

```
## Error in library(scatterplot3d): there is no package called 'scatterplot3d'  
## Error in library(car): there is no package called 'car'  
## Error in library(plotly): there is no package called 'plotly'  
## Error in eval(expr, envir, enclos): could not find function "plotly"
```

1 Intro

Dyadic data are often used by family researchers who are interested in comparing the attitudes, behaviors, and opinions of husbands and wives, parents and children, and romantic partners towards each other. The dyadic structure can come from inter-individual reporting (e.g. comparison of husband's report of his support toward his wife to the wife's report of her support towards the husband), or it can come from intra-individual reporting (e.g. comparison of an observer's report of husband's hostility to an observer's report of the wife's hostility).

One of the most common methods in analyzing dyadic data is difference scores. Difference scores are used to see how well two variables/components 'fit' together. Some of the most common difference scores use the algebraic, absolute, and squared difference between the two variables as shown by Equations 1, 2, and 3 respectively. These difference scores are often modeled to predict for an outcome variable labeled Z in the equations below,

$$Z = \beta_0 + \beta_1(X - Y) + \epsilon \quad (1)$$

$$Z = \beta_0 + \beta_1|X - Y| + \epsilon \quad (2)$$

$$Z = \beta_0 + \beta_1(X - Y)^2 + \epsilon \quad (3)$$

There are many methodological issues associated with difference scores. Dr. Edwards describes many of these issues in his paper *Problems with the Use of Profile Similarity Indices in the Study of Congruence in Organizational Research* (Edwards 1993) [2]. Perhaps the most important reason why statisticians avoid using difference scores is that it makes it difficult to identify the underlying mechanism. This happens because in using difference scores, two distinct measures are turned into one entity so it becomes difficult to tease out the effect of one variable relative to another. These individual effects are what we are interested in. Difference scores fail to tell the researcher which variables contribute the most

to the response. We also lose information when using difference scores. There are assumptions used with difference scores namely, that the difference between two entities to the outcome variable are assumed to be symmetric and that the outcome is constant at all points where the two entities are equal. These are assumptions that we need to look at by examining the absolute level of both measures and its direction.

In his paper, *On the Use of Polynomial Regression Equations as an Alternative to Difference Scores in Organizational Research*, Dr. Edwards suggested using "polynomial regression equations containing the...[two variables]...composing the difference and higher-order terms." (Edwards 1993)[1] as an alternative to difference scores. An example would be to take the two predictor variables (X, Y) , their squares (X^2, Y^2) , and their products (XY, X^2Y, XY^2) . A successful application of polynomial regression is presented by Phillips, Diefenbach, Kronish, Negron, and Horowitz (2014) in their paper titled *The Necessity-Concerns Framework: a Multidimensional Theory Benefits from Multidimensional Analysis*[5]. They use polynomial regression to assess stroke survivors' concerns about medication and their necessity beliefs effect on their adherence to medication whose goal is stroke prevention. Before looking at the polynomial model Edwards proposed, we will look at some components which are not common to normal linear regression: the model form, stationarity points, the principal axis, and the predictive surface (which we will look through 3D graphs). Polynomial regression equations avoid many of the issues with difference scores, but they do have their own set of problems. Mainly these problems have to do with interpreting the estimates of the model and interpreting two components we have yet to define: stationarity and principal axis.

2 The Model

The linear regression model (Eq 4) is familiar to most researchers; it is often used in research due in part to the ease and familiarity of the interpretations of the parameters and the ease of computation. A slight tweak to simple linear models is to include two explanatory variables as in Eq . Often these models also include their interactions e.g $z = a + bx + cy + dxy$. We call Eq the linear difference model; this particular model does not contain an interaction.

$$Z = \beta_0 + \beta_1 X + \epsilon \quad (4)$$

$$Z = \beta_0 + \beta_1 (X - Y) + \epsilon \quad (5)$$

$$Z = \beta_0 + \beta_1 X + \beta_2 Y + \beta_3 X^2 + \beta_4 Y^2 + \beta_5 XY + \epsilon \quad (6)$$

An extension of the linear regression model is the quadratic regression model (Eq). It's form is similar to the form of a linear regression model with the addition of a quadratic term of the explanatory variable. It is very similar to the linear model in terms of computation and ease of interpret-ability. In this project, we consider a composition of these two ideas; a polynomial(quadratic) regression using two predictors and the interaction (Eq 2). We will be using the basic framework of response methodology to look two important concepts at stationary points and principal axes. Using the response surface methodology will allow us to take advantage of some key properties developed by that field. So looking how to develop the stationary points and principal axes will be key because they will be very useful in making inference on the predicted surface. In the pages that follow, we demonstrate how this framework has been used and then apply these methods our own data.

