Mrinal Chaudhari ***Applied Statistic-614***

Final Exam 21

1. A clinical trial was run to assess the usefulness of a new drug. The subjects in the study had a particular infection. The goal of the study was to see if there was association between curing the infection and whether or not the subject got the drug or a placebo. This trial was run at eight different medical centers. The data is given in a table below.

	Success	Failure
Drug	55	75
Control	47	96

a. Do a preliminary analysis and draw conclusions

The goal of the study is only to find an association between curing the infection and whether or not the subject got the drug or the placebo therefore, I have discarded the Centers column and have summed up the Successes and Failures of them

We can conclude that there are more Failures than Success from the above table.
From the above data, We can say that there is no association between curing the infection and whether the subject received the drug or the placebo.

b.**Give the null and alternative hypotheses for this case**

H_0: X *is independent of* Y; H_A: X *is NOT independent of* Y; $\alpha = 0.05$;

c.Test the hypothesis from part b

$$\chi^2_{Stat} = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \; ; \; E_{i,j} = \frac{R_i C_j}{N} \; ;$$

$E_{1,1} = 48.57$; $E_{1,2} = 81.43$ $E_{2,1} = 53.43$; $E_{2,2} = 89.57$;

$Therefore,\ since\ each\ expected\ count\ is\ greater\ than\ 5,\ the\ Chi - Square\ test\ is\ valid.$

$\chi^2_{Stat} = 2.593;\ df = 1;$

$p - value = P(\chi^2_{df} \geq \chi^2_{Stat}) = 1 - P(\chi^2 \leq \chi^2_{Stat}) = 0.1073$

$Therefore,\ since\ p - value > \alpha,\ we\ fail\ to\ reject\ H_0$

d.Give the real-world results

$We\ can\ conclude\ from\ the\ above\ result\ that\ there\ is\ NO\ association\ between\ curing\ the\ infection\ and\ whethe$
$the\ subject\ got\ the\ drug\ or\ the\ placebo.$

2.

a. Do a preliminary analysis and draw conclusions

From the question , it is clear that this is an experimental study to understand the delay in responding to stimuli.There are 2 types of cues and the human subject is informed/warned about the incoming cue on 3 different intervals beforehand (5, 10 and 15 seconds). However, the question does not help us to understand if the human subject is also informed about the type of incoming cue, also it does not help us understand how many human subjects the experiment was run on. Additionally, from the given data and the explanation of the experiment provided, we do not have enough information to understand what the human subject has to provide as an action or a response to the incoming cues.

b.**Give the null and alternative hypotheses for this case**

H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_6$; H_A: $\mu_i \neq \mu_j$ $for\ at\ least\ one\ (i, j)\ pair\ where\ i < j$

c.Test the hypothesis from part b (including if necessary, deciding which means are different)

$a = 6;\ n = 3$

$N = a * n = 6 * 3 = 18;\ \alpha = 0.10$

$F_{stat} = \frac{MS_{Treatment}}{MS_{Error}};\ SS_T = 0.029;\ SS_M = 0.0255;\ SS_E = 0.0035;$

$df_T = 17;\ df_M = 5;\ df_E = 12;$

$MS_M = 0.0051$

$MS_E = 0.0003$

$F_{stat} = 17.661$

$p - value = 1 - P(F_{df_{Treat}, df_E} \geq F_{Stat}) = 0.000037$

$p - value < \alpha,\ We\ reject\ H_0\ in\ favour\ of\ the\ alternate\ hypothesis.$

$Since\ 0.000037 < 0.10 -\ we\ now\ need\ to\ run\ Tukey's\ All - Pairwise\ tests\ to\ see\ which\ mean\ pairs\ are\ di$

$The\ table\ below\ gives\ the\ confidence\ bounds\ and\ adjusted\ p - value\ for\ each\ pair.$

$If\ the\ p - values\ are\ less\ than\ \alpha\ or\ the\ confidence\ intervals\ do\ not\ cover\ 0,\ then\ the\ means\ are\ different.$

