

**Ethics Pledge**

**Consistent with the above statements, all homework exercises, tests and exams that are designated as individual assignments MUST contain the following signed statement before they can be accepted for grading.**

I pledge on my honor that I have not given or received any unauthorized assistance on this assignment/examination. I further pledge that I have not copied any material from a book, article, the Internet or any other source except where I have expressly cited the source.

Signature: Haodong Zhao Date: Mar 5th 2019

Please note that assignments in this class may be submitted to www.turnitin.com, a web- based anti-plagiarism system, for an evaluation of their originality.

1. **You are conducting a study to see if students do better when they study all at once or in intervals. One group of 12 participants took a test after studying for one hour continuously. The other group of 12 participants took a test after studying for three twenty minutes sessions. The first group had a mean score of 75 and a variance of 120. The second group had a mean score of 86 and a variance of 100. Assuming the normal populations, independent samples, and equal population variances conditions hold, are the mean test scores of these two groups significantly different at the 0.05 level?**

**Answer：**

α = 0.05

n1 = 12, mean1 = 75, s12 = 120

n2 = 12, mean2 = 86, s22 = 100

– = -11,

H0 :

H1 :

Since it’s normal distribution and independent samples, and equal population variances conditions, we use Pooled-Variance t test.

= 110

= -2.569

And the d.f. = (12 + 12 – 2) = 22

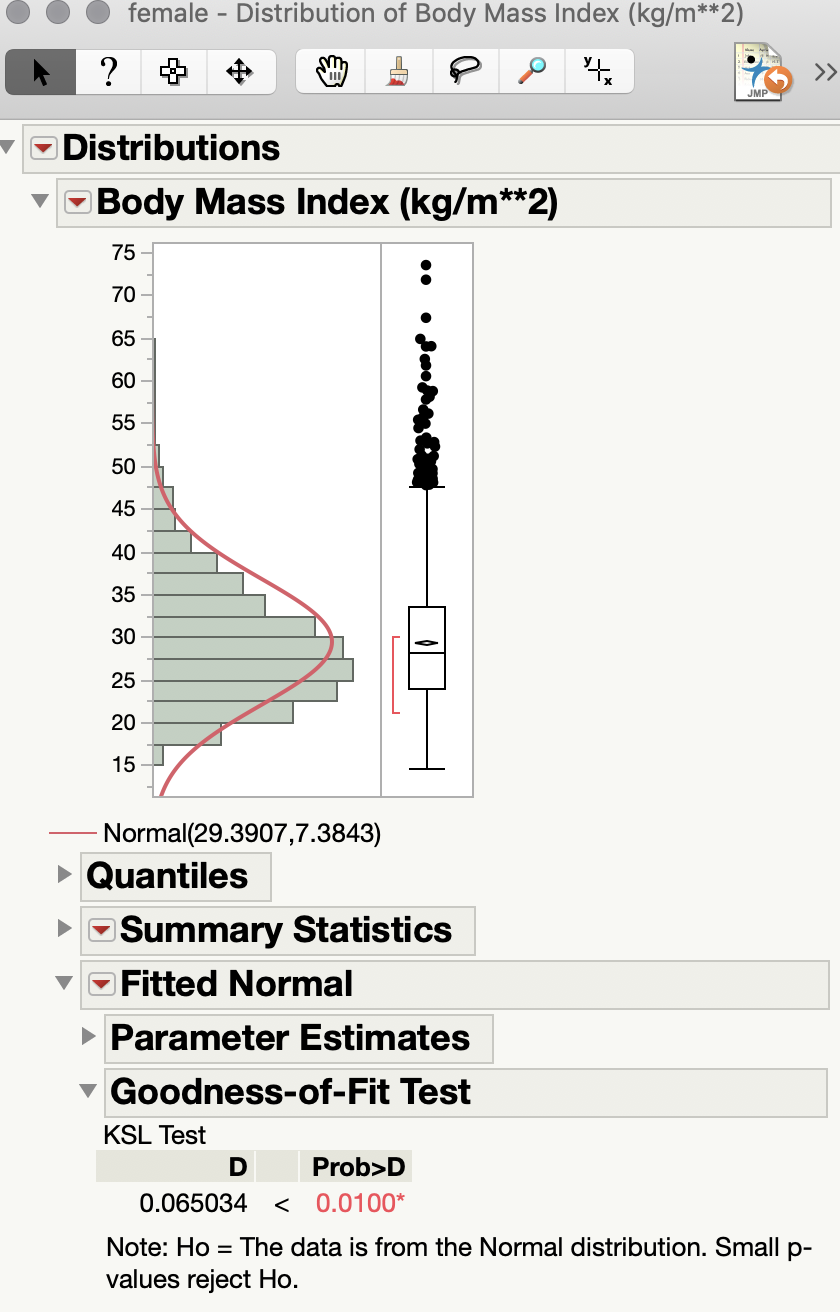
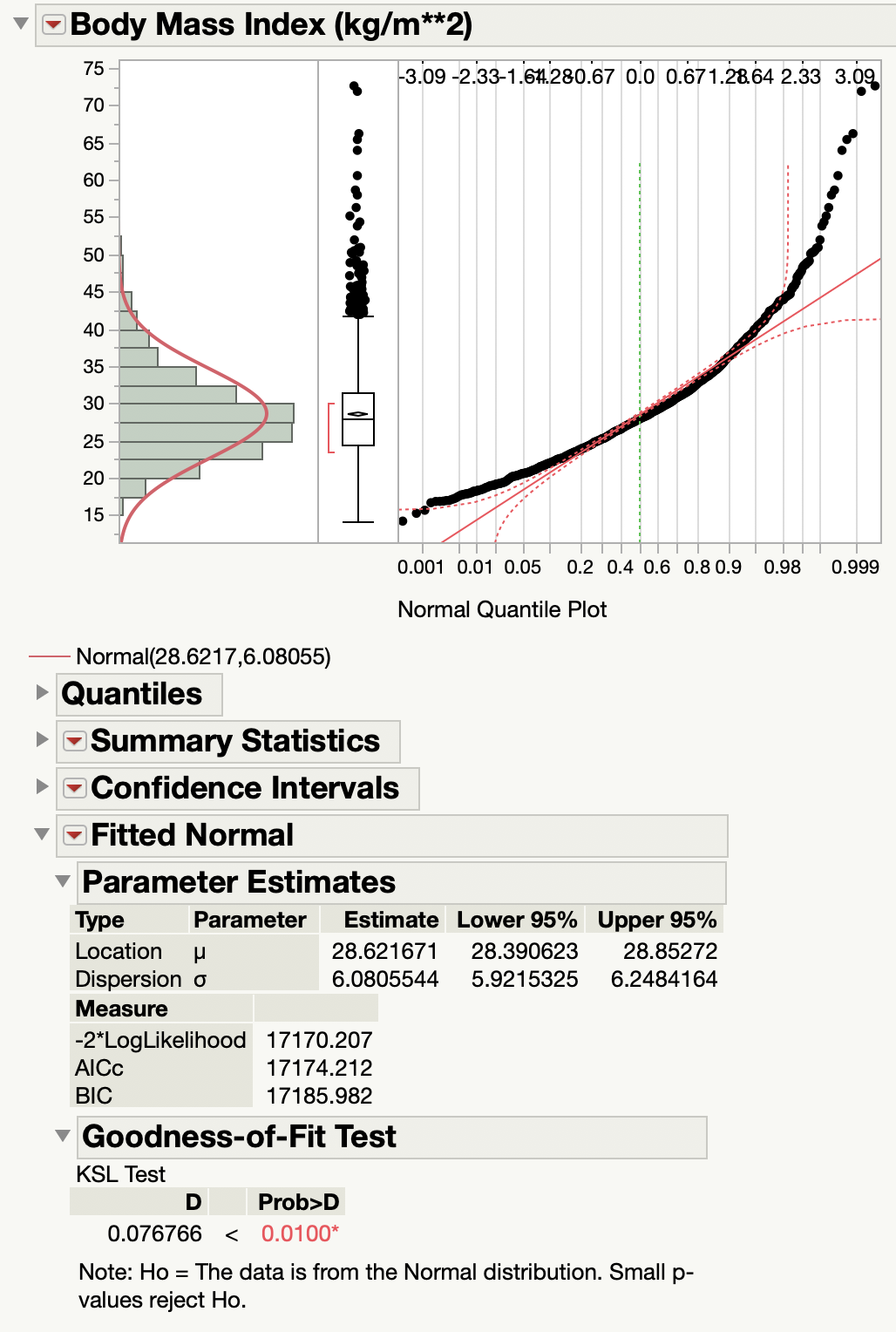
For α = 0.05 and d.f. = 22, the t critical = 2.0739

Since = -2.569 < -2.0739 = t critical, we reject H0, the mean test scores of these two groups are significantly different at 0.05 level.