2.1 Form

Let us consider a model with 2 explanatory variables and a single response. We can write the model as follows:

$$z = \beta_0 + \beta_1x + \beta_2y + \beta_3x^2 + \beta_4xy + \beta_5y^2 + \epsilon \quad (7)$$

And the fitted model:

$$\hat{z} = b_0 + b_1x + b_2y + b_3x^2 + b_4xy + b_5y^2 \quad (8)$$

The fitted model in matrix notation is:

$$\hat{z} = b_0 + \mathbf{d}'\mathbf{b} + \mathbf{d}'\mathbf{B}\mathbf{d} \quad (9)$$

where

$$\mathbf{d} = \begin{bmatrix} x \\ y \end{bmatrix} \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} b_3 & b_4/2 \\ b_4/2 & b_5 \end{bmatrix}$$

For further derivations, we will use matrix notation since it condenses the algebra.

We can fit a model like this in R using the `lm()` function. This setup corresponds to Eq [?] where Z is the response variable, and X and Y are the explanatory variables.

```
QuadFit <- lm(Z ~ X + Y + I(X^2) + I(X*Y) + I(Y^2), data=d)
```

2.2 Analyze/Properties

2.2.1 Stationary Points: How and Why?

One of the features of our model (Eq 2) we are interested in are the stationary points. The stationary points are the points where the slope of the response is zero no matter which direction you take the derivative. Recall how to find the minimum and maximum of a function from calculus; it is a similar concept except now we have a three dimensional surface rather than just a two-dimensional surface. We are interested in the stationarity points because they allow us to see which values of our explanatory variables provide the “best” fit for the response (“best” meaning minimum or maximum). The benefit of knowing the stationary points is it gives us a goal to reach or not reach depending on the situation. We will see examples of the usefulness of stationary points later. Using our intuition from calculus, we take the derivatives of Equation 8 with respect to x and y , set the derivatives equal to zero, and solve for x and y in terms of \mathbf{b} and \mathbf{B} to find the stationarity points (we’ll refer to these points as x_0 and y_0).

To find the stationarity points, let’s begin with taking the partial derivatives with respect to x and y

$$\begin{bmatrix} \frac{dz}{dx} = b_1 + 2b_3x + b_4y \\ \frac{dz}{dy} = b_2 + b_4x + 2b_5y \end{bmatrix} = \mathbf{b} + 2\mathbf{B}\mathbf{d}$$

Next, we set the derivatives to zero and solve for \mathbf{d}_0 , where \mathbf{d}_0 is:

$$\begin{aligned} \mathbf{d}_0 = \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} &= -\frac{\mathbf{B}^{-1}\mathbf{b}}{2} \\ &= -\frac{1}{2} \frac{1}{b_5b_3 - \frac{b_4^2}{2}} \begin{bmatrix} b_5 & -b_4/2 \\ -b_4/2 & b_3 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} \\ &= -\frac{2}{4b_5b_3 - b_4^2} \begin{bmatrix} b_5b_1 - b_2b_4/2 \\ -b_1b_4/2 + b_2b_3 \end{bmatrix} \\ &= \begin{bmatrix} \frac{b_2b_4 - 2b_5b_1}{4b_5b_3 - b_4^2} \\ \frac{b_1b_4 + 2b_2b_3}{4b_5b_3 - b_4^2} \end{bmatrix} \end{aligned}$$

So our stationary points are

$$x_0 = \frac{b_2 b_4 - 2b_5 b_1}{4b_5 b_3 - b_4^2} \quad (10)$$

$$y_0 = \frac{b_1 b_4 + 2b_2 b_3}{4b_5 b_3 - b_4^2} \quad (11)$$

These points represent the values in our predictors that will optimize our response. Optimizing the response either means a minimum or maximum depending on the surface shape.

