Assignment 3 Report

12032373 徐浩翔

1.1 A significant exceeding was observed in seeding group compared to the un-seeding group.

1.2 P-value of Kolmogorov-Smirnov test, 0.01905, is smaller than 0.05, which means the two samples are not from one same distribution. So the seeding makes a different which is 277.4 by taken the first order origin moment estimation.

2. P-value of ANOVA, 1.01e-06, is smaller than 0.05. So the temperature of bones are not the same, which means several bones were formed under different temperature. T-rex is a cold-blooded dinosaur.

3. P-value of t-test (0.8812) and ANOVA (0.925) both indicate the diet of vegetarian has nothing related to zinc concentration when pregnant.

4. The lapse rate here is 9.3 ℃/km.

5.1 There seems to be a positive relationship between the two parameters.

5.3

Intercept: 0.3990982

Slope: 0.0013729

If Hubble's *Big Bang Theory* is correct, the intercept need to be zero because zero shift should be observed at zero distance. The slope means distance per unit shifting velocity. Assume we have two nebulas on the regression line, with coordinate:

$$(v_1, l_1) \text{ and } (v_2, l_2).$$

According to the Hubble's *Big Bang Theory*, the two nebulas were at almost the same position at the beginning, which means the relative distance was negligible. After a certain time they separated from each other. If the two nebulas are moving along the same direction away from the same center, this certain time can be calculate from

$$time = \frac{Distance\ difference}{Relative\ velocity\ difference}$$

And can be written as

$$time = \left|\frac{l_1 - l_2}{v_1 - v_2}\right|$$

This is precisely the slope of the regression line. And the time that separated the two nebulas from the initial position to the current condition is just the age of the universe, because the initial condition is a singularity. Thus, the slope is the age of the universe.

The age of the universe is

$$slope = 0.0013729 \times \frac{10^6 \times 30.9 \times 10^{12} km}{km/s} = 1.35 \times 10^9\ years$$

This is 10 times smaller than the age of universe we known as 1.38×10^{10} years. And the Hubble constant ($67.8\ km \cdot s^{-1}/Mpc$), or saying the reciprocal slope, is far away from the regression result ($\frac{1}{slope} = 728.4\ km \cdot s^{-1}/Mpc$). This indicate the data probably has some error.

5.4 The largest velocity is nearly 15 times larger than the smallest one, while this relationship for distance is larger than 60 times. This indicate the distance has a larger range and it contribute more to the regression slope. Thus an accurate measurement would lead to a more-close result of regression line.

6.1

$$pref = 0.059syct + 0.017mmin + 0.0055mmax + 0.46cach - 1.21chmin + 1.95chmax - 64.11$$

6.2

Mean bias: -4.842515

Correlation coefficient: 0.8341233

7.1

Data was from two AQ stations of PM_{10} in μg/m^3. Due to the natural PM concentration is not follow the standard normal distribution, I didn't directly apply the PM_{10} concentration to t-test. Instead I use a mean value to do the t-test. One station was selected here and sample randomly for 1000 times and stored their mean as one value. Repeated this process for 2000 times to get two group of 1000 values. Based on the statistical theory, mean value of a sample should follow normal distribution and convergence to the real average in probability. A t-test was applied to test if the two group were from the same origin (Of course they are from the same distribution).

P-value of several test (because sample might change the values, I had ran this script more than once) all larger than 0.05, which means they are from the same distribution.

7.2

The two stations are CN_1365A Meisha (梅沙) Station and CN_1360A Yantian (盐田) Station, both in Shenzhen. An ANOVA test was applied whether a distance of stations would cause a different in AQ concentration within a city in years.

Result: p-value <2e-16, which means different AQ station have different concentration variation. One station cannot represent air quality of the entire city.

7.3

Data file '*PS3_7_data3.csv*' (unit: μg/m^3) was from two PM_{10} sensors measured the same place within the same time period. One is already calibrated while the other need calibration. A linear function was adapted to calibrate the un-calibrated sensor. This require a linear regression. A good R^2 indicate the un-calibrated sensor has a stable working condition. On the contrast, we should doubt the stability of the sensor if we received a not-good R^2 value.

Result: $calibrated\_result = sensor\_result \times 0.45587 + 2.62609$

$R^2 = 0.2224$

This sensor is not stable.