1. **Find a file NHANES.dat. It contains a large data set related to patients’ profile and their health status. Suppose we are interested in whether BMI (Body Mass Index) is different between Male and Female groups. Is there an evidence of equal variance at significance level 0.05? Perform an appropriate two-sample t-test.**

**Answer:**

α = 0.05

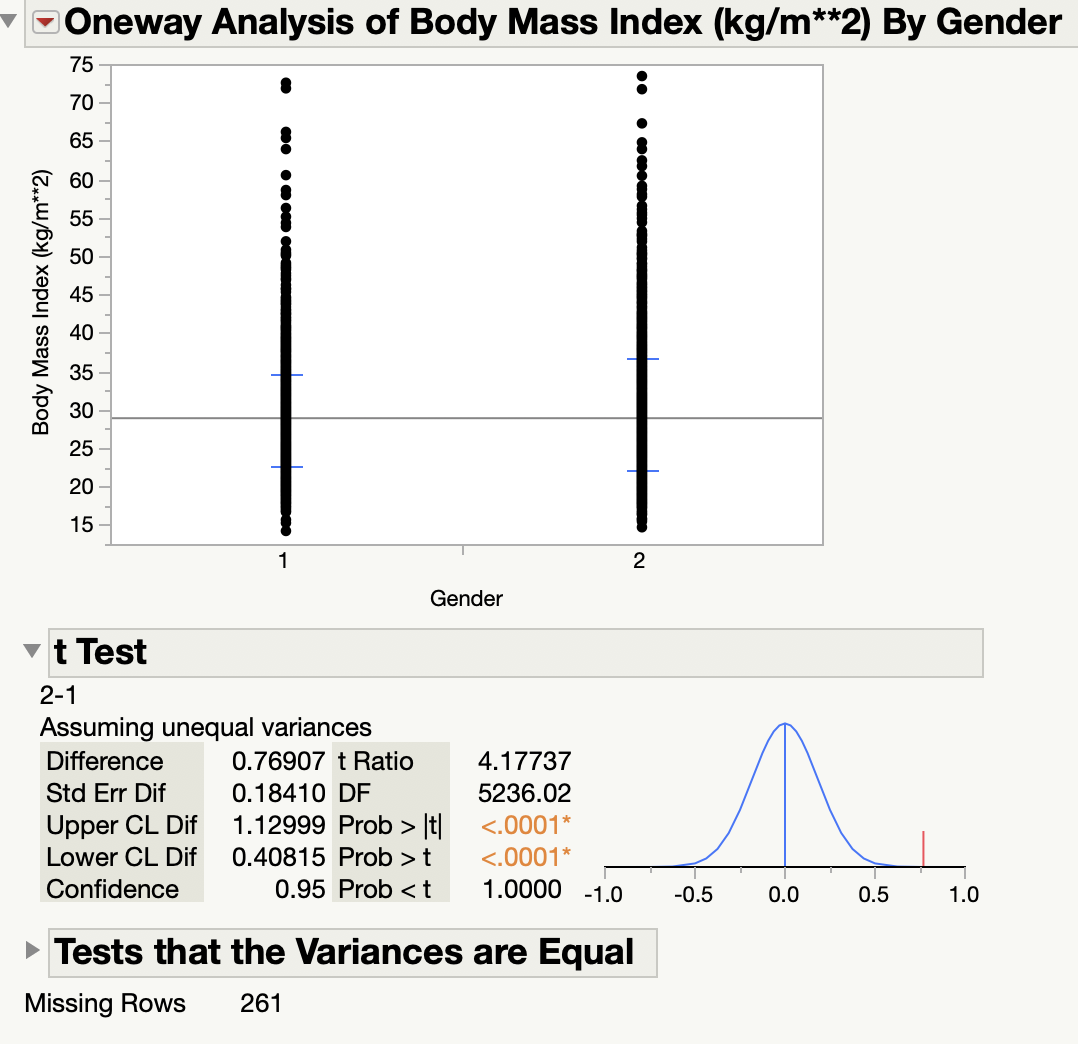
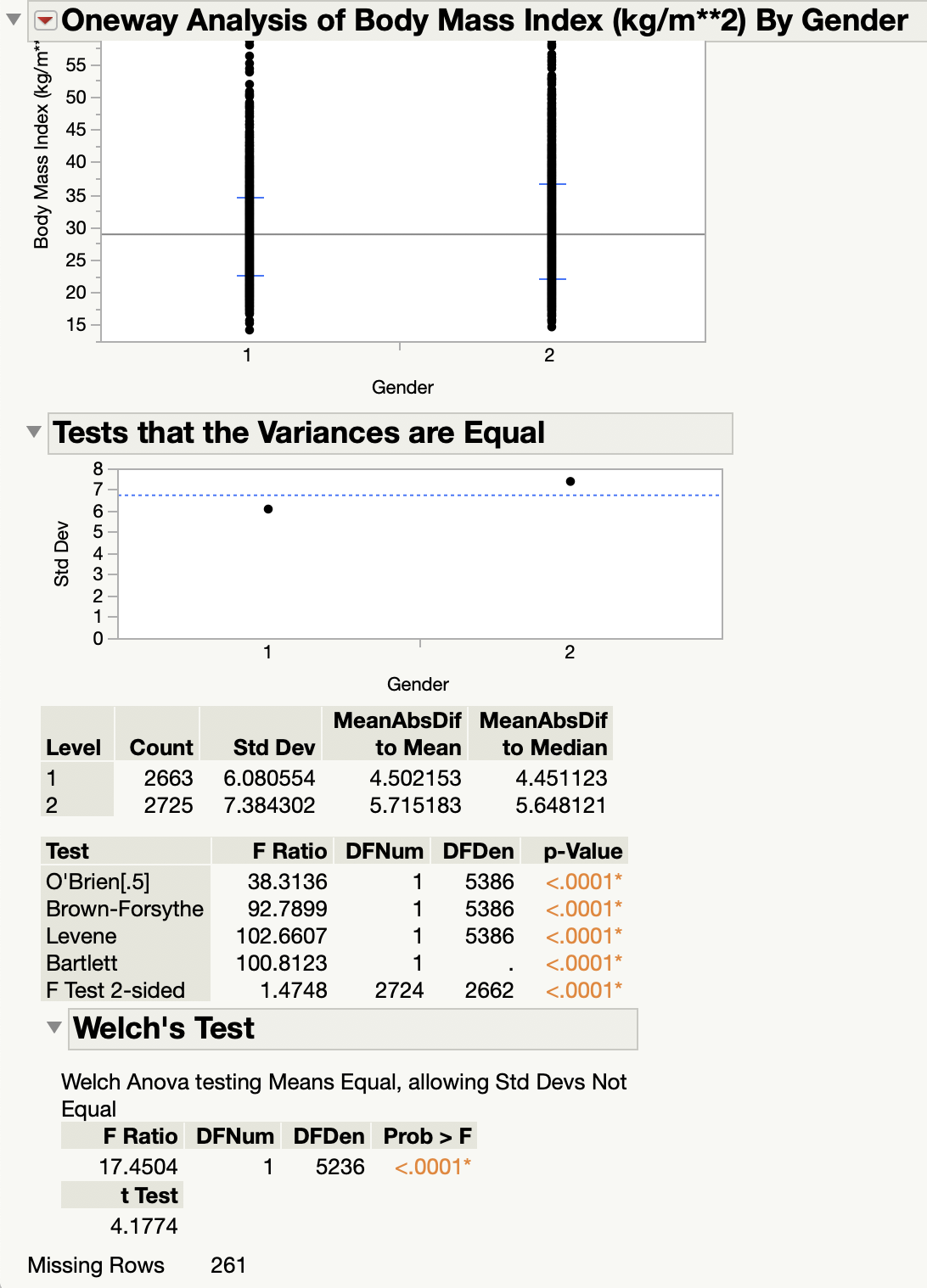