Using our output from the `lm` function noted earlier in R, we can get the stationary points. Note we can use other computing programs as well as R (e.g. SAS, MPlus, etc.) The R code is as follows:

```
statpts <- function(lm){
  # Takes a quadratic lm function
  # coef(lm) gives a vector of all the parameter estimates in the linear model
  # So here we are grabbing the individual parameter estimates from that vector
  b0 <- as.numeric(coef(lm)[1])
  b1 <- as.numeric(coef(lm)[2])
  b2 <- as.numeric(coef(lm)[3])
  b3 <- as.numeric(coef(lm)[4])
  b4 <- as.numeric(coef(lm)[5])
  b5 <- as.numeric(coef(lm)[6])

  # Stationary Pts. using the formulas in Eq 10 & 11
  x0 <- (b2*b4 - 2*b1*b5)/(4*b3*b5 - b4^2)
  y0 <- (b1*b4 - 2*b2*b3)/(4*b3*b5 - b4^2)

  # Output
  out <- matrix(c(x0, y0), ncol=2)
  colnames(out) <- c("x0", "y0")
  out <- data.frame(x0=x0, y0=y0)
  return(out)
}
```

2.2.2 Predicted Response at Stationary Points

Once we have the stationary points, we can get the predicted responses at the stationary points by simply plugging in the stationary points into Equation 8. The predicted response at the stationary points give us the optimized value for the response variable i.e. where the response surface reaches its maximum or minimum. In matrix notation, the predicted response is given as:

$$\begin{aligned}\hat{z}_0 &= b_0 + \mathbf{d}'_0 \mathbf{b} + \mathbf{d}'_0 \mathbf{B} \mathbf{d}_0 \\ &= b_0 + \mathbf{d}'_0 \mathbf{b} - \frac{\mathbf{d}'_0 \mathbf{B} \mathbf{B}^{-1} \mathbf{b}}{2} \\ &= b_0 + \mathbf{d}'_0 \mathbf{b} - \frac{\mathbf{d}'_0 \mathbf{b}}{2} \\ &= b_0 + \frac{\mathbf{d}'_0 \mathbf{b}}{2}\end{aligned}$$

It is easy to get the predicted response in R:

```
x0 <- statpts(lm)$x0
y0 <- statpts(lm)$y0
predstatpts <- coef(lm)[1] + coef(lm)[2]*x0 + coef(lm)[3]*y0 + coef(lm)[4]*x0^2 + coef(lm)[5]*x0*y0 +
predstatpts
```

2.2.3 Principal Axes

The purpose of principal axes is to make interpretations of the model easier by rotating the axes so that we can remove all the cross-product terms. Khuri and Cornell (1996), *Response Surfaces: Designs and Analyses* [3], lays out the process of deriving the canonical equations. The canonical equations come from transforming the response system to its canonical form (this is the name of the form where the surface lives with its cross-product terms are removed); this process occurs because it represents the response centered around the stationarity points. These canonical equations are then used in finding the principal axes. In order to complete this process, new axes are created which the surface will be on. This maybe useful in some applications, but the problem is it becomes difficult to relate the response to our original variables. Essentially it becomes difficult to interpret how our original variables are affecting the response. So we focus on the process of transforming these canonical equations back to functions of the original variables i.e. transform our principal axes back to the original variables. By doing this we get

the benefits of the principal axes related to our original variables.

To solve for principal axes, we need to first solve for eigenvalues. The equations for solving for the eigenvalues comes from Khuri and Cornell (1996) and is as follows:

$$\begin{aligned}
|\mathbf{B} - \lambda \mathbf{I}| = 0 &\Rightarrow \left| \begin{bmatrix} b_3 & b_4/2 \\ b_4/2 & b_5 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} \right| = 0 \\
&\Rightarrow \left| \begin{bmatrix} b_3 - \lambda & b_4/2 \\ b_4/2 & b_5 - \lambda \end{bmatrix} \right| = 0 \\
&\Rightarrow \det \begin{pmatrix} b_3 - \lambda & b_4/2 \\ b_4/2 & b_5 - \lambda \end{pmatrix} = 0 \\
&\Rightarrow (b_3 - \lambda)(b_5 - \lambda) - \left(\frac{b_4}{2}\right)\left(\frac{b_4}{2}\right) = 0 \\
&\Rightarrow b_3b_5 - b_3\lambda - b_5\lambda + \lambda^2 - \frac{b_4^2}{4} = 0 \\
&\Rightarrow \lambda^2 - (b_3 + b_5)\lambda + \left(b_3b_5 - \frac{b_4^2}{4}\right) = 0 \\
&\Rightarrow \lambda = \frac{b_3 + b_5 \pm \sqrt{(b_3 + b_5)^2 - 4\left(b_3b_5 - \frac{b_4^2}{4}\right)}}{2}
\end{aligned}$$

