

Control t-test

Word count: 2052

1928613

1 Learning objectives

In this lecture, you will first be able to understand the concept of control t-test and how to use it, or in other words, when do you need to use the control t-test. Then, you will learn to apply it on R studio, and try to interpret the return result. Furthermore, you can comprehend how to choose the method in different backgrounds, especially in abnormal distribution or variance homogeneity it not proved.

2 Definition and usage

2.1 Definition

Control t-test is a kind of post-hoc test. After doing ANOVA, if the p-value is less than the significant level that was selected, we can reject the null hypothesis, in the other words, we have abundance adequate evidence to say that there is at least one group have different means with other groups. Thus the next step is compare each two group to find the difference, control t-test is a marvelous method that can be applied to this situation. The word ‘control’ in the title means this method will set a control group first, and then compare the control group with other groups to find which groups have significant difference with the control group.

2.2 Usage

In the definition, the purpose of the control t-test is to find out the specific level of difference between the two groups. However, in real life, the selection object of the control group is often the most common experimental group with relatively few variables, or the special group that the experimenter wants to compare. For example, people hope to compare the purification effects of several new reagents on industrial wastewater. If a group of sewage not applicable to any reagent or sewage using the older generation reagent is used as the control group, and the differences between other groups and the control group are compared, the results reflect whether other reagents have the effect of purifying sewage. If there is a significant difference and the index of sewage decreases, it indicates that the new reagent has played a role, which is helpful to purify sewage and protect the environment.

2.3 Basic Assumptions

Before using ANOVA, there must have at least two assumptions(Ruxton and Beauchamp, 2008): First, the sample should be in normal distribution. Second, the variance in each sample should be the same. On the premise of normal distribution, it can be determined that the difference between the two groups of samples is caused by mean and variance. Ensuring the homogeneity of variance further ensures that the difference between sample groups is caused by the difference of mean value. After doing ANOVA, we will have a basic

result that whether at least one group's mean is different from any of the other groups from compare the p-value with the significant level. When these two hypotheses are not tenable, nonparametric tests (described later) are required.

2.4 Why ANOVA

Why use ANOVA but not directly use t-test to compare the differences between each two groups. The most important reason is that the direct use of t-test to compare more than two groups of quantitative data will greatly increase the probability of the first type of error(Brown, 2008). We set the significance level of t-test as $\alpha = 0.05$, then the probability of making the first type of error in each pairwise t-test is $\alpha = 0.05$, the probability of not making mistakes is 0.95, which seems very low. However, when the number of samples is increased to 3, the calculation formula will be 0.95^3 , therefore, under the three comparisons, the probability of not making a kind of error is about 0.86, while the probability of making an error is about 0.14, which is obviously greater than the maximum significance level acceptable for the test. Further, when the number of samples increases to 5, the number of pairwise comparisons is 10, and the probability of the first type of error is 0.95^{10} , which is about 0.6. In other words, when we make 10 comparisons, we can believe that the number of results is only slightly higher than half, so the comparison of multiple samples is not suitable for direct use of t-test.

2.5 Two application conditions of control t-test

When we make a basic judgment on the data, there will be two situations: the first is that the sample data follows the normal distribution and the variance is homogeneous; The second case indicates that the data does not follow the normal distribution and the variance is not homogeneous. When the data follow the normal distribution and the variance is homogeneous, we need to use parameter test method. We use t-test to test the two groups of data, or use ANOVA to compare multiple groups of data. If there is significant difference, Dunnett t-test is used to compare data with the same sample size, while Bonferroni correction can be used to compare data with different sample sizes. In the case of non normal distribution or uneven variance, nonparametric methods need to be used. Welch test should replace ANOVA to test whether there is a significant difference in the mean value between groups(Hilton and Armstrong, 2006). Game Howell test was used instead of Dunnett test or Bonferroni test to compare the differences between the two sample groups(Hilton and Armstrong, 2006).The flow chart is made according to these paths to facilitate the judgment of the methods to be used in different situations(Appendix 1)

2.5.1 Bonferroni correction

Bonferroni correction is a very simple but effective method to adjust the threshold p-value to In the Mathematical formula, control the first type error.

$$P_b = P/k$$

P_b presents the corrected threshold, and k is the number of comparisons(Holland and Copenhaver, 1987). in this way, when there is 5 comparisons, the Error rate will be $1-(1-0.01)^5=0.049$, which is very close to the significant level, thus the Bonferroni correction make a well performance in ensuring the confidence to reject a null hypothesis.

