Main Research Question: How do pick-up lines and a person's scent influence relationship initiation?
Dataset: "PickUpLines.sav"

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?

'PickUp': {1.0: 'Cute-Direct (Experimental Condition)',
2.0: 'Direct-Direct (Control Condition)'
'Scent': {1.0: 'Spray (Experimental Condition)',
2.0: 'No Spray (Control Condition)'
'Gender': {1.0: 'male', 2.0: 'female'}
'Ethnicity': {1.0: 'eastern or asian',
2.0: 'black or african',
3.0: 'hispanic, latino, spanish',
4.0: 'white or european',
5.0: 'other'}
'YearInSchool': {1.0: 'freshman', 2.0: 'sophomore', 3.0: 'junior', 4.0: 'senior'
'Contraceptive': {1.0: 'No', 2.0: 'Yes'}
'filter_$': {0.0: 'Not Selected', 1.0: 'Selected'}

1. What is your dependent variable?
   It's receptivity.

2. What is(are) your independent variable(s)?
   Pick-up method.

3. Is there independence of observations?
   Yes, it is. Because all of the Participants are unique people.

4. Are there any significant outliers?
   Using the box plot we found no significant outliers.

5. How is your dependent variable distributed in each cell?
   Receptivity with the direct-direct approach is normally distributed.
   Shapiro-Wilk test shows pvalue: 0.87 which is > then 0.05

   Receptivity with the cute-direct approach isn't normally distributed.
   Shapiro-Wilk test shows pvalue: 0.00085 which is < then 0.05
   The data is positively skewed. We should try to transform it.

6. Do you need to perform any transformations?
   Let's try the log10 transformation to normalize the data. P-value: 0.019 which is not enough to consider it as a normal distribution.
   We can try box-cox transformation. P-value: 0.06 > 0.05
   Now cute-direct data is normally distributed.

7. Is there homogeneity or heterogeneity of variance?
   We also transform direct-direct values with box-cox technique.
   With the Levene's test P-value: 0.0001 < 0.05 we found heterogeneity of variance.

8. What is the mean score of receptivity in the experimental condition?
   3.86

9. What is the mean score of receptivity in the control condition?
   3.65

10. What is your answer to RQ1? Report on the findings (no less than 150 words). Don't forget to mention the assumptions.
    To sum up:
    - The sampling distributions are normally distributed; t-test/non-parametric t-test
    - Data are continuous; t-test/non-parametric t-test
    - Scores are independent as they come from unique people; t-test/non-parametric t-test
    - Variances in these populations aren't equal – heterogeneity of variance (Levene's test < 0.05). t-test/non-parametric t-test

One of the assumptions to conduct independent mean t-test is violated, so it's better to use non-parametric Mann Whitney test.

H_0: No significant difference between two samples
H_1: Significant difference between two samples

P-value: 0.01 < 0.05 => we accept the null hypothesis
As a result: We **didn't find evidence** to say that the cute-direct pick-up approach will lead to more relationship receptivity than the direct-direct approach.

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

11. What is your dependent variable?
    It's the receptivity.

12. What is(are) your independent variable(s)?
    Scent.

13. Is there independence of observations?
    Yes, it is. Because all of the Participants are unique people.

14. Are there any significant outliers?
    There is one significant outlier in no spray group.

We can drop it, because it'll ruin group comparison.

15. How is your dependent variable distributed in each cell?
    Receptivity with the spray isn't normally distributed.
    Shapiro-Wilk test shows pvalue: 0.0364 which is < then 0.05
    The data is positively skewed coef: 0.08
    Kurtosis : -0.75

    Receptivity without the spray isn't normally distributed.
    Shapiro-Wilk test shows pvalue: 0.0362 which is < then 0.05
    The data is positively skewed coef: 0.16
    Kurtosis: 0.74

    Taking into account that each cell isn't distributed normally we can close our eyes on it because the distributions are similarly skewed. Will consider it as normal.

16. Do you need to perform any transformations?
    We consider that cells are normally distributed, so it is not necessary to apply transformation here.

17. Is there homogeneity or heterogeneity of variance?
    Levene's test p-value: 0.0004 < then 0.05
    We found heterogeneity of variance.