Now let λ_1 be the first eigenvalue:

$$\begin{aligned}
\lambda_1 &= \frac{(b_3 + b_5) + \sqrt{(b_3 + b_5)^2 - 4\left(b_3b_5 - \frac{b_4^2}{4}\right)}}{2} \\
&= \frac{(b_3 + b_5) + \sqrt{(b_3 + b_5)^2 - 4b_3b_5 + b_4^2}}{2} \\
&= \frac{(b_3 + b_5) + \sqrt{(b_3 - b_5)^2 + b_4^2}}{2}
\end{aligned}$$

and λ_2 be the first eigen value:

$$\begin{aligned}
\lambda_2 &= \frac{(b_3 + b_5) - \sqrt{(b_3 + b_5)^2 - 4\left(b_3b_5 - \frac{b_4^2}{4}\right)}}{2} \\
&= \frac{(b_3 + b_5) - \sqrt{(b_3 + b_5)^2 - 4b_3b_5 + b_4^2}}{2} \\
&= \frac{(b_3 + b_5) - \sqrt{(b_3 - b_5)^2 + b_4^2}}{2}
\end{aligned}$$

We will define a matrix \mathbf{M} whose columns are the egienvectors associated with λ_1 and λ_2 :

$$\mathbf{M} = \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix} = \begin{bmatrix} \mathbf{m}_1 & \mathbf{m}_2 \end{bmatrix}$$

To solve for the elements of \mathbf{m}_i we use eigen vector associated with λ_i and then normalize it s.t. $\mathbf{m}_i' \mathbf{m}_i = 1$. Here we complete that process using λ_1 :

$$(\mathbf{B} - \lambda_1 \mathbf{I}) \mathbf{m}_1 = \mathbf{0}$$

$$\begin{bmatrix} b_3 - \lambda_1 & b_4/2 \\ b_4/2 & b_5 - \lambda_1 \end{bmatrix} \begin{bmatrix} m_{11} \\ m_{21} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

\Rightarrow

$$b_3 m_{11} - \lambda_1 m_{11} + b_4 m_{21}/2 = 0$$

$$b_4 m_{11}/2 + b_5 m_{21} - \lambda_1 m_{21} = 0$$

Next, we need to normalize it i.e. $m_{11}^2 + m_{21}^2 = 1$. A simple way of doing this it to set $m_{11} = 1$ and solve for m_{21} :

$$\begin{aligned} m_{21} &= \frac{b_4}{2} \div (b_5 - \lambda_1) \\ &= \frac{b_4}{2(b_5 - \lambda_1)} \\ &= \frac{b_4}{2b_5 - b_3 - b_5 - \sqrt{(b_5 - b_3)^2 + b_4^2}} \text{ plug in } \lambda_1 \\ &= \frac{b_4}{b_5 - b_3 - \sqrt{(b_5 - b_3)^2 + b_4^2}} \end{aligned}$$

Now we begin our process of tranforming the principal axes back to the original variables. Let

$z = d - d_0$ and define:

$$\begin{aligned} \mathbf{w} = \mathbf{m}'\mathbf{z} &= \begin{bmatrix} \mathbf{m}_1 \\ \mathbf{m}_2 \end{bmatrix} (\mathbf{d} - \mathbf{d}_0) = \begin{bmatrix} m_{11} & m_{21} \\ m_{12} & m_{22} \end{bmatrix} \begin{bmatrix} x - x_0 \\ y - y_0 \end{bmatrix} \\ &= \begin{bmatrix} m_{11}(x - x_0) + m_{21}(y - y_0) \\ m_{12}(x - x_0) + m_{22}(y - y_0) \end{bmatrix} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \end{aligned}$$

w_1 and w_2 are the variables the canonical equations use so in this next step, we will remove them.

