National Technical
University of Athens

HOMEWORK 4

Kreativstorm Data Analysis Course

Full Name: Antonios Mitsis

Email: anmitsis@hotmail.com

Group: B

RQ1: Is there any evidence to suggest that the cute-direct pick-up approach will lead to more relationship receptivity than the direct-direct approach?

Q1. What is your dependent variable?
A1. Our dependent variable is the receptivity, as we are interested to see whether different pick-up lines affect the value of receptivity.

Q2. What is(are) your independent variable(s)?
A2. Our independent variable is the variable PickUp which has two levels. The first level is the Cute-Direct pick-up line and the second level is the Direct-Direct pick-up line.

Q3. Are the scores in each cell independent?
A3. Yes, the scores in each cell are independent. We can check this by fitting the regression model between the two variables and clicking on the Durbin-Watson statistic. The value of this statistic should be between 1.5 and 2.5 in order to assume independence and in our data the value is 1.864, therefore the scores in each cell are independent.

Q4. Are there any significant outliers?
A4. No, there are no significant outliers in the receptivity variable. We calculate the z-score of this variable by selecting transform variable and subtracting the mean and dividing by the standard deviation of the variable, both values can be found in the explore menu. After that, we look for a z-score with an absolute value greater than 3. There is none as the highest z-score is 2.29 and the lowest is -2.6. So, we conclude that there are no significant outliers. For the PickUp variable there is no point in checking as the variable is a factor and only has two levels.

Q5. How is your dependent variable distributed in each cell?
A5. We can see in the following histogram how our dependent variable is distributed. It looks like a normal distribution, however it is not as we see from the Shapiro-Wilk test that has a significance value of 0.031. Our data also has a small positive skewness, with a value of 0.248.

Histogram

Mean = 3.78
Std. Dev. = .531
N = 194

x-axis: Receptivity
y-axis: Frequency
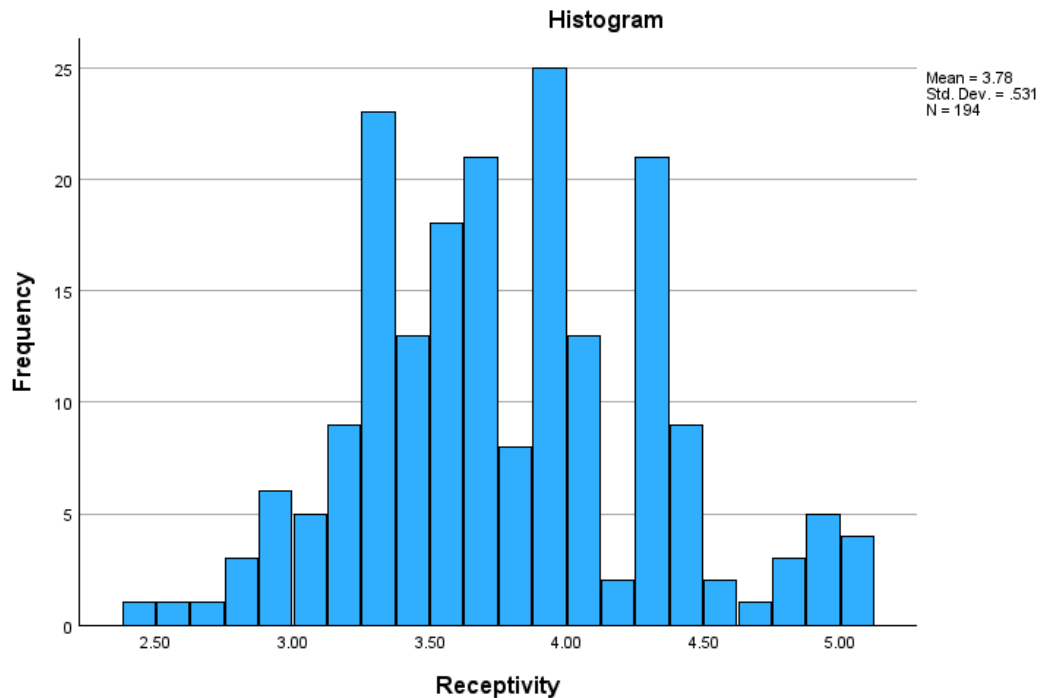
Q6. Do you need to perform any transformations?