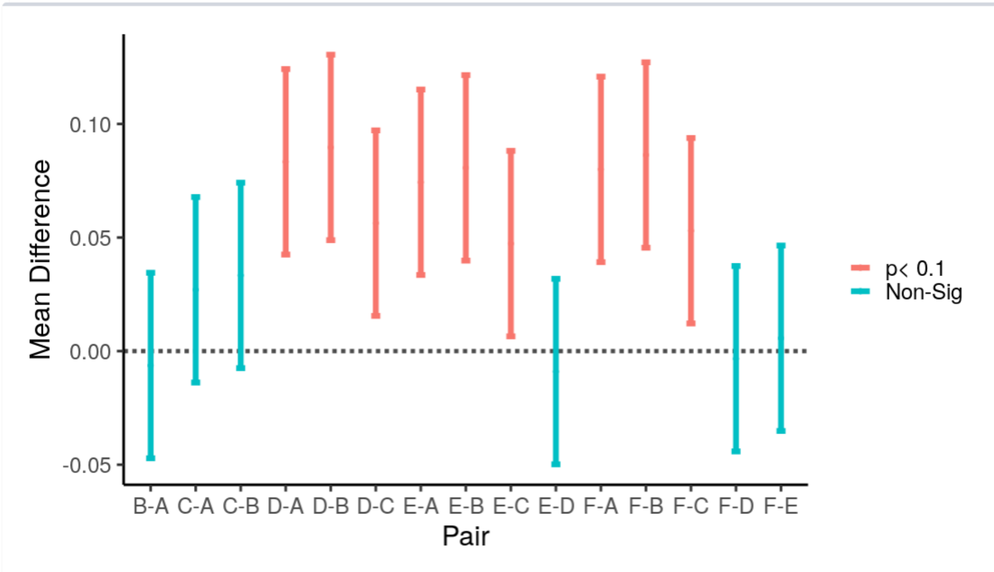
ANOVA Table

	SS	df	MS	F	P-value
X	0.03	5.00	0.01	17.66	0.000037
Error	0.00	12.00	0.00	NA	NA
Total	0.03	17.00	NA	NA	NA

Tukey's Test Results Table

Pairs	Mean Difference	Lower CB	Upper CB	P-Value
B-A	-0.0063	-0.0471	0.0345	0.9969
C-A	0.027	-0.0138	0.0678	0.4235
D-A	0.0833	0.0425	0.1241	0.0007
E-A	0.0743	0.0335	0.1151	0.0018
F-A	0.08	0.0392	0.1208	0.001
C-B	0.0333	-0.0075	0.0741	0.23
D-B	0.0897	0.0489	0.1305	0.0003
E-B	0.0807	0.0399	0.1215	0.0009
F-B	0.0863	0.0455	0.1271	0.0005
D-C	0.0563	0.0155	0.0971	0.0154
E-C	0.0473	0.0065	0.0881	0.046
F-C	0.053	0.0122	0.0938	0.0231
E-D	-0.009	-0.0498	0.0318	0.9845
F-D	-0.0033	-0.0441	0.0375	0.9999
F-E	0.0057	-0.0351	0.0465	0.9982

Tukey's Test Results Graph



d.Give the real-world results
The pairs with the same type of cue (audio-audio & visual-visual) do not possess a significant difference.Every other pair where the type of cue is a mix of both audio and visual contain a significant difference.When the human is trying to respond to the same type of cue - there is not a significant difference.However, when the human is trying to respond to both types of cues - there exists a significant difference in the means. Therefore, codes 2-1, 3-1,3-2,5-4,6-4,6-5 are all the pairs which do not contain a significant difference in their means.