Let $w_1 = 0$ and solve the first equation above:

$$\begin{aligned} m_{11}x - m_{11}x_0 + m_{21}y - m_{21}y_0 &= 0 \\ \rightarrow m_{21}y &= -x + x_0 + m_{21}y_0 \\ \rightarrow y &= -\frac{x}{m_{21}} + \frac{x_0}{m_{21}} + y_0 \\ \rightarrow y &= -P_{21}x + P_{21}x_0 + y_0 \end{aligned}$$

$$y = -P_{21}x + P_{20} \tag{12}$$

$$\text{where } P_{20} = y_0 + P_{21}x_0 \text{ and } P_{21} = \frac{b_5 - b_3 - \sqrt{(b_5 - b_3)^2 + b_4^2}}{b_4}.$$

Now we have transformed our principal axes back to functions of our original variables. This will make them easier to plot and interpret. In a concave surface, the first principal axis (Eq 12) tells us where the surface has the lowest downward curvature. This is where the two explanatory variables create a maximized surface that decreases the least as we move along the line. The second principal axis is the opposite of the first principal axis. The second principal axis is the line in the surface where the outcomes decreases the most. It is the line in the surface where the two explanatory variables create the greatest decrease in the response. The simplest case would be when the first principal axis is parallel to the line in the surface where $X = Y$ and the second principal axis is parallel to the line in the surface where $X = -Y$. In this case, when the variables agree ($X = Y$) the response would be good, and when they do not agree, we would expect the response to behave poorly. If these axes are parallel to these lines, then our intuition about our explanatory variables and their congruence, or lack of, would be correct.

We can easily create a function in R to take the `lm` output to get the principal axis (Eq 12):

```

paxis <- function(lm){
  # Takes a quadratic lm function just as the function before and just as before
  # coef(lm) gives a vector of all the parameter estimates in the linear model
  # So here we are grabbing the individual parameter estimates from that vector
  b0 <- as.numeric(coef(lm)[1])
  b1 <- as.numeric(coef(lm)[2])
  b2 <- as.numeric(coef(lm)[3])
  b3 <- as.numeric(coef(lm)[4])
  b4 <- as.numeric(coef(lm)[5])
  b5 <- as.numeric(coef(lm)[6])

  # Stationary Pts.
  x0 <- (b2*b4 - 2*b1*b5)/(4*b3*b5 - b4^2)
  y0 <- (b1*b4 - 2*b2*b3)/(4*b3*b5 - b4^2)

  # First Principal axis
  p11 <- (b5 - b3 + sqrt((b3 - b5)^2 + b4^2))/b4 #slope
  p10 <- y0 - p11*x0 #intercept

  # Second Principal axis
  p21 <- (b5 - b3 - sqrt((b3 - b5)^2 + b4^2))/b4
  p20 <- y0 - p21*x0

  # Output
  out <- data.frame(p10=p10, p11=p11, p20=p20, p21=p21)
  return(out)
}

```

Now that we have developed some intuition about the model, we will look at some examples of their usage. First, we will recreate an example of the model being used in a paper then we will apply this methodology to our own data.

3 Recreation of Previous Results

To check our understanding of how to run the model and make sure that our results are consistent with previous works, we reproduced the results in Phillips et al. (2012), *Congruence research in behavioral medicine: methodological review and demonstration of alternative methodology* [4]. In this paper, they were interested in seeing what created good situations in which patients followed through with their doctors orders. To study this they looked at patient-reported adherence a month after the doctor visit and its correspondence to the physician's rating of patient's health and the physician's estimate of how the patient would rate of patient's own health. We'll begin by first outlining the variables used here:

$Z_{1,i}$ – patient-reported adherence a month after the dr. visit for patient i

$Z_{2,i}$ – physician's percieved agreement with patient i on the illness treatment

X_i – physician's rating of patient i 's health

Y_i – physician's estimate of how patient i would rate of their own health

Presented in the paper are two different components of using polynomial regression: Confirmatory and Exploratory. The first component (confirmatory) checks whether or not it is appropriate to use polynomial regression, and the second component (exploratory) is in the name. Essentially exploratory analysis is where you would check hypothesis you have in mind, or just explore the model to see what you can learn.

