**Introduction**

We are conducting a generalized randomized block design with two factors, two levels each, to understand whether a job applicant's eye contact (Yes/ No) and a personnel officer's gender (Male/ Female) have an effect on the likelihood of job applicant success. We will conduct this test using Two-way ANOVA. This is measured by the personnel officer rating the job applicant on a scale of 0 (complete failure) to 20 (complete success). 10 male and 10 female job applicants were chosen at random. Once grouped by gender, half of the males and females in each group were chosen at random to receive a version of the photograph in which the applicant made eye contact with the camera lens while the other half received a version in which no eye contact with the camera lens was made.

Organizing the Data

* I create a dataframe to read the data.

ECE <- as.data.frame(read.table(url("https://users.stat.ufl.edu/~burrdoss/Courses/4211/Data-and-R/Datasets/CH19PR12.txt")))

* Here I am naming factors and factor levels to begin and then setting them as factors.

names(ECE) <- c("Success","Contact","Gender","Group")

ECE["Gender"][ECE["Gender"] == 1] <- "M"

ECE["Gender"][ECE["Gender"] == 2] <- "F"

ECE["Contact"][ECE["Contact"] == 1] <- "Yes"

ECE["Contact"][ECE["Contact"] == 2] <- "No"

ECE$Contact <- as.factor(ECE$Contact)

ECE$Gender <- as.factor(ECE$Gender)

* I will now print 4 lines from the dataset, one from each cell.

subset(ECE, ECE['Group'] == 1)

A picture containing text, device, gauge

Description automatically generated

Analyzing the Data

* We will now find the Cell Means and Cell Standard deviations

cellmeans <- t(tapply(ECE$Success, list(ECE$Contact, ECE$Gender), mean))

cellsds <- t(tapply(ECE$Success, list(ECE$Contact, ECE$Gender), sd))

This code below will create the tables for the Cell Means and Standard Deviations by factor.

CMT <- list()

CMT[[1]] <- "Table of Cell Means"

CMT[[2]] <- cellmeans

CMSD <- list()

CMSD[[1]] <- "Table of Cell Standard Deviations"

CMSD[[2]] <- cellsds

CMT ; CMSD

A screenshot of a computer

Description automatically generated with low confidence

* Below I create the Plot of Factor Level Means for Gender and Eye Contact with Mean rate of Success as the response.

plot.design(Success ~ Contact + Gender, data=ECE)

Chart, box and whisker chart

Description automatically generated

* Now I create the Interaction plot between Gender and Eye Contact (Contact) with Likely Job Success of Applicant as the response.

interaction.plot(x.factor = ECE$Contact, trace.factor = ECE$Gender, response = ECE$Success,

fun = mean,

ylab = "Likely Job Success of Applicant",

xlab = "Contact",

trace.label = "Gender",

col = c("#0198f9", "#f95801"),

lwd = 3)

Chart, line chart

Description automatically generated

There is no important interaction between Eye Contact and Gender in the interaction plot as there is no crossing or significant change from one factor level to the next.

The four means (5 observations each) are shown to be Y.bar\_22 = 9.2, Y.bar\_21 = 13, Y.bar\_12 = 13.6, and Y.bar\_11 = 16.4

16.4 - 13 = 3.4 (Eye Contact: No) and 13.6 - 9.2 = 4.4 (Eye Contact: Yes)

The change in effect from male to female (3.4 to 4.4) is not a very large change in effect.

The standard deviations / variances are all relatively similar.

The mean success rates of eye contact and gender are about the same according to the factor level means plot.

Below I state the model:

i = 1,...,a ; j = 1,...,B ; k = 1,...,n

Yijk = Mij + Eijk

E(Yijk) = Mij = M.. + ai + Bj + (aB)ij

ai is the effect of factor A (Contact) on the cell means as we go from one level to the other (No eye contact to Eye contact and vice versa) holding all other factors constant.

Bi is the effect of factor B (Gender) on the cell means as we go from one level to the other (Male to Female and vice versa) holding all other factors constant.

M.. is the grand mean of the two-way table.

Yijk represents the sample means after adding the effects of Contact and Gender.

Mij represents the population mean after adding the effects of Contact and Gender.

(aB)ij represents an interaction effect between Contact and Gender.

