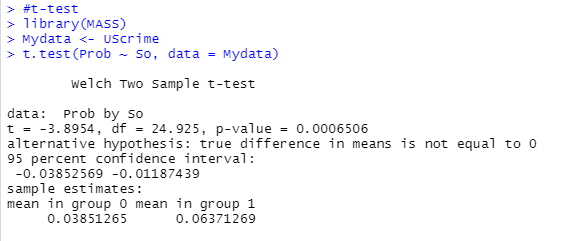
6.2 将最佳回归模型应用于测试集，计算出测试值和真实值之间的偏差：



PS3\_7：

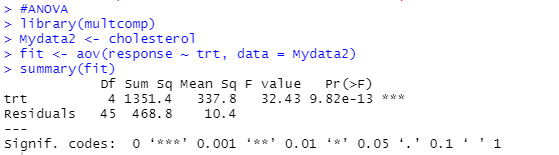
由于课题组没有合适的数据，我找了R语言库里的数据做以下测试，参考《R语言实战》：

7.1 使用MASS包中的UScrime数据集。定义的问题为：美国南方和非南方（so）的监禁概率（prob）是否相同？

****

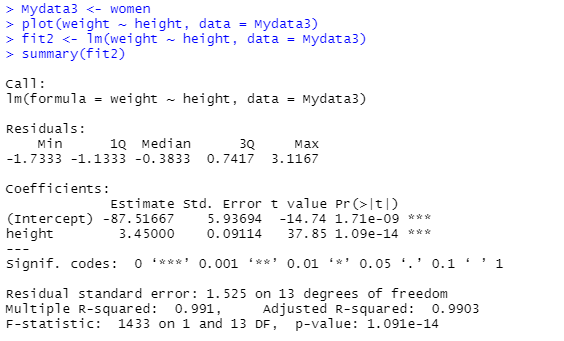
对监禁概率（Prob）相对于该州是否位于南方（So）做t-teat。从结果可以看出，p-value<0.01，可认为南方各州和非南方各州不具有相同监禁概率。

7.2 使用mulcomp包中的cholesterol数据。定义的问题为：不同的疗法（trt）对降低胆固醇的效果（response）如何？



对降低胆固醇的效果（response）相对于不同的疗法（trt）做anova分析。从结果可以看出，p<0.0001，F检验非常显著，说明五种疗法的效果不同。

7.3 使用基础安装中的数据集women，数据集中为15个成年女性的身高及体重信息。定义的问题为：建立成年女性身高和体重的线性关系？



从结果可以得出该样本身高和体重的线性关系为：

Weight = -87.52 + 3.54 \* height

做出拟合图：