In the NHANES dataset, we assume the gender 1 is male and gender 2 is female.

As we can see, by using Goodness-of-Fit Test, the BMI distribution of male and female are not normal distributed. Therefore, we cannot use F-test. We have to use Levene’s Test.

H0 :

H1 :



Levene’s test Two sample t-test

From the Levene’s test, the p-Value is less than 0.0001 < 0.05. Therefore, reject H0, we have significant evidence to prove that the variances are different between male’s and female’s BMI at significance level 0.05. And by using the same hypothesis to try two sample t-test, the p-value is also less than 0.0001 < 0.05. Then also reject H0, there are significant evidence to prove the variances are different at 95% confidence interval.

1. **The female cuckoo lays her eggs into the nests of foster parents. The foster parents are usually deceived, probably because of the similarity in the sizes of the eggs. Lengths of cuckoo eggs (in millimeters) found in the nests of hedge sparrows, robins, and wrens are shown below:**

**Hedge sparrow: 22.0, 23.9, 20.9, 23.8, 25.0, 24.0, 21.7, 23.8, 22.8, 23.1, 23.1, 23.5, 23.0, 23.0**

**Robin: 21.8, 23.0, 23.3, 22.4, 23.0, 23.0, 23.0, 22.4, 23.9, 22.3, 22.0, 22.6, 22.0, 22.1, 21.1, 23.0**

**Wren: 19.8, 22.1, 21.5, 20.9, 22.0, 21.0, 22.3, 21.0, 20.3, 20.9, 22.0, 20.0, 20.8, 21.2, 21.0**

**It is believed that the size of the egg influences the female cuckoo in her selection of the foster parents. Do the data support this hypothesis? Test whether or not the mean lengths of cuckoo eggs found in the nests of the three foster-parent species are the same. (Here, don’t forget to check each of the underlying ‘assumptions.’)**

**Answer:**

To test whether the mean lengths of eggs found are same or not, we’ll use ANOVA.

Assume α = 0.05

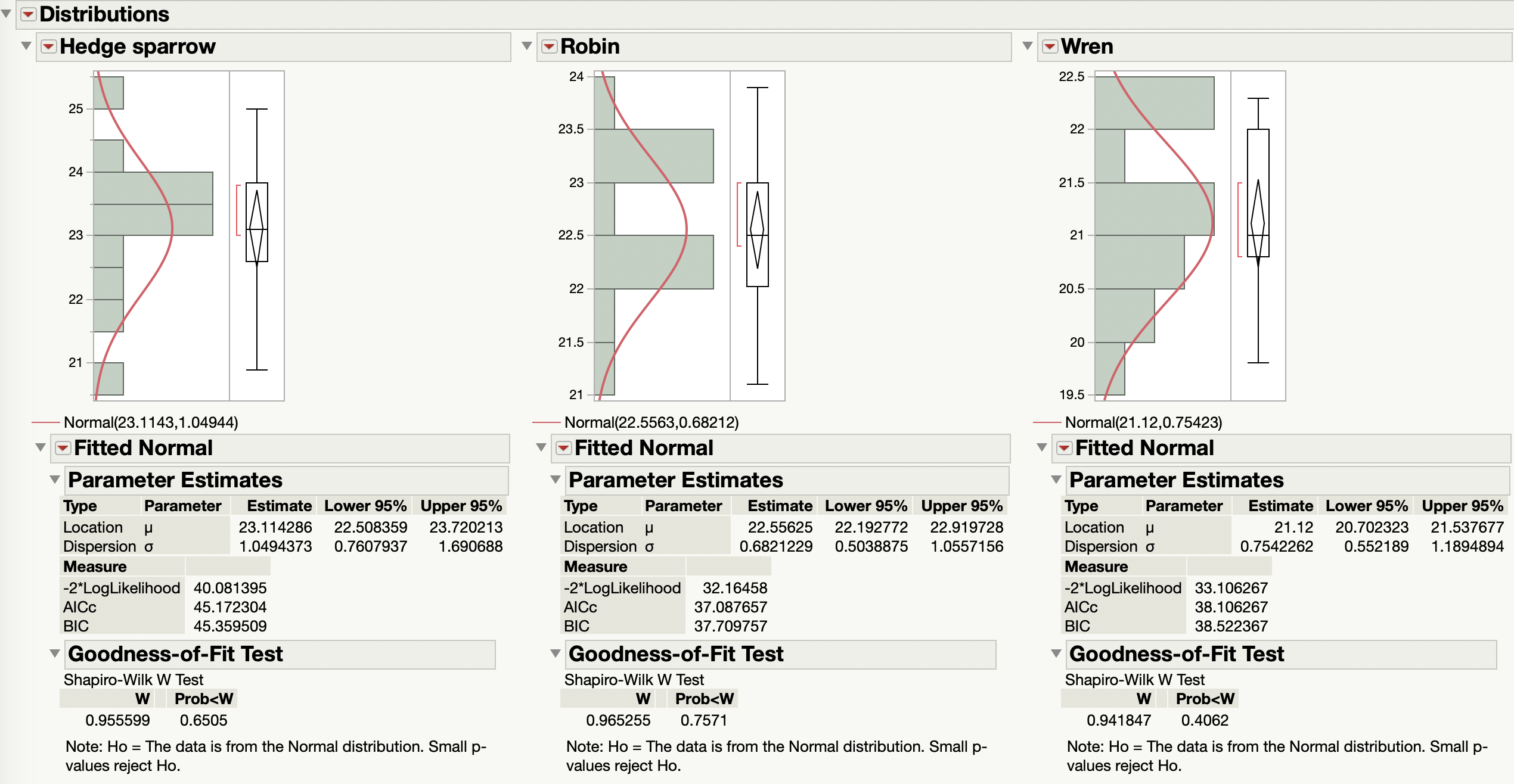
The underlying ‘assumptions’ of ANOVA are:

* 1. Normally distributed
  2. Equal variances

Testing for normal distribution:

H0 : The data are normally distributed

H1 : The data are not normally distributed



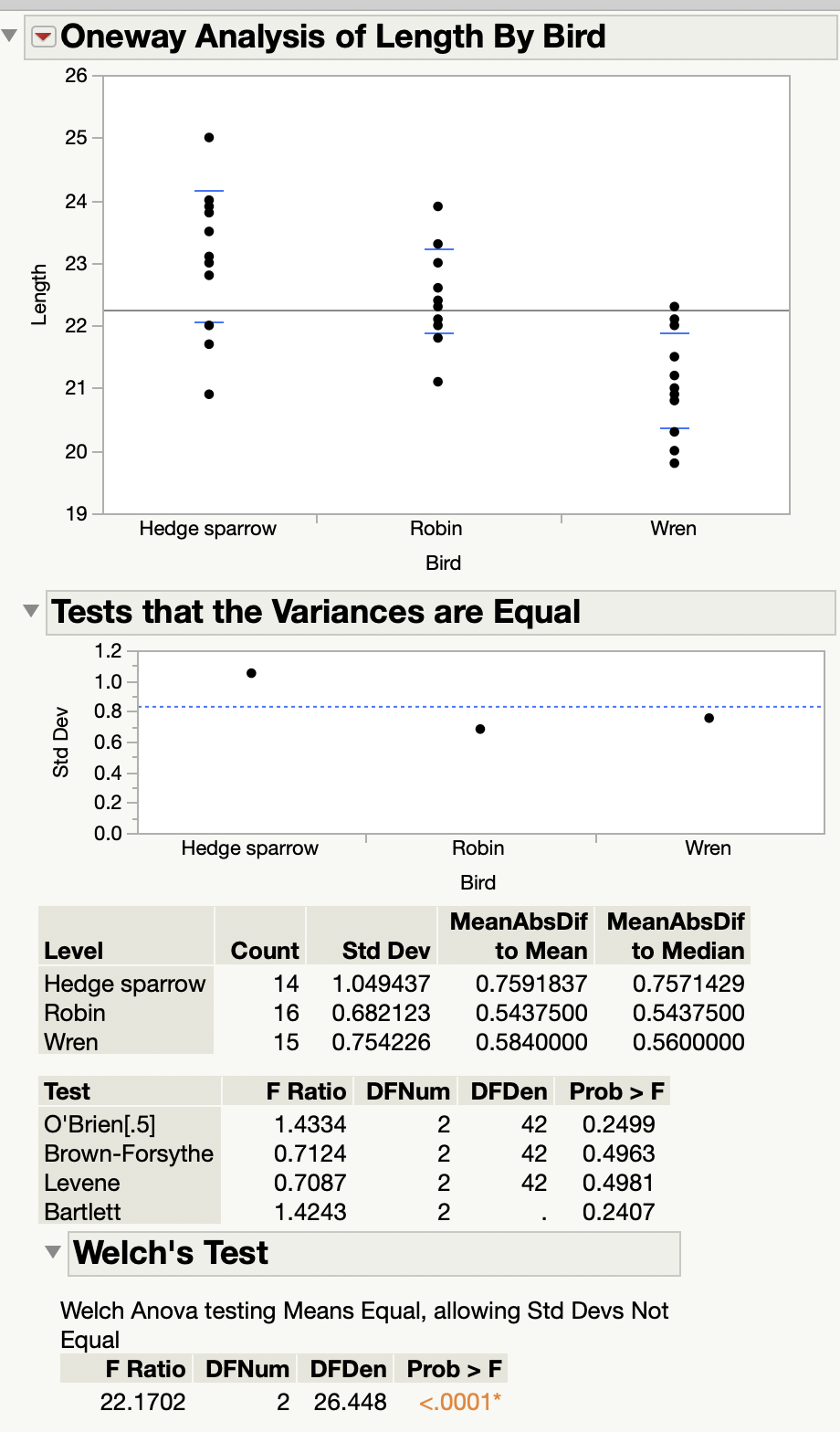
Test whether or not the eggs lengths are normally distributed by using Goodness-of-Fit Test, we can find all the p-value are greater than 0.05, we cannot reject the hypothesis in 95% confidence interval, so lengths of cuckoo eggs found in the nests of the three foster-parent species are normally distributed.