3. An analyst expects that the percentages of car accidents per season are equal. The data for a given year is below.
a. **Do a preliminary analysis and draw conclusions**
From the given data we can conclude that we have to set 25% as the percentage of accidents per season. Since there are 4 seasons and the data provided is for a complete year.Therefore, the number of accidents in each season - Winter, Spring, Summer and Fall, contribute 25% each to the total 100% (N = total number of accidents) for the entire year.

b. **Give the null and alternative hypotheses for this case**
H_0: $\pi_{Winter} = 0.25$; $\pi_{Spring} = 0.25$; $\pi_{Summer} = 0.25$; $\pi_{Fall} = 0.25$; $H_A: At\ least\ one\ \pi_i \neq \pi_{i,0}\ for\ all\ i$;

C.Test the hypothesis from part b

$$\alpha = 0.05;\ \chi^2_{Stat} = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i};\ E = N\pi_i;\ E_{i,j} = \frac{R_i C_j}{N};\ N = 328 + 334 + 372 + 327 = 1361;$$

$$E_{Winter} = 1361 * 0.25 = 340.25;\ E_{Spring} = 1361 * 0.25 = 340.25;$$

$$E_{Summer} = 1361 * 0.25 = 340.25;\ E_{Fall} = 1361 * 0.25 = 340.25;$$

The assumption that each expected count to be greater than 5 *is met.*

$we\ can\ proceed\ with\ the\ Chi\ -\ Square\ test.$

$$\chi^2_{Stat} = \frac{(328-340.25)^2}{340.25} + \frac{(334-340.25)^2}{340.25} + \frac{(372-340.25)^2}{340.25} + \frac{(327-340.25)^2}{340.25} = 0.4410 + 0.1148 + 2.9627 + 0.5159 = 4.03$$

$$df = (a - 1) = 4 - 1 = 3;\ p - value = P(\chi^2_{df} \geq \chi^2_{Stat}) = 1 - P(\chi^2 \leq \chi^2_{Stat}) = 0.2582$$

$p - value > \alpha \Rightarrow We\ fail\ to\ reject\ H_0.$

$There\ is\ not\ enough\ evidence\ to\ support\ the\ claim\ of\ alternate\ hypotheses.$

d. Give the real-world results

The analyst was correct in his assumption/expectation that the percentage of car accidents per season are equal.

4.

The analyst is looking to run a study to understand if there is an affiliation between party and race, therefore I have to proceed to test this with the Test for Independence.
We use a test for independence to see if there is a relationship between a categorical response variable and a categorical input variable. We hypothesize that there is no relationship between the two. Then we collect data and compare the observed counts to the expected counts considering that the null is true. Here, I will be trying to test to see whether the parameter of interest - race - is independent of the political party parameter.

$H_0:\ X\ is\ independent\ of\ Y;\ H_A:\ X\ is\ NOT\ independent\ of\ Y;\ \alpha = 0.05;$

$$\chi^2_{Stat} = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}\ ;\ E_{i,j} = \frac{R_i C_j}{N}\ ;\ N = 980;$$

$$E_{1,1} = \frac{129 * 444}{980} = 58.444\ ;\ E_{1,2} = \frac{129 * 120}{980} = 15.795\ ;\ E_{1,3} = \frac{129 * 416}{980} = 54.759$$

$$E_{2,1} = \frac{851 * 444}{980} = 385.555\ ;\ E_{2,2} = \frac{851 * 120}{980} = 104.204\ ;\ E_{2,3} = \frac{851 * 416}{980} = 361.240$$

$Therefore,\ since\ each\ expected\ count\ is\ greater\ than\ 5,\ the\ Chi\ -\ Square\ test\ is\ valid.$

$$\chi^2_{Stat} = \frac{(103-58.444)^2}{58.444} + \frac{(15-15.795)^2}{15.795} + \frac{(11-54.759)^2}{54.759} + \frac{(341-385.555)^2}{385.555} + \frac{(105-104.204)^2}{104.204} + \frac{(405-361.240)^2}{361.240}$$

$$\chi^2_{Stat} = 33.968 + 0.040 + 34.968 + 5.148 + 0.006 + 5.3010 = 79.431$$

$df = (i - 1)(j - 1) = (2 - 1)(3 - 1) = 1(2) = 2$

$p - value = P(\chi^2_{df} \geq \chi^2_{Stat}) = 1 - P(\chi^2 \leq \chi^2_{Stat}) = 0$

$Therefore,\ 0 < 0.05 - we\ reject\ H_0\ in\ favour\ of\ H_A$

From the above conclusion, we can conclude that there is a relationship between party affiliation and race.I will report to the republican party analyst that there is a clear relationship between party affiliation and race.