2.5.2 Dunnett t-test

Another method for performing control t-test is Dunnett's test, which use Dunnett's critical value to limit the largest difference among groups(Koenig et al., 2008).

$$\text{Dunnett's critical value} = t_D \times \sqrt{2 \times MS_E/n}$$

t_D can be found in Dunnett's Table(ZACH,2020) (Appendix 2), through a significant level, the number of groups, and the sample size. MS_E is an abbreviation of Mean square error, it is the expected value of the square of the difference between the estimated value of the parameter and the true value of the parameter.

3 Case study 1

To investigate New York's air quality from May to September, the researchers recorded four consecutive indicators: ozone concentration (Roosevelt Island, unit: ppb), solar radiation frequency (Central Park, unit: angstrom), wind speed (LaGuardia Airport, unit: MPH) and temperature (LaGuardia Airport, Fahrenheit and Celsius). Finally, the data are summarized and processed(Chambers et al., 1983).

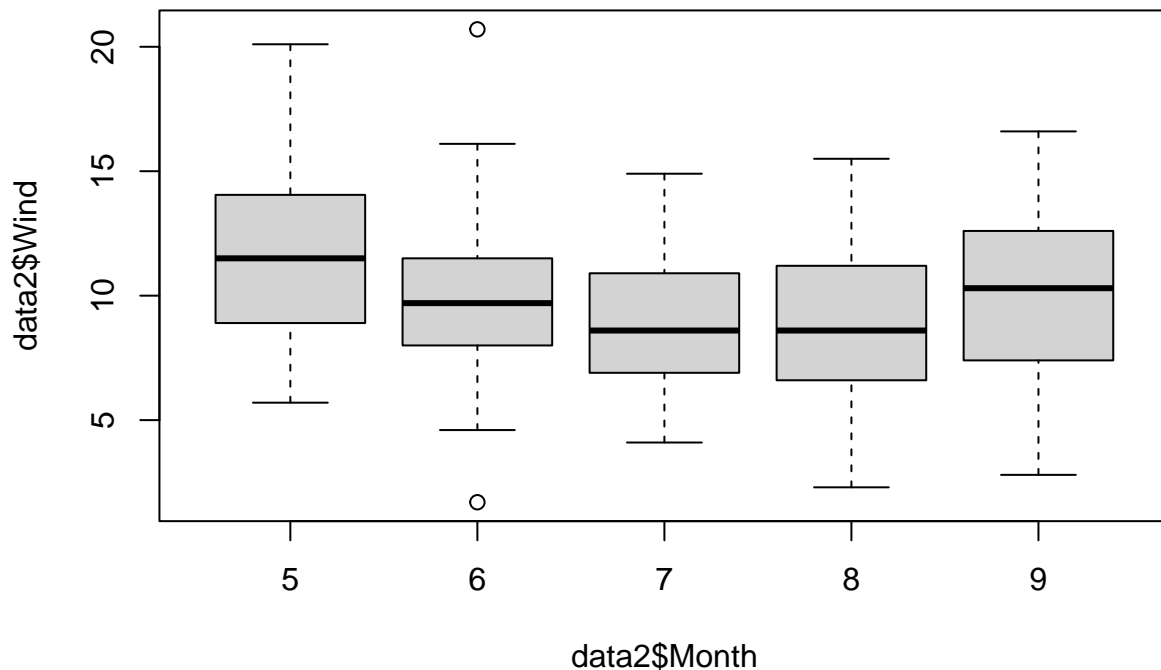
3.1 Scientific question

Compared with may, which months in New York state have obvious differences in wind speed from May.

