Homework №19

Author: David Oniani
Instructor: Dr. Eric Westlund

May 06, 2019

27.3 (a) Below is the stemplot for the Group 1 (Never looged).

```
1 | 6   9   9
2 | 0   1   2   4   7   7   8   9
3 | 3
```

Below is the stemplot for the Group 2 (1 year ago).

```
0 | 2   9
1 | 2   2   4   4   5   7   7   8   9
2 | 0
```

Below is the stemplot for the Group 3 (8 years ago).

```
0 | 4
1 | 2   2   5   8   8   9
2 | 2   2
```

For all stemplots, the key is $X \mid Y = XY$. For instance, $2 \mid 0 = 20$.

(b) It appears that there are more trees in Group 1 (never logged) than in other groups. We have, $\overline{x}_1 = 23.75$, $\overline{x}_2 = 14.0833$, and $\overline{x}_3 = 15.7778$. Then, it is easy to see that $\overline{x}_1 > \overline{x}_2 > \overline{x}_3$.

(c) $H_0 = \mu_1 = \mu_2 = \mu_3$, in other words, null hypothesis is that all means are the same. We test the null against H_a : not all means are the same (or at least one mean is different). Our F-statistic is $F = 11.43$ with df (between groups) $= 2$ and df (within groups) $= 30$ with $P = 0.0002$ and therefore, we conclude that the differences are statistically significant. In other words, the mean number of trees per plot is significantly lower in logged areas (not that Group 1 is the area that is never logged, and it has the highest mean).

27.4 (a) We have, $\overline{x}_1$ (Conservative) $= 51.09$, $\overline{x}_2$ (Moderate) $= 46.98$, and $\overline{x}_3$ (Liberal) $= 46.80$. From the statistics above, we can see that the mean is the highest in the first group (Conservative) and that the mean is lowest in the third group (Liberal).

(b) $H_0 = \mu_1 = \mu_2 = \mu_3$, in other words, null hypothesis is that all means are the same. We test the null against H_a : not all means are the same (or at least one mean is different). Our F-statistic is $F = 11.69$ with df (between groups) $= 2$ and df (within groups) $= 1871$ with $P < 0.0001$ and therefore, we conclude that the differences are statistically significant. In other words, the mean number of people who are Conservatives is significantly higher than the mean number of people in either of the other two groups (Moderates and Liberals).

27.7 (a) The distributions are approximately normal with each being independent SRS. Notice that $s_1^2 = 25.6591$, $s_2^2 = 24.8106$, and $s_3^2 = 33.1944$. Therefore, we get $s_1 = 5.065$, $s_2 = 4.981$, and $s_3 = 5.761$. Now, the ratio of the largest standard

deviation over the lowest standard deviation is $\dfrac{s_3}{s_2} = \dfrac{5.761}{4.981} \approx 1.16 < 2$ and therefore, the conditions are satisfied.

(b) The distributions are approximately normal with each being independent SRS. Notice that $s_C = 17.42$, $s_M = 18.13$, and $s_L = 17.41$. Now, the ratio of the largest standard

deviation over the lowest standard deviation is $\dfrac{s_M}{s_L} = \dfrac{18.13}{17.41} \approx 1.04 < 2$ and therefore, the conditions are satisfied.

27.8 After calculating the standard deviations, I got that $s_1 = 4.372$, $s_2 = 4.5$, and $s_3 = 3.529$. The ratio of the largest standard deviation over the lowest standard deviation is $\dfrac{s_1}{s_3} = \dfrac{4.5}{3.529} \approx 1.28 < 2$. In this case, however, the normality condition is not satisfied as the

datasets seem to be skewed and there are not enough datapoints for CLT to make the distributions approximately normal. Therefore, all the conditions are not satisfied.

27.10 (a) I is the number of groups in the data. In this case, $I = 3$.
       n_i is the sample size of the group i. In this case, $n_1 = 12$, $n_2 = 9$, and $n_3 = 12$.
       N is the total sample size. In this case, $N = n_1 + n_2 + n_3 = 12 + 9 + 12 = 33$.

       (b) df (between groups) $= I - 1 = 3 - 1 = 2$.
       df (within groups) $= N - I = 33 - 3 = 30$.
       These values indeed match those in Figure 27.5.

27.12 (a) The sample size is large enough for the CLT (Central Limit Theorem) to make the distributions approximately normal.

       (b) The ratio of the largest standard deviation over the lowest standard deviation is $\dfrac{s_1}{s_3} = \dfrac{3.11}{1.60} \approx 1.94 < 2$. Therefore, the standard deviations indeed satisfy the guidelines for ANOVA inference.

       (c) The $\bar{x}$ is $\bar{x} = 1.31$.
       MSG $= 178.07$.
       MSE $= 5.123$.
       $$F = \frac{\text{MSG}}{\text{MSE}} = \frac{178.07}{5.123} = 34.759.$$

       (d) df (numerator) $= I - 1 = 3 - 1 = 2$.
       df (denominator) $= N - I = (244 + 734 + 364) - 3 = 1342 - 3 = 1339$.

       Therefore, I would use the F-distribution with 2 degrees of freedom in the numerator and 1339 degrees of freedom in the denominator. Since, $P < 0.001$, and the commonly used significance level α is 0.05, we reject the null and accept the alternative as $P < 0.05$ $(0.001 < 0.05)$.