Creating the Models and Plotting:

* We will now create the model of interaction between Gender and Eye Contact.

ECE.aov <- aov(Success~ Contact\*Gender, data=ECE)

* After creating the model, we create a Residuals vs Fitted plot and a Q-Q plot to test normality.

plot(ECE.aov, which = 1:2)

Chart, line chart, scatter chart

Description automatically generated

The Q-Q plot shows a short-tailed distribution.

The ends of the data have a few outliers (labeled as 4, 18, and 19) that could be an issue in the Normal Q-Q plot.

It mostly satisfies the law of constant variance according to the Residuals vs Fitted plot. In this plot, other than the aforementioned outliers the data looks relatively normal.

The model starts to lie somewhat above the line at the values near 0 sd in the Q-Q plot meaning the model may not be the absolute best fit, but it does not invalidate our T and F distributions.

Creating the ANOVA Table.

anova(ECE.aov)

Text

Description automatically generated

Ho: There is no interaction between Contact and Gender

Ha: There is an interaction between Contact and Gender

The F-test statistic for interaction is 0.2058 and interaction has a p-value of 0.656202

Do not reject the null hypothesis. There is no significant interaction between Contact and Gender.

We are able to conclude this because given a significance level of a = 0.05 there is a p-value of 0.656202, above the stated significance level.

* Confidence Intervals using the Bonferroni Method:

Formula for CI of D1: D1 +/- t\_(N-k, 1- a/(2\*k)) \* sqrt(Varhat\_Lhat\_D1)

Formula for CI of D2: D2 +/- t\_(N-k, 1- a/(2\*k)) \* sqrt(Varhat\_Lhat\_D2)

b <- 2 # number of means

a <- 2 # number of means

n <- 5 # number of observations per cell

k <- 2 # number of comparisons

df <- 20-2 # Formula for degrees of freedom: df = N-k

* The formula for D1 = Mu2. - Mu1.

D1 <- ((13.6+16.4)/2) - ((13+9.2)/2)

* The formula for D2 = Mu.2 - Mu.1

D2 <- ((16.4+13.0)/2) - ((13.6+9.2)/2)

* B is the quantile we find using the adjusted probability under Bonferonni P = 1 - (a /2\*k)

B <- qt(1-(.05/(2\*k)), df=18)

* The formula for Varhat\_Lhat\_D1 = (sigma\_hat^2 / (b\*n))\*sum(ci^2)

Varhat\_Lhat\_D1 <- ((6.075) / (b\*n))\*(((-1)^2) + (1)^2)

* The formula for Varhat\_Lhat\_D2 = (sigma\_hat^2 / (a\*n))\*sum(ci^2)

Varhat\_Lhat\_D2 <- ((6.075) / (a\*n))\*(((-1)^2) + (1)^2)

Varhat\_Lhat\_D1 == Varhat\_Lhat\_D2



* since a = b, the variances are equal. Therefore, we can use the same Standard Error for both.
* Formula for Standard Error = sqrt(Varhat\_Lhat\_D2)

SE <- sqrt(Varhat\_Lhat\_D2)

* Margin of Error

ME <- B\*SE

* Confidence Interval for D1:

D1\_CI <- D1 + c(-1, 1)\*ME

* Confidence Interval for D2:

D2\_CI <- D2 + c(-1, 1)\*ME

noquote(paste("Confidence Interval for D1: (", D1\_CI[1], ",", D1\_CI[2], ")"))



noquote(paste("Confidence Interval for D2: (", D2\_CI[1], ",", D2\_CI[2], ")"))



**Conclusion**

We found that a job applicant's eye contact and a personnel officer's gender have an effect on the mean likelihood of job applicant success. No Eye Contact and Female Gender are different from Eye Contact and Male Gender respectively. We also found that there is not a significant interaction between Eye Contact and Gender with Success as the response variable.

A drawback is that it requires us to use a post-hoc test to test for the significance of interaction as well after we initially tested which can be time consuming. Another drawback of the two-way ANOVA table design is that there is no special interpretation for the significance of the means. A useful recommendation since we found that there is a difference may be an additional test using multiple gender identities of personnel officers including those outside of the normal binary to see if there is additional statistical significance.