3.2 Procedure

First, import data and visualize the data

```
data("airquality")
data2=airquality
data2$Month<-as.factor(data2$Month)
boxplot(data2$Wind~data2$Month)
```



From the box diagram, we find that the wind speed is relatively similar in several months, and the wind speed is the highest in May, but this does not allow us to determine whether the difference between them is significant, so we need to carry out the next analysis. Firstly, we test the normality of the data and the homogeneity of variance to ensure the accuracy of the results after data analysis.

```
shapiro.test(data2$Wind)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  data2$Wind
## W = 0.98575, p-value = 0.1178
```

```
#library(car) it contains the leveneTest function
leveneTest(Wind~Month,data=data2)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##      Df F value Pr(>F)
## group  4  0.1835 0.9467
##      148
```

The p value obtained by normality test is $0.1178 > 0.05$, so it can explain the normal distribution of the data. The p value obtained by variance homogeneity test is $0.9467 > 0.05$, so it can be explained that the variance is homogeneous. On this basis, ANOVA can be selected for analysis of variance.

```
mod<-aov(Wind~Month,data=data2)
summary(mod)
```

```
##              Df Sum Sq Mean Sq F value  Pr(>F)
## Month          4   164.3    41.07   3.529 0.00879 **
## Residuals     148  1722.3    11.64
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The p value obtained from ANOVA is $0.00879 < 0.05$, so we are confident that there are significant differences between these sample groups, that is, the wind speed in New York state is not equal in each month. Under the previous conditions, we finally chose Dunnett test to make pairwise comparison between the control group and other groups.

```
#library(DescTools) it contains the DunnettTest function
DunnettTest(x=data2$Wind,g=data2$Month)
```

```
##
##  Dunnett's test for comparing several treatments with a control :
##    95% family-wise confidence level
##
## $'5'
##      diff      lwr.ci      upr.ci    pval
## 6-5 -1.355914 -3.512983  0.8011548 0.3424
## 7-5 -2.680645 -4.819960 -0.5413304 0.0086 **
## 8-5 -2.829032 -4.968347 -0.6897175 0.0051 **
```

```
## 9-5 -1.442581 -3.599649 0.7144881 0.2891
##
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

3.3 Results

Finally, compared with the data in May, there is no significant difference between June ($P = 0.3424 > 0.05$) and September ($P = 0.2891 > 0.05$) and may. However, the difference between July and may is significant ($P = 0.0086 < 0.05$), and there is also a significant difference between August and may ($P = 0.0051 < 0.05$).

3.4 Conclusion

Compared with the wind speed in May, the wind speed in June and September in New York state did not change significantly. However, the wind speed in August and September is more obvious than that in May. Combined with the box diagram, it can be seen that the wind speed in August and September can be said to have decreased significantly. These data can provide better help for human production and life.

4 Case study 2

In Environmental Science, crops or vegetation in a region are often destroyed by pests. Therefore, the prevention and control of regional diseases and pests has become an important factor to protect a regional environment. For example, China took diseases and pests as an important indicator of the forest evaluation system in 1992(Jun and Xing-yao, 2005). A group of experimenters studied the effects of different pesticides on insects in agricultural experimental units, and made statistics on the results(Beall, 1942). A total of six insecticides were used in the experiment. The same kind of insects were killed in the same experimental environment. Finally, the surviving insects were recorded.

4.1 Scientific question

Among these six insecticides, compared with type A insecticides, which insecticides have similar effects to type A insecticides and which insecticides have obvious differences in the effects on insects. Is the effect better or worse than type A insecticides? If it is better than type A insecticides, is it more conducive to killing insects in the natural environment to protect vegetation and farmers?