A6. Yes, the transformation we need to perform is the square root. We select compute variable and then create a new variable that is the square root of the receptivity. The Shapiro-Wilk statistic value of the new variable is 0.132 so our new data are normally distributed. We can also take the log with base 10 of the variable as well as the inverse but these transformations do not yield the same results in our data.

Q7. Is there homogeneity or heterogeneity of variance?

A7. There is homogeneity of variance. We select Analyze -> Compare means and proportions -> One-Way ANOVA and in the options menu select homogeneity of variance test. The test performed is the Levene's test and our significance value is 0.282 so we accept the null hypothesis, which is that variance in each level is the same, or in other words there is homogeneity of variance.

Q8. What is the mean score of receptivity in the experimental condition?

A8. The mean score of receptivity in the experimental condition is 3.8624. The experimental condition is the cute-direct pickup line. To get this result we select the cases where pickup is equal to 1 and then in the analyze menu select explore.

Q9. What is the mean score of receptivity in the control condition?
A9. The mean score of receptivity in the control condition is 3.6545. The control condition is the direct-direct pickup line. To get this result we select the cases where pickup is equal to 2 and then in the analyze menu select explore.

Q10. What is your answer to RQ1? Report on the findings (no less than 150 words). Don't forget to mention the assumptions.
A10. Yes, there is evidence to suggest that the cute-direct pick-up approach will lead to more relationship receptivity than the direct-direct approach. The method to tell If there is a difference in the means of two groups is the t-test. There are 3 types of t-tests but we can only use the independent samples t-test as there are two experimental conditions and different participants are assigned to each condition. The assumptions for the independent samples t-test are:
1. Normal Distribution in our data.
2. Data are measured at least at the interval level
3. Scores are independent (they come from different people)
4. Homogeneity of variance

On Q5 we saw that our dependent variable was indeed not distributed normally so we performed the transformation and the new transformed variable follows the normal distribution. The measurement of our data is at the interval level, which means that our data have the three following properties. They have order, they have equal and consistent intervals between values, and lastly the value of 0 does not mean absence of measured attribute. In Q3 we found that the scores are independent and the homogeneity of variance is also true in our data, according to Q7. Therefore, we are able to perform the independent samples t-test with dependent variable the transformed receptivity and factor the pickup variable. The statistic of the test is 2.802 with a significance value of 0.006. This means that we can reject the null hypothesis which is that the means of the groups are equal and we accept the alternative hypothesis that means in each group are different. The difference of the means is 0.0555 and it is statistically significant.

RQ2: Is there any evidence to suggest that the presence of androstadienone spray will lead to more relationship receptivity than no spray?

Q11. What is your dependent variable?
A11. Our dependent variable is the receptivity, as we are interested to see whether the presence of androstadienone spray will affect the value of receptivity.

Q12. What is(are) your independent variable(s)?
A12. Our independent variable is the variable scent. It has two levels 1 and 2, where 1 represents spray and 2 is no spray.

Q13. Are the scores in each cell independent?

A13. Yes, the scores in each cell are independent. The Durbin-Watson statistic is 1.809 and is between 1.5 and 2.5 as we want it to be.

Q14. Are there any significant outliers?

A14. Since our independent variable is a factor there is no point in checking for outliers. The outliers for the dependent variable have been checked in Q4 and there are none significant outliers.

Q15. How is your dependent variable distributed in each cell?

A15. The answer is the same as in Q5.

Q16. Do you need to perform any transformations?

A16. The answer is the same as in Q6.

Q17. Is there homogeneity or heterogeneity of variance?

