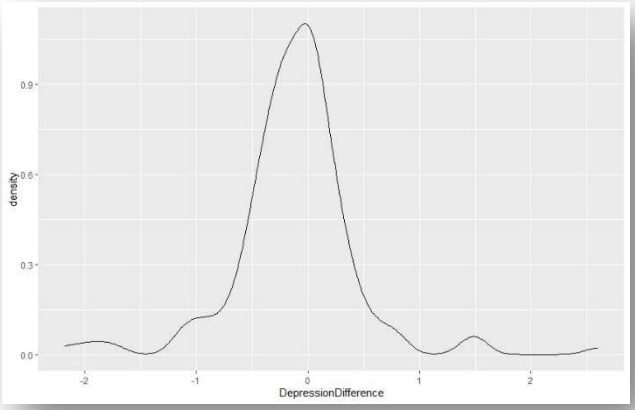
Intro to Applied Stats for Grads
Module 5 Group Work
Stewart Kabis

1. Use this new difference variable to answer the following questions:
    a. Plot your data using the geom_density() ggplot function. Does it look normal?



Yes, the data look like they are normally distributed based only on this density plot.

2. Using your descriptive statistics skills in R, fully describe this "difference" data (include whatever metrics you think are appropriate to fully describe the nature of this data).
    - Mean = -0.118
    - SD = 0.553
    - Variance = 0.305
    - IQR = 0.476
    - Skewness = 0.208
    - Kurtosis = 5.981

3. Is this data normally distributed? Make note of the mean and standard deviation for this data. You will need it moving forward.

    Mean = -0.118, SD = 0.553

    The data are not normally distributed. Results of the Shapiro Wilks test P value = 9.23E-10. The null hypothesis is that the test data are normally distributed. If the p value is less than or equal to 0.05, we can reject the null hypothesis. In this case, the p value is less than 0.05, so we can conclude that these data are not normally distributed.

4. Now use your data description results to answer the following questions:
    a. What is the probability of seeing a person in the general population with an increase of more than 1 depression unit (i.e. measured value (X) for your difference column is p(X>1).

pnorm(1, mean=-0.118, sd=0.553, lower.tail= FALSE)

Pnorm function in R returns 0.0216, or 2.16% chance.

b. Based on this data, what is the probability of seeing someone decrease in their depression rating (i.e. measured value (X) for your difference column is p(X< 0).

pnorm(0, mean=-0.118, sd=0.553, lower.tail= TRUE)
Pnorm function in R returns 0.584, or a 58.4% chance.

c. Now, let's consider the power of this test. If we wanted to flag people whose depression increased significantly (using a 1-tailed 0.05 alpha threshold) what is the depression value we would use as the threshold to flag people for additional follow-up?

Using the power t test in R, setting n=162, sig.level=0.05, power = 0.8, type=one sample; it returns a value of d = 0.2214079.

$$d = \frac{|\mu_1 - \mu_2|}{\sigma}$$

$\sigma = 0.553$ and $\mu_1 = -0.118$, solve for μ_2.
$\mu_2 = -0.240$. Therefore, the threshold we should use is a change in depression (since that is what the DepressionDifference column is showing) of -0.240.