4.2 Procedure

First, import the data and conduct basic analysis on the data.

```
data('InsectSprays')
k1<-InsectSprays
k1$spray<-factor(k1$spray,ordered=T)
tapply(k1$count,k1$spray,mean)
```

```
##           A           B           C           D           E           F
## 14.500000 15.333333  2.083333  4.916667  3.500000 16.666667
```

```
summary(k1)
```

```
##      count      spray
## Min.   : 0.00    A:12
## 1st Qu.: 3.00    B:12
## Median : 7.00    C:12
## Mean   : 9.50    D:12
## 3rd Qu.:14.25    E:12
## Max.   :26.00    F:12
```

It can be found from the results that the number of surviving insects in group A, B and F is relatively close, while the number of surviving insects in group C,D,E is much less than that in the first three groups. So can we say that the effects of A,B,F three insecticides are the same? This requires further testing. Firstly, we test the normality and homogeneity of variance of sample data.

```
shapiro.test(k1$count)
```

```
##
## Shapiro-Wilk normality test
##
## data:  k1$count
## W = 0.9216, p-value = 0.0002525
```

```
leveneTest(count~spray,data=k1)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##      Df F value  Pr(>F)
## group  5  3.8214 0.004223 **
##      66
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Through the test of variance and normal distribution, we find that the p-value in Shapiro-wilk normality test is $0.0002525 < 0.05$, and the p-value in Levene's Test for Homogeneity of Variance is $0.004223 < 0.05$, they proves that this data set is not normally distributed, and the variance is not homogeneity, so we need to change our methods to find the significant difference.

```
oneway.test(count~spray,k1,var.equal = F)
```

```
##
## One-way analysis of means (not assuming equal variances)
##
## data:  count and spray
## F = 36.065, num df = 5.000, denom df = 30.043, p-value = 7.999e-12
```

Through the Welch's test, we can determine that there are significant differences in the number of insects under the influence of different insecticides in each group by comparing $P\text{-value} = 7.999e-12 < 0.05$. So what are the groups causing the difference? This requires the use of games-Howell test.

```

# Tukey
# Games-Howell
tukey <- function( data,
                  group,
                  method=c("Tukey", "Games-Howell"))
{
  OK <- complete.cases(data, group)
  data <- data[OK]
  group <- factor(group[OK])
  n <- tapply(data, group, length)
  a <- length(n)
  phi.e <- sum(n)-a
  Mean <- tapply(data, group, mean)
  Variance <- tapply(data, group, var)
  result1 <- cbind(n, Mean, Variance)
  rownames(result1) <- paste("Group", 1:a, sep="")
  method <- match.arg(method)
  if (method == "Tukey") {
    v.e <- sum((n-1)*Variance)/phi.e
    t <- combn(a, 2, function(ij)
      abs(diff(Mean[ij]))/sqrt(v.e*sum(1/n[ij])) )
    p <- ptukey(t*sqrt(2), a, phi.e, lower.tail=FALSE)
    Tukey <- cbind(t, p)
    rownames(Tukey) <- combn(a, 2, paste, collapse=":")
    return(list(result1=result1, Tukey=Tukey, phi=phi.e, v=v.e))
  }
  else {
    t.df <- combn(a, 2, function(ij) {
      t <- abs(diff(Mean[ij]))/sqrt(sum(Variance[ij]/n[ij]))
      df <- sum(Variance[ij]/n[ij])^2/sum((Variance[ij]/n[ij])^2/(n[ij]-1))
      return(c(t, df))} )
    t <- t.df[1,]
    df <- t.df[2,]
    p <- ptukey(t*sqrt(2), a, df, lower.tail=FALSE)
    Games.Howell <- cbind(t, df, p)
    rownames(Games.Howell) <- combn(a, 2, paste, collapse=":")
    return(list(result1=result1, Games.Howell=Games.Howell))
  }
}
tukey(k1$count,k1$spray,method=c('Tukey','Games-Howell'))

```