A17. There is heterogeneity of variance in our data. This means that the variances in the two groups (spray and no spray) are significantly different. We reached this conclusion with the Levene's test that has a significance value of less than 0.01. This means that we reject the null hypothesis than variances between groups are equal.

Q18. What is the mean score of receptivity in the experimental condition?

A18. The mean score of receptivity in the experimental condition is 3.8569. The experimental condition is the presence of androstadienone spray. We select Data -> Select cases and the condition is if scent = 1. After that we select explore to get the mean value of receptivity.

Q19. What is the mean score of receptivity in the control condition?

A19. The mean score of receptivity in the control condition is 3.6992. The control condition is the absence of androstadienone spray and to get the mean value the condition is scent = 2 and then explore to get the value.

Q20. What is your answer to RQ2? Report on the findings (no less than 150 words). Don't forget to mention the assumptions.

A20. Yes, there is evidence to suggest that the presence of androstadienone spray will lead to more relationship receptivity. The method to tell If there is a difference in the means of two groups is the t-test. There are 3 types of t-tests but we can only use the independent samples t-test as there are two experimental conditions and different participants are assigned to each condition. The assumptions for the independent samples t-test are:
1. Normal Distribution in our data.
2. Data are measured at least at the interval level
3. Scores are independent (they come from different people)

4. Homogeneity of variance

On Q15 we saw that our dependent variable was indeed not distributed normally so we performed the transformation and the new transformed variable follows the normal distribution. The measurement of our data is at the interval level, which means that our data have the three following properties. They have order, they have equal and consistent intervals between values, and lastly the value of 0 does not mean absence of measured attribute. In Q13 we found that the scores are independent and there is heterogeneity of variance in our data, according to Q17. When we perform the independent samples t-test it is important to look for the significance value that represents equal variances not assumed. The value of the statistic is 2.134 and the significance value is 0.034 which is smaller than 0.05. Therefore, we procced to reject the null hypothesis that the mean values on each group are equal and we accept the alternative hypothesis that the mean values are different, with the mean value with the presence of spray being bigger. So, in conclusion we are able to say that the presence of androstadienone spray will lead to more relationship receptivity.

RQ3: Is there any evidence to suggest that the impact of the androstadienone spray on attractiveness effect will be enhanced by the pick-up approach?

Q21. What is your dependent variable?
A21. The dependent variable is receptivity, as we are interested to see how it's value changes with the effects of pick-up lines and androstadienone spray.

Q22. What is(are) your independent variable(s)?
A22. The independent variables are two, PickUp and Scent. Both are factors of two levels and the levels are cute-direct approach for 1 in pickup and direct-direct approach for 2 in pickup and spray for 1 in scent and no spray for 2 in scent.

Q23. Are the scores in each cell independent?
A23. Yes, the scores in each cell are independent. To get this answer we fit the regression model based on our dependent and independent variables in the regression menu and select Durbin-Watson statistic in the options menu. The value of the statistic is 1.866 and it is between 1.5 and 2.5 so we can say that the scores are independent.

Q24. Are there any significant outliers?
A24. No, there are no significant outliers in the receptivity variable. We calculate the z-score of this variable by selecting transform variable and subtracting the mean and dividing by the standard deviation of the variable, both values can be found in the explore menu. After that, we look for a z-score with an absolute value greater than 3. There is none as the highest z-score is 2.29 and the lowest is -2.6. So, we conclude that there are no significant outliers. For the PickUp

and the scent variables there is no point in checking as the variables are factors and only have two levels.

Q25. How is your dependent variable distributed in each cell?
A25. The answer is the same as Q5.

Q26. Do you need to perform any transformations?
A26. The transformed variable is the same as in Q6 and it is the square root of receptivity.

Q27. Is there homogeneity or heterogeneity of variance?
A27. By performing Levene's test to test for homogeneity of variance among the groups we see that the significance value is 0.02 so we would say that there is heterogeneity of variance. However, since the samples sizes are equal and large, we can check if the ratio of the largest group variance to the smallest is less than 3 and then we can say that the two-way ANOVA is somewhat robust to heterogeneity of variance. The largest standard deviation is 0.15428 so the largest variance is 0.0238 and the smallest standard deviation is 0.08905 so the smallest variance is 0.0079. The ration of these two values is 3.001. The value is right on the limit so we could both say that the ANOVA is and is not robust to heterogeneity of variance. We will assume that our answers are somewhat reliable so we can answer the research question.