Testing for equal variance:

H0 : All the variance are equal

H1 : Not all the variance are equal

We can use Levene’s test to test whether the variances are equal or not:

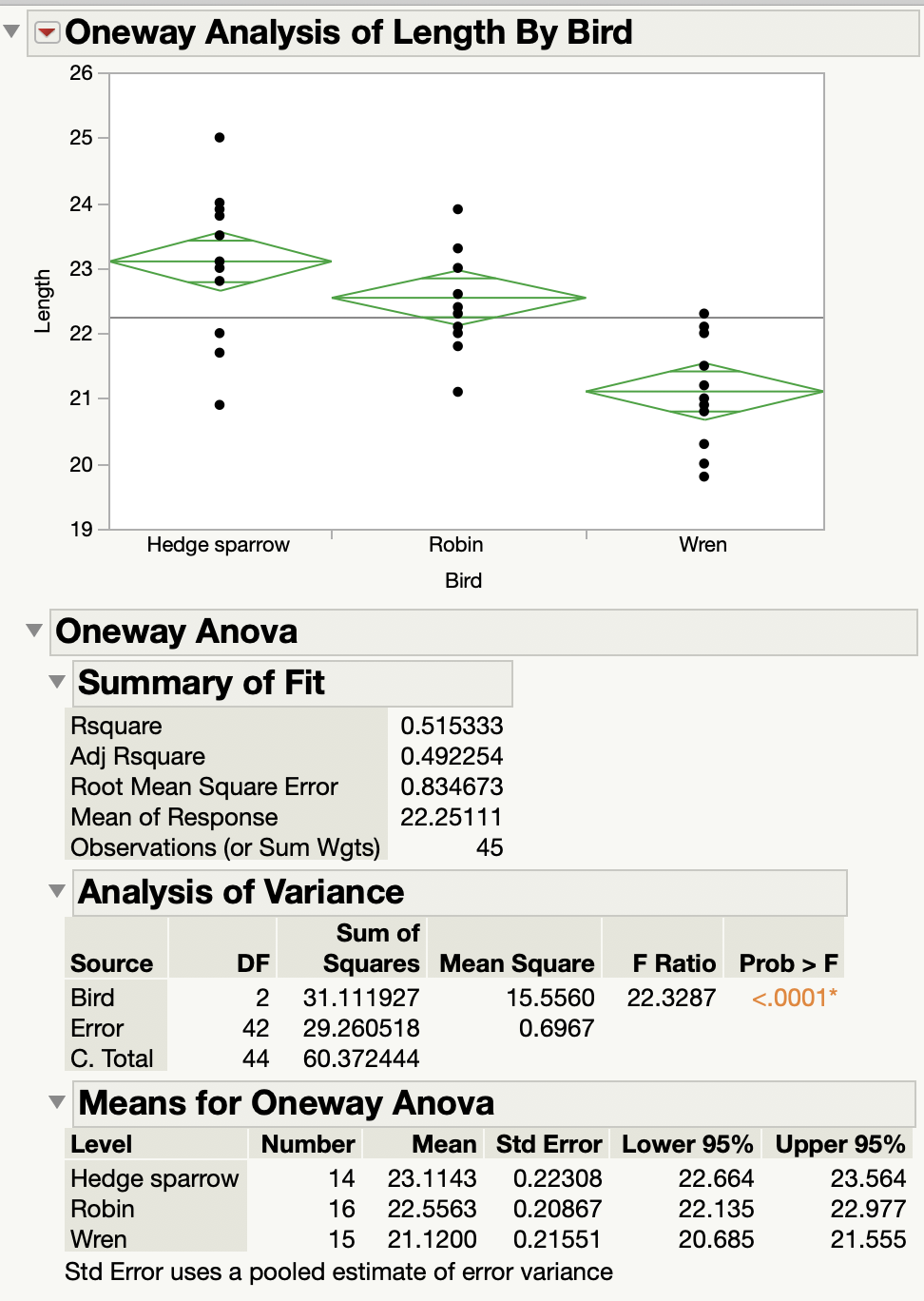


From Levene’s test, we can find the p-value are greater than 0.05, so we cannot reject the hypothesis at 95% confidence interval. Therefore, the variances are equal.