```

## $result1
##      n      Mean  Variance
## Group1 12 14.500000 22.272727
## Group2 12 15.333333 18.242424
## Group3 12  2.083333  3.901515
## Group4 12  4.916667  6.265152
## Group5 12  3.500000  3.000000
## Group6 12 16.666667 38.606061
##
## $Tukey
##      t      p
## 1:2 0.5204724 9.951810e-01

```

```
## 1:3 7.7550382 1.075055e-09
## 1:4 5.9854322 1.442013e-06
## 1:5 6.8702352 4.086244e-08
## 1:6 1.3532281 7.542147e-01
## 2:3 8.2755106 1.184582e-10
## 2:4 6.5059045 1.793911e-07
## 2:5 7.3907075 4.842290e-09
## 2:6 0.8327558 9.603075e-01
## 3:4 1.7696060 4.920707e-01
## 3:5 0.8848030 9.488669e-01
## 3:6 9.1082663 0.000000e+00
## 4:5 0.8848030 9.488669e-01
## 4:6 7.3386603 5.998531e-09
## 5:6 8.2234633 1.488213e-10
##
## $phi
## [1] 66
##
## $v
## [1] 15.38131
```

4.3 Results

Finally, through the Games-Howell test, we can compare the differences between each two groups through the p value. As we set the Group A as control group, we can find it has significant difference with group C($p=1.075055e-09<0.05$), D($1.442013e-06<0.05$) and E($4.086244e-08<0.05$), but has not significant difference with group B($7.542147e-01>0.05$) and group F($7.542147e-01>0.05$).

4.4 Conclusion

It can be concluded that insecticide A,B,F has similar effects on insects, but insecticide C,D,E has better effects than insecticide A,B,F. Therefore, it can be said that when encountering diseases and pests, the use of C, D or e insecticides can more effectively kill vegetation or diseases and pests in crops. However, it can not be directly determined that these types of insecticides are the most suitable. Therefore, it is also necessary to study whether insecticides have destructive effects on vegetation and water sources. Based on these, we can finally determine the most suitable type of insecticides, which can also be used as the object of future research.

5 Exercises

5.1 Exercises 1

Sunlight is the main radiation of the earth. According to the research, the sun can affect the earth's climate system (Nazari-Sharabian and Karakouzian, 2020). Sunspots are one of the important solar activities. Studying the number of sunspots is helpful to analyze the earth's climate. Therefore, researchers counted the changes of sunspots between 1749 and 1983 (Andrews and Herzberg, 2012). If we take each decade as a cycle, is there any significant change in the average number of sunspots in each subsequent decade compared with the first decade?