Q28. Is there any interaction between the two factors?
A28. Yes, there is significance interaction between the two factors. In the ANOVA table toe interaction coefficient is 0.147 and its significance value is 0.04, meaning that we can reject the null hypothesis, which is that the true value of the interaction is 0 and accept the alternative hypothesis that the value of interaction is significant and not equal to 0.

Q29. What is your answer to RQ3? Report on the findings (no less than 200 words). Don't forget to mention the assumptions. You can use η^2 instead of ω^2.
A29. Yes, there is evidence to suggest that the impact of the androstadienone spray on attractiveness effect will be enhanced by the pick-up approach. In order to answer this question, we need to perform a two0way ANOVA so we can see the impact each factor has on the dependent variable as well as the interaction of the two independent variables. The assumptions for the two-way ANOVA are:
1. No significant outliers
2. Independent scores in each cell
3. Normal distribution of dependent variable
4. Homogeneity of variance of dependent variable

All conditions are satisfied as answered before and even though we do not have homogeneity of variance, we assumed our ANOVA model to be robust enough so it can produce reliable data. Continuing, to fit the ANOVA model we select Analyze -> General Linear Model -> Univariate and we input our dependent and independent variables. We also select the plots of scent*pickup and

pickup*scent, the estimates of effect sizes and descriptive statistics. The significance value of the ANOVA is less than 0.01 so it is significant. The interaction value as answered before is also significant and has a positive value, this means that the impact of the spray is greater when the best pickup approach is chosen. The η^2 for the interaction is 0.044 which is small to medium effect. Another way to check for the interaction is with the interaction plots we selected. There is significance interaction between the factors if the lines are intersecting and that is what is happening in the following plots.