Then we use one-way ANOVA to test whether the mean lengths of eggs found are same or not.

H0 : All the mean of lengths are same

H1 : Not all the mean of lengths are same



From this ANOVA test, we can find the p-value is less than 0.0001 < 0.05, so we can reject the hypothesis in 95% confidence interval. Therefore, the mean lengths of cuckoo eggs found in the nests of the three foster-parent species are not same. The data support following hypothesis: the size of the egg influences the female cuckoo in her selection of the foster-parents.

1. **Recall the class note example on three promotions and sales volume difference percentage data; see Lecture 4-2.pdf lecture slide, page 7. Perform one-way ANOVA procedure to test**

**H0 :μ1 =μ2 =μ3, H1 : Not H0.**

**(a) Carry this out via hand calculations (it is worth doing this practice at least once!).**

|  |  |  |  |
| --- | --- | --- | --- |
|  | 1 | 2 | 3 |
| Sales vol. diff. % | 9.5 | 8.5 | 7.7 |
|  | 3.2 | 9.0 | 11.3 |
|  | 4.7 | 7.9 | 9.7 |
|  | 7.5 | 5.0 | 11.5 |
|  | 8.3 | 3.2 | 12.4 |

Assume α = 0.05

n1 = 5, n2 = 5, n3 = 5, c = 3, n = 15

mean1 =6.64, mean2 = 6.72, mean3 = 10.52, mean\_total = 7.96

variance1 = 6.82, variance2 = 6.28, variance3 = 3.43

H0 :μ1 =μ2 =μ3

H1 : Not H0

SST = SSB + SSW

SSB = 5 \* (6.64 – 7.96)2 + 5 \* (6.72 – 7.96)2 + 5 \* (10.52 – 7.96)2

= 8.712 + 7.688 + 32.768 = 49.168

MSB = SSB / (c – 1) = 49.168 / 2 = 24.584

SSW = (9.5 – 6.64)2 + (3.2 – 6.64)2 + (4.7 – 6.64)2 + (7.5 – 6.64)2 + (8.3 – 6.64)2 +

(8.5 – 6.72)2 + (9.0 – 6.72)2 + (7.9 – 6.72)2 + (5.0 – 6.72)2 + (3.2 – 6.72)2 +

(7.7 – 10.52)2 + (11.3 – 10.52)2 + (9.7 – 10.52)2 + (11.5 – 10.52)2 + (12.4 – 10.52)2

= 27.272+ 25.108 + 13.728 = 66.108

MSW = SSW / (n – c) = 66.108 / 12 = 5.509

FSTAT = MSB / MSW = 4.463

df1 = c – 1 = 2

df2 = n – c = 12

From F table, when df1 = 2, df2 = 12 and α = 0.05, Fα = 3.89. Since FSTAT = 4.463 > Fα = 3.89, reject H0 at 95% confidence interval, not all the mean values are same.

**(b) Now, do it again via statistical software package and compare with the result in (a). Remarks. We recall some logic behind ANOVA and also follow-up on the in-class questions.**

* 1. **A One-Way Analysis of Variance is a way to test the equality of three or more means at one time by using variances. There are the between group variation (SSB) (not variance yet) and the within group variation (SSW). The whole idea behind the ANOVA is to compare the ratio of between group variance (that is, variation/degrees of freedom, similar to sample variance formula s2) to within group variance (SSW/degrees of freedom).**
  2. **I agree to some extent that the term ‘between-group’ variation is a misnomer. This is because SSB does not measure something like (Xi − Xj)2, which seems to be more consistent with the term ‘between-group’ variation. However, notice that under the null H0, all means are the same (μ1 = μ2 = μ3). To test this equality of three (or more) means simultaneously, one wants to measure an aggregated variation caused by each group effect (that is, deviation from the overall mean); if this number is large enough, then it is an evidence that goes against H0.**
  3. **If the variance caused by the between group effect (that is, due to Xi − X(overall mean), i = 1,...,c) is much larger when compared to the variance that appears within each group, then it is likely because the means aren’t the same.**

**Answer:**

H0 :μ1 =μ2 =μ3

H1 : Not H0

Use one-way ANOVA test:



From ANOVA test, the p-value is 0.0356 < 0.05, so reject H0 at 95% confidence interval, not all mean values are same.