```
data("sunspots")
```

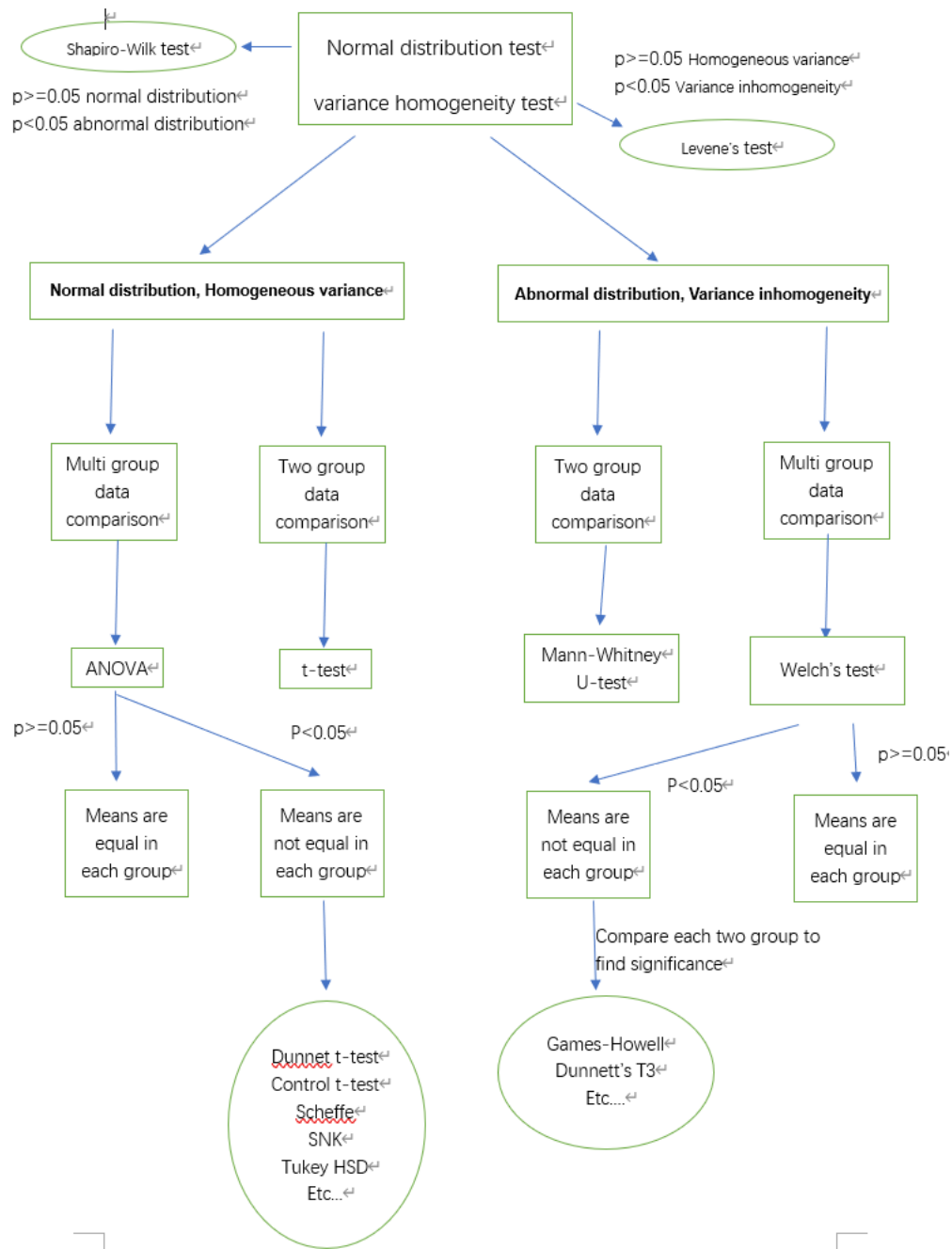
5.2 Exercise 2

This famous (Fisher or Anderson's) iris data set gives four indicators of three Iris species (Becker et al., 1988): bristly iris, variegated iris and Virginia iris, that is, the measurement of sepal length and sepal width, and petal length and petal width, in centimeters. The whole data set has 150 rows and 5 columns. So are there significant differences between the four measurements of the three Iris species?

