1. Although the mean for two groups appears drastically different: Ctl – 18.8 and Ex – 5.7, the p-values indicated by T-test is 0.14 and not statistically significant to reject the null hypothesis, which is the true means of these two groups are the same.

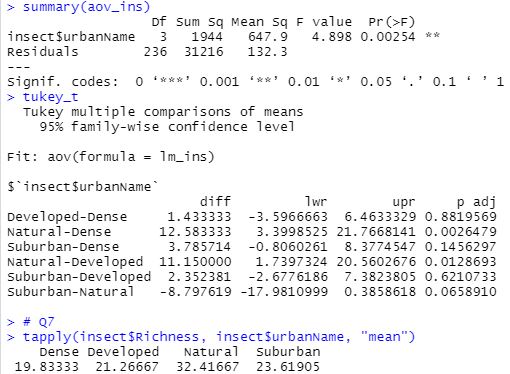
2. A one sided T-test’s null hypothesis is that the sample mean is the same as the true mean. With only one group of data, the function needs a given true mean, as shown in the example: t.test(senior\_data$age, mu = 22).

3. H0 is that the true means for species richness across all types of urban area are the same. Ha is that the true means for species richness among urban areas are the different.

4. First we assume that the data are all collected independently, and all factors are independent from each other. Then, I ran the Shapiro test on each of the urbanType to see if all groups of data have a normal distribution, here are the resulting P-values for type1, 3, 8, 9 respectively: 0.25, 0.25, 0.06, 0.10. If we are using 0.05 as a hard cut-off point for p-statistical significance, then all the tests fail to reject the null hypothesis that the data are distributed normally. Lastly, the Bartlett test is used to check on the variances, and with a P-value of 0.75, the test fails to reject the null hypothesis where the groups have similar variances . (For detailed codes and full results please refer to “Q4” section in R script.)

5. In my opinion, the F-ration would be higher as shown in graph A. For graph a, the distance between group sample means and the overall sample means are large, while within each group the data are relative closely distributed around the group mean, meaning the numerator (MS among groups) is large and the denominator (MS within groups) is low, resulting in a big fraction and high ratio. On the other hand, in graph B, the group sample means are close to the overall sample mean, while the data in each group are quite disperse. This would result in a small numerator (MS among groups) and a large denominator (MS within groups), thus the ratio here is small.

6. Natural – X; Suburban – Y; Developed – Y, Z; Dense – Z.

7.

With the given results above, it is safe to conclude that there are differences in insect species richness between different urban land surfaces. The general ANOVA’s P-value is statistically significant to reject the test’s null hypothesis that the true means for all groups are the same. If we look at the post hoc analysis, it is quite easy to observe the urban groups that differs greatly in terms of the insect richness. The sample-mean difference between natural and developed area has a 95% CI above 0 with an adjusted P-value of 0.01; similarly, for natural and dense area the difference is even greater above 0 and has an adjusted P-value of 0.002. The rest of the P-values are not statistically significant to reject the null hypothesis that the two groups in comparison has same mean. But looking closely, suburban-and-natural and suburban-and-dense are both very close to reject the null hypothesis, hinting that there might be an underlying difference.

8. Total declined = 18 + 8 = 26, Total Stable/Increase = 15 + 32 = 47. No legal impact = uniform distribution (0.5).

|  |  |  |
| --- | --- | --- |
| [**Expected**] | Not Protected | Protected |
| Declining | 26 \* 0.5 = 13 | 26 \* 0.5 = 13 |
| Stable/Increase | 47 \* 0.5 = 23.5 | 47 \* 0.5 = 23.5 |

9. The test result is: X-squared = 7.9642, df = 1, p-value = 0.004771.

With a P-value of 0.005, it is clear to see that the χ² value in above the accepted range, and thus we can reject the null hypothesis that legal protection has no impact on the species’ population status. The conclusion is that legal protection has impact on whether the species population is in decline or in stable/increase status.

10.

<https://github.com/guozhaosengzs/ENVDS/blob/master/activity3/activity3_script.R>