1. Suppose the National Transportation Safety Board (NTSB) wants to examine the safety of compact cars, midsize cars, and full-size cars. It collects a sample of three for each of the treatments (cars types). Using the hypothetical data provided below, test whether the mean pressure applied to the driver’s head during a crash test is equal for each types of car.

|  |  |  |
| --- | --- | --- |
| Compact cars | Midsize cars | Full-size cars |
| 643 | 469 | 484 |
| 655 | 427 | 456 |
| 702 | 525 | 402 |

* State the hypotheses & list necessary assumptions.
* Run the test using alpha of 5% in R & fill ANOVA table manually.
* Provide your decision and conclusions.

1. H0: mu (CC) == mu (MS) == mu (FC)

Ha: at least one car has different average pressure

1. P = 0.00121 < *a* = 0.05 (yes) Reject H0, Ha is accepted
2. Conclusion: at least one car has different average pressure
3. Girls from four different soccer teams are to be tested for mean goals scored per game. The entries in Soccer.csv file are the goals per game for the different teams.

* Indicate SSF, MSF and numerator degree of freedom.
* Indicate SSE, MSE and denominator degree of freedom.
* What is the F statistic?

1. H0: mu (T1) == mu (T2) == mu (T3) == mu (T4)

Ha: at least one team has different average goals

p = 0.000485 < a = 0.01(yes) => Reject H0, Accept Ha

Conclusion: Ha: at least one team has different average goals

1. SSF = 25.75, MSF = 8.583, NDF = 3
2. SSE = 13.20, MSE = 0.825, DDF = 16
3. F = 10.4
4. A grassroots group opposed to a proposed increase in the gas tax claimed that the increase would hurt working-class people the most, since they commute the farthest to work. Suppose that the group randomly surveyed 24 individuals and asked them their daily one-way commuting mileage. The results are given in the Workshop data.

**Using a 5% significance level, test the hypothesis whether the three mean commuting mileages are significantly different. What were the list of assumptions you applied for analysis?**

H0: mu (WC) == mu (MC) == mu (RC)

Ha: at least one class has different average distance

p = 0.00143 < a = 0.05, Reject H0, Accept Ha

Conclusion: Not all averages are equal

1. Suppose the US Golf Association (USGA) wants to compare the mean distances reached of four different brands of golf balls struck with a driver. A completely randomized design is employed, with Iron Byron, the USGA’s robotic golfer, using a driver to hit a random sample of 10 balls of each brand in a random sequence. The distance is recorded for each hit, and the results are shown in workshop data, organized by brand.

* Set up the test to compare the mean distances for the four brands. Use a = 0.10.
* Use R to obtain the test statistic and p-value. Give the appropriate conclusion.

H0: mu (A) = mu (B) = mu (C) = mu (D)

Ha: at least one class has different average distance

P = 3.97 \* 10-12 < a = 0.10(yes) => Reject H0, Accept Ha

Conclusion: not all mean distances are equal

1. [**HOMEWORK**] Assume, you approached random sample of students from 4EC1, 4BIS2, 4BAMB15 and 4FIN5 groups and asked about their ISDS midterm exam marks. The imaginary results are given in gpa.csv file. Using a significance level of 1%, can we conclude that the students’ performances differ among the groups?
2. [**HOMEWORK**] Using in-built ***mtcars*** dataset, test whether mile per gallon differ among cars with different types of gear. Use significance level at 1%.
3. [**HOMEWORK**] Import college data from cloud and conduct following tests:
   1. Test whether average SAT differ among region at 1% significance level.
   2. Test whether tuition fee differs among states at 5% significance level.