Assignment3

Amrit Bhat

2/10/2022

Question 1. The first goal is to compare the mean IQ scores for males and females. Use a 2-sample t-test for this comparison. What is the p-value?

```
lead_df <- read.csv("/Users/amrit/Documents/Courses/Applied Statistics & Experimental Design/Assignments
var(lead_df[lead_df$SEX==1,]$IQ)
```

```
[1] 222.9298
```

```
var(lead_df[lead_df$SEX==2,]$IQ)
```

```
[1] 184.554
```

The variance of above two groups are not equal, therefore we consider the Welch t-test

```
welch_t_test_results <- with(lead_df, t.test(IQ[SEX==1],IQ[SEX==2],var.equal = F))
welch_t_test_results
```

```
    Welch Two Sample t-test

data:  IQ[SEX == 1] and IQ[SEX == 2]
t = 0.1541, df = 146.58, p-value = 0.8777
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -4.258377  4.978637
sample estimates:
mean of x mean of y
 91.23684  90.87671
```

Question 2. State the conclusion from your test.

As the p-value = 0.878 > 0.05, it is statistically not siginificant and therefore we do not reject the null hypothesis that the means are equal.

Question 3. Are the independence assumptions valid for the t-test in this situation? Give a brief explanation.

I do not believe that the independence assumptions hold for this test as the children were chosen randomly from the same area around the lead smelter. Relying on the fact that "You are an average of the 5 people around you", there is a high possibility of the IQ of the children being dependent upon each other.

Question 4. The second goal is to compare the mean IQ scores in the 3 groups. State in words the null hypothesis for this test.

Null Hypothesis: Let the mean Sepal Width for the 3 groups by μ_1, μ_2, μ_3, for *Group1*, *Group2*, and *Group3*, respectively. Suppose that we want to test the null hypothesis $H_0 : \mu_1 = \mu_2 = \mu_3$

Question 5. State in words the alternative hypothesis for this test.

Alternative Hypothesis: μ_1, μ_2, μ_3 are not all equal to each other

Question 6. What method should be used to perform the test?

ANOVA test should be used to compare the means of the three groups at the same time.

Question 7. Perform the test. Report the p-value.

```r
summary(aov(IQ ~ factor(GROUP), data=lead_df))
```

```
               Df Sum Sq Mean Sq F value Pr(>F)
factor(GROUP)   2   1491   745.4   3.816 0.0242 *
Residuals     146  28522   195.4
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Question 8. State your conclusion about the evidence for an association between lead exposure and IQ.

From the p-value we deduce that two or more groups don't have the same mean. Therefore there might be an association between lead exposure and IQ

Question 9. Are there strong reasons to believe that the assumptions of this test are not met? Briefly justify your answer.

Assumptions of ANOVA are Independence (of samples and of observations within each sample), Equal variances, Large sample sizes *or* normal distributions

Checking for variance first,

```
[1] "Group 1 variance:  242.434643143545"
```

```
[1] "Group 2 variance:  90.2973790322581"
```

```
[1] "Group 3 variance:  168.033333333333"
```

As the variances aren't equal, we can conclude that the assumptions of this test do not hold

Question 10. Conduct all pairwise comparison of group means. Report the p-values.

```
[1] "Group 1 & Group 2 mean comparison p-value:  0.0120505968735434"
```

```
[1] "Group 2 & Group 3 mean comparison p-value:  0.813103552005519"
```

```
[1] "Group 3 & Group 1 mean comparison p-value:  0.0230031898393195"
```

Question 11. What conclusion about the association between lead and IQ would you draw from the pairwise comparisons of group means? Does it agree with the conclusion in Q8? (Consider the issue of multiple testing in your answer.)

Upon bonferroni correction, we know that the significance level for each comparison would be 0.016667. As the p-value for the comparison between group 1 and group 2 is less than the significance level after bonferroni correction we reject the null hypothesis. This agrees with our conclusion in Q8 that there might be an assoication between lead and IQ.

Question 12. Now we wish to compare the 3 group means for males and females separately. Display some appropriate descriptive statistics for this analysis.