Estimated Marginal Means of sqrt_recept

Estimated Marginal Means of sqrt_recept
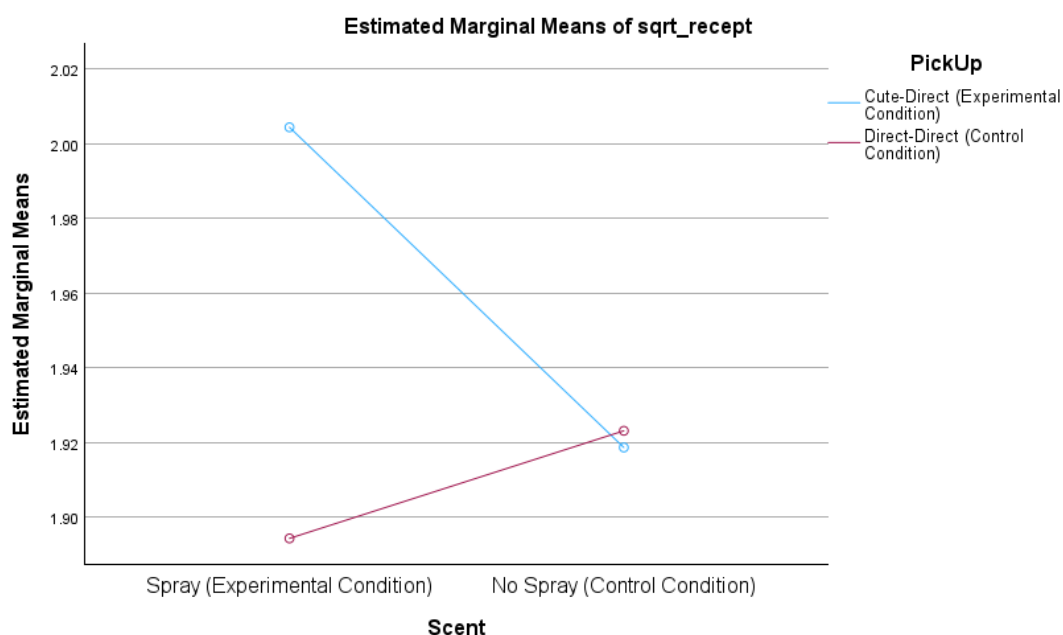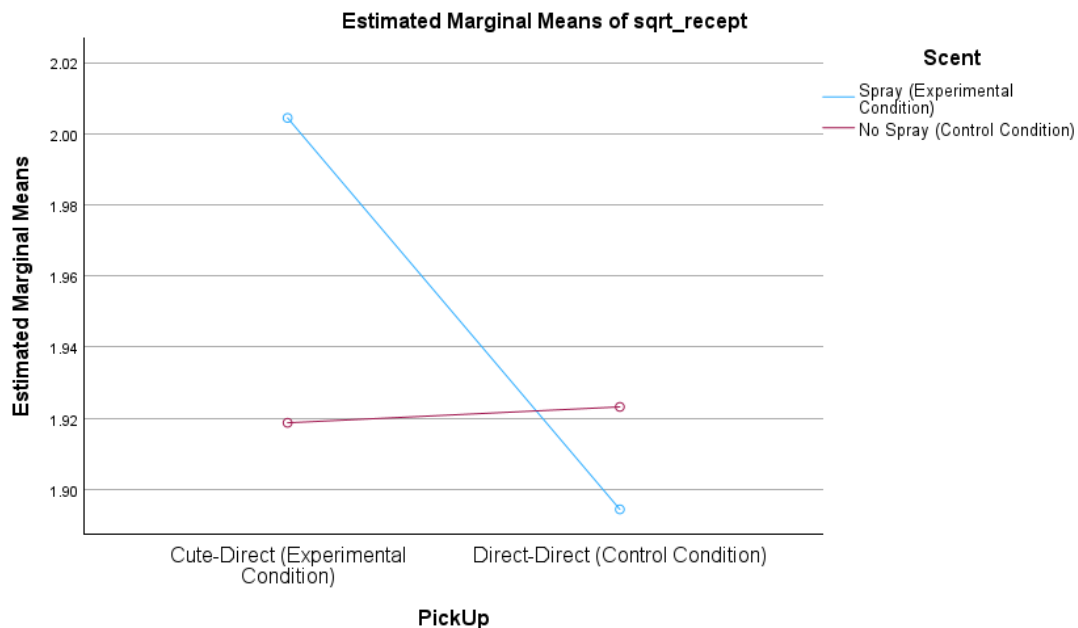
Q30. Answer the main research question by taking the above findings into account (no less than 200 words). Don't forget to mention the assumptions.
RQ: How do pick-up lines and a person's scent influence relationship initiation?

Answer.

A two-way ANOVA was conducted to examine the effects of pickup and androstadienone spray on relationship initiation. Residual analysis was performed to test for the assumptions of the two-way ANOVA. Outliers were assessed by z-scores, normality was assessed using Shapiro-Wilk's normality test for each cell of the design and homogeneity of variances was assessed by Levene's test. There were no outliers, residuals were normally distributed (p > .05) after the square root transformation and there was heterogeneity of variances (p = 0.02), however the ANOVA was considered to be robust due to large sample size.

There was a statistically significant interaction between pickup and scent on receptivity, F =8.669 p = 0.04, partial η^2 =0 .44. We were not able to perform any post-hoc comparisons as the levels of the factors were too few. We also performed two independent samples t-test to test for the effect of each factor alone and both had a significant difference between the groups. The assumptions for the t-tests were the same and homogeneity of variance was assumed.

What was concluded from the three tests is that the cute-direct pickup approach has increased relationship receptivity as well as the presence of androstadienone spray. Moreover, the cute-direct pickup approach increases the impact of the spray as the interaction of the two variables is positive and significant. The significance level for all the models were below 0.05 so all the conclusions made are reliable.