```
data('iris')
```

6 Appendix

6.1 1.Flow Chart



6.2 2.Dunnett's table

n	α	Total number of groups (including control group)						
		2	3	4	5	6	7	8
5	0.05	2.57	3.03	3.29	3.48	3.62	3.73	3.82
	0.01	4.03	4.63	4.98	5.22	5.41	5.56	5.69
6	0.05	2.45	2.86	3.1	3.26	3.39	3.49	3.57
	0.01	3.71	4.21	4.51	4.71	4.87	5	5.1
7	0.05	2.36	2.75	2.97	3.12	3.24	3.33	3.41
	0.01	3.5	3.95	4.21	4.39	4.53	4.64	4.74
8	0.05	2.31	2.67	2.88	3.01	3.13	3.22	3.29
	0.01	3.36	3.77	4	4.17	4.29	4.4	4.48
9	0.05	2.26	2.61	2.81	2.95	3.05	3.14	3.2
	0.01	3.25	3.63	3.85	4.01	4.12	4.22	4.3
10	0.05	2.23	2.57	2.76	2.89	2.99	3.07	3.14
	0.01	3.17	3.53	3.74	3.88	3.99	4.08	4.16
11	0.05	2.2	2.53	2.72	2.84	2.94	3.02	3.08
	0.01	3.11	3.45	3.65	3.79	3.89	3.98	4.05
12	0.05	2.18	2.5	2.68	2.81	2.9	2.98	3.04
	0.01	3.05	3.39	3.58	3.71	3.81	3.89	3.96
13	0.05	2.16	2.48	2.65	2.78	2.87	2.94	3
	0.01	3.01	3.33	3.52	3.65	3.74	3.82	3.89
14	0.05	2.14	2.46	2.63	2.75	2.84	2.91	2.97
	0.01	2.98	3.29	3.47	3.59	3.69	3.76	3.83
15	0.05	2.13	2.44	2.61	2.73	2.82	2.89	2.95
	0.01	2.95	3.25	3.43	3.55	3.64	3.71	3.78
16	0.05	2.2	2.42	2.59	2.71	2.8	2.87	2.92
	0.01	2.92	3.22	3.39	3.51	3.6	3.67	3.73
17	0.05	2.11	2.41	2.58	2.69	2.78	2.85	2.9
	0.01	2.9	3.19	3.36	3.47	3.56	3.63	3.69
18	0.05	2.1	2.4	2.56	2.68	2.76	2.83	2.89
	0.01	2.88	3.17	3.33	3.44	3.53	3.6	3.66
19	0.05	2.09	2.39	2.55	2.66	2.75	2.81	2.87
	0.01	2.86	3.15	3.31	3.42	3.5	3.57	3.63
20	0.05	2.09	2.38	2.54	2.65	2.73	2.8	2.86
	0.01	2.85	3.13	3.29	3.4	3.48	3.55	3.6
30	0.05	2.04	2.32	2.47	2.58	2.66	2.72	2.77
	0.01	2.75	3.01	3.15	3.25	3.33	3.39	3.44
40	0.05	2.02	2.29	2.44	2.54	2.62	2.68	2.73
	0.01	2.7	2.27	3.09	3.19	3.26	3.32	3.37

7 References

- ZACH, Dunnett's table. *Statology*. Available at: <https://www.statology.org/dunnetts-table/> [Accessed May 8, 2022].
- Andrews, D.F., Herzberg, A.M., 2012. Data: A collection of problems from many fields for the student and research worker. Springer Science & Business Media.
- Beall, G., 1942. The transformation of data from entomological field experiments so that the analysis of variance becomes applicable. *Biometrika* 32, 243–262.
- Becker, R.A., Chambers, J.M., Wilks, A.R., 1988. The new s language. Wadsworth & brooks. Cole.[Google Scholar].
- Brown, J.D., 2008. The bonferroni adjustment. *Statistics* 12.
- Chambers, J., Cleveland, W., Kleiner, B., Tukey, P., 1983. Graphical methods for data analysis. Wadsworth int'l. Group, Belmont, CA.
- Hilton, A., Armstrong, R.A., 2006. Statnote 6: Post-hoc ANOVA tests. *Microbiologist* 2006, 34–36.
- Holland, B.S., Copenhaver, M.D., 1987. An improved sequentially rejective bonferroni test procedure. *Biometrics* 417–423.
- Jun, L., Xing-yao, Z., 2005. Ecological control of forest pest: A new strategy for forest pest control. *Journal of Forestry Research* 16, 339–342.
- Koenig, F., Brannath, W., Bretz, F., Posch, M., 2008. Adaptive dunnett tests for treatment selection. *Statistics in medicine* 27, 1612–1625.
- Nazari-Sharabian, M., Karakouzian, M., 2020. Relationship between sunspot numbers and mean annual precipitation: Application of cross-wavelet transform—a case study. *J* 3, 67–78.
- Ruxton, G.D., Beauchamp, G., 2008. Time for some a priori thinking about post hoc testing. *Behavioral ecology* 19, 690–693.