```
men_df <- lead_df[lead_df$SEX==1,]
women_df <- lead_df[lead_df$SEX==2,]

men_mean=with(men_df,tapply(IQ,GROUP,mean))
men_sd=with(men_df,tapply(IQ,GROUP,sd))
men_sample_length=with(men_df,tapply(IQ,GROUP,length))
men_descriptive_stats = data.frame(men_mean,men_sd,men_sample_length)
men_descriptive_stats
```

```
  men_mean   men_sd men_sample_length
1 92.93478 15.42351                46
2 90.17647 11.13124                17
3 86.61538 17.32791                13
```

```
women_mean=with(women_df,tapply(IQ,GROUP,mean))
women_sd=with(women_df,tapply(IQ,GROUP,sd))
women_sample_length=with(women_df,tapply(IQ,GROUP,length))
women_descriptive_stats = data.frame(women_mean,women_sd,women_sample_length)
women_descriptive_stats
```

```
  women_mean  women_sd women_sample_length
1   94.60976 15.877465                  41
2   84.80000  6.471917                  15
3   87.23529  8.898942                  17
```

Question 13. Perform tests to compare the mean IQ scores in the 3 groups for males and females separately. Report the p-values from the two tests.

```
summary(aov(IQ ~ factor(GROUP), data=men_df))
```

```
              Df Sum Sq Mean Sq F value Pr(>F)
factor(GROUP)  2    429   214.7   0.962  0.387
Residuals     73  16290   223.2
```

```
summary(aov(IQ ~ factor(GROUP), data=women_df))
```

```
              Df Sum Sq Mean Sq F value Pr(>F)
factor(GROUP)  2   1351   675.3    3.96 0.0235 *
Residuals     70  11937   170.5
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Question 14. What can you conclude about the association between lead and IQ from these tests? Does it agree with the result in Q8 and Q11? (Consider multiple testing.)

There is no association between lead and IQ in men (disagreeing with Q8 and Q11), whereas there is an association between lead and IQ in women (agreeing with Q8 & Q11)

Question 15. Now perform all 3 pairwise comparisons of groups for males and females separately. Report the p-values from these tests?

```r
print("Pariwise comparison for Men")
```

```
[1] "Pariwise comparison for Men"
```

```r
men_group_1_IQ <- men_df[men_df$GROUP == 1,]$IQ
men_group_2_IQ <- men_df[men_df$GROUP == 2,]$IQ
men_group_3_IQ <- men_df[men_df$GROUP == 3,]$IQ

t.test(men_group_1_IQ,men_group_2_IQ, var.equal = F)$p.value
```

```
[1] 0.4391938
```

```r
t.test(men_group_2_IQ,men_group_3_IQ, var.equal = F)$p.value
```

```
[1] 0.5258587
```

```r
t.test(men_group_3_IQ,men_group_1_IQ, var.equal = F)$p.value
```

```
[1] 0.2502756
```

```r
print("Pariwise comparison for Women")
```

```
[1] "Pariwise comparison for Women"
```

```r
women_group_1_IQ <- women_df[women_df$GROUP == 1,]$IQ
women_group_2_IQ <- women_df[women_df$GROUP == 2,]$IQ
women_group_3_IQ <- women_df[women_df$GROUP == 3,]$IQ

t.test(women_group_1_IQ,women_group_2_IQ, var.equal = F)$p.value
```

```
[1] 0.001830775
```

```r
t.test(women_group_2_IQ,women_group_3_IQ, var.equal = F)$p.value
```

```
[1] 0.3796352
```

```r
t.test(women_group_3_IQ,women_group_1_IQ, var.equal = F)$p.value
```

```
[1] 0.02927457
```

Question 16. What do you conclude about the association between lead and IQ from the results in Q13? Does your conclusion change from previous conclusions made in Q8, Q11 and Q14?

As can be seen from the results, the pairwise t-test results for all groups in men indicates that the null hypothesis is not rejected, hence there is no association between lead and IQ in men (similar to conclusion in Q14, disagreeing with Q8 and Q11).

In women, results of one of comparison between two of the groups is statistically significant, therefore we can reject the null hypotheis (similar to conclusion in Q14). Therefore, there is an association between lead and IQ in women (agreeing with Q8 & Q11)