18. What is the mean score of receptivity in the experimental condition?
    3.86

19. What is the mean score of receptivity in the control condition?
    3.70

20. What is your answer to RQ2? Report on the findings (no less than 150 words). Don't forget to mention the assumptions.
    To sum up:
    - The sampling distribution is approximately normally distributed. It has similar skewness, so that assumption is quite a truth; t-test/non-parametric t-test
    - Data are continuous; t-test/non-parametric t-test
    - Scores are independent as they come from unique people; t-test/non-parametric t-test
    - Variances in these populations aren't equal – heterogeneity of variance (Levene's test < 0.05). t-test/non-parametric t-test

One of the assumptions to conduct independent mean t-test is violated, so it's better to use non-parametric Mann Whitney test.

H_0: No significant difference between two samples
H_1: Significant difference between two samples

P-value: 0.03 < 0.05 => we accept the null hypothesis

As a result: We **didn't find evidence** to say that the presence of androstadienone spray will lead to more relationship receptivity than no spray.

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?

21. What is your dependent variable?
    Attractiveness (Receptivity).

22. What is(are) your independent variable(s)?
    Spray, pick-up method.

23. Is there independence of observations?
    Yes, it is. Because all of the Participants are unique people.

24. Are there any significant outliers?
    There are no significant outliers in any group.

25. How is your dependent variable distributed in each cell?
    In **direct-direct with spray** group p value: 0.12 > 0.05 => normal
    In **cute-direct with spray** group p value: 0.06 > 0.05 => normal
    In **direct-direct with no spray** group p value: 0.44 > 0.05 => normal
    In **cute-direct with no spray** group p value: 0.06 > 0.05 => normal

26. Do you need to perform any transformations?
    I've tried log, boxcox transformation but it didn't work. Data still was not normal. Leave everything as it is.

27. Is there homogeneity or heterogeneity of variance?
    Levene's test p value: 0.0007 < than 0.05 => there is heterogeneity of variance.
    BUT:
    The biggest group is **cute-direct with no spray** with 60 samples (std = 0.34), and the smallest is **direct-direct with no spray** with 30 samples (std = 0.55). Their variance ratio is 0.61 which is less than 3. Because of that we can ignore heterogeneity of variance.

N.B. If group sample sizes are equal or approximately equal and large, there is normality and the ratio of the largest group variance to the smallest group variance is less than 3, the two-way ANOVA is somewhat robust to heterogeneity of variance in these circumstances (Jaccard, 1998).
Reminder: Standard deviation is square root of variance ($SD = \sqrt{Variance}$).

28. Is there any interaction between the two factors?

	df	sum_sq	mean_sq	F	PR(>F)
C(PickUp)	1.0	1.999082	1.999082	7.821627	0.005693
C(Scent)	1.0	1.515279	1.515279	5.928695	0.015820
C(PickUp):C(Scent)	1.0	2.329234	2.329234	9.113383	0.002886
Residual	190.0	48.560936	0.255584	NaN	NaN

P value for interaction PickUp and Scent: 0.003 < 0.05 => there is significant interaction between these two factors.

29. 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.
    To sum up:
    - no significant outliers in our groups;
    - independence of observations;
    - normality of dependent variable in groups;
    - heterogeneity of variance, but we proved that it's not critical in that case.

Here we can use two-way ANOVA to determine if there is significant interaction between two factors (PickUp, Scent). It is.

Conducting one-way ANOVA we got p value: 0.00001 < 0.05 => one of the groups is significantly different from others. Using Dunn's post hoc test we got:

	1	2	3	4
1	1.000000	0.000373	0.505985	0.505266
2	0.000373	1.000000	0.014106	0.001664
3	0.505985	0.014106	1.000000	0.910111
4	0.505266	0.001664	0.910111	1.000000

Second group (cute-direct with perfume) is significantly different from others. Means are 1: 3.61, **2: 4.03**, 3: 3.71, 4: 3.68

30. Answer the main research question by taking the above findings into account (no less than 200 words). Don't forget to mention the assumptions.

We didn't get evidence that pick-ups methods and spray are influence relationship initiation separately. But together that variables make impact on relationship initiation. The cute-direct method with perfume is the best way to initiate a relationship. To prove that, we firstly checked if pick-up methods are significantly important to get higher rates with non-parametric Mann Whitney test, it was insignificant. Secondly we checked if the presence of perfume affects relationship success with non-parametric Mann Whitney test, it was insignificant too. Thirdly, we checked if the impact of spray on attractiveness effect is enhanced by the pick-up approach, we got that there is significant interaction between these two factors with a two-way ANOVA test. Eventually, we found the best combination (cute-direct with perfume) with a one-way ANOVA test.