First discussed is the use of polynomial regression in a confirmatory approach. We are looking to see whether or not we are supposed to use polynomial regression. First, we will check the assumptions of normality for all the models namely looking for points with high leverage and potential influential points. No points showed any major signs of concern although there was one point which consistently had a higher leverage and Cook's D value compared to the rest of the points in the data set. This is not a major issue, but rather something we need to keep in the back of our mind and be aware of while the analysis is continued. Next, in the paper, four criterion are presented which help decide whether using a polynomial surface is appropriate. The four criterion are:

1. Check whether the unconstrained model explained a significant amount of the variance in the data.
2. Check whether the regression coefficients for the unconstrained model are in expected directions and whether they are significant to the model.

3. Check whether if the unconstrained model is significantly better than the constrained model.
4. Check whether if a higher order model i.e. polynomial regression is significantly better than the constrained model.

The first criterion checks the basic linear model assumption which is whether the model is better than a basic mean(intercept) only model. The second criterion checks whether the coefficients are lined up in the way we expect them too and whether they add anything to the model. The fourth criterion is intertwined with the second in that they check whether a simple model or polynomial model should be used. It can very well happen that maybe the variable are significant but none of the higher order terms are in which case the response surface would be a plane. The third criterion checks to see if there is benefit in moving from the simpler difference models to more complex polynomial models.

We have replicated Table 2 found in Phillips et al. (2012) in Table 1 to help with checking these assumptions. We can see from Table 1 when we compare to the algebraic difference score that under the unconstrained model both outcomes explained a significant amount of variation in the data; so criterion 1 was satisfied for both outcomes. However, criterion 2 was only satisfied for the outcome dealing with patient adherence as the regression coefficients for physician perceptions were not going in opposite directions and were not both significant. Criterion 3 and 4 were only met for the second outcome i.e. the physician perceptions. So by these criteria, we may consider using polynomial regression for the outcome dealing with physician perceptions while it looks as if the model using patient adherence as its outcome will be sufficiently explained using an algebraic difference score model.

We also checked the criterion using the absolute-value difference score model. Again we can refer to Table 1 when checking the criterion. Like with the difference score model, the unconstrained model explained a significant amount of the variation for both outcomes. Criterion 2 was satisfied for both outcomes. The unconstrained model using the outcome regarding physician perception was the only unconstrained model that explained the variation in the data better than the absolute-value difference score model, but it reversed for the test to see if a higher-order polynomial model (in this case a quadratic model) to explain the variation in the data better. We can conclude that the absolute-value difference score model was not an adequate fit for either of the outcome variables, the difference score is an adequate model for the outcome concerning patient adherence, and that the polynomial regression would be the best model for the data using the physician perception outcome.

Now we will continue the analysis using an exploratory approach. The name implies the type of analysis to be completed; this is used when there is not a specific hypothesis in mind. We can perform change of

Table 1: This table is a replication of the Table 2 in Phillips et al. (2012) with the exception that we are only looking at the outcome regarding patient adherence where $*p < 0.05$; $**p < 0.01$; $***p < .001$

Outcome	Constrained Model		Unconstrained Model				R^2	F_c	F_h
	$(X - Y)$	R^2	X	Y					
Patient Adherence	-0.35299	0.0982957	-0.3336364	0.3918518			0.1043567	0.8391374	0.8290231
Physician perception of patient agreement	-0.0452847	0.0011012	0.0965281	0.0965281			0.1821114	54.8858755	17.3864194
	$ X - Y $	R^2	X	Y	W	WX	WY	R^2	
Patient Adherence	-0.4052581	0.0964611	-0.4587246	0.4578297	0.1433036	0.5721369	-0.4452852	0.1573678	2.1865135
Physician perception of patient agreement	-0.3182014	0.0384911	-0.0635237	0.4426716	0.097664	0.5053487	-0.5696705	0.209204	13.2223343
									7.4520056

sum of squares F-tests to check that we should use a quadratic model. If we test linear model vs the quadratic model, we see that the quadratic model explains the data significantly better than the linear model with a p-value of 0.0065404, but when we do not see a significant improvement when using the cubic vs the quadratic (p-value 0.673799).

So recall the fitted model from above:

$$\hat{z} = 3.82 + 0.13x + 0.18y - 0.15x^2 + 0.43xy - 0.22y^2$$

We can use our functions to get the stationary points and the principal axes.

```
statpts(lm2Square)

##           x0           y0
## 1 -2.481127 -2.003207

paxis(lm2Square)

##           p10           p11           p20           p21
## 1 0.08433488 0.8413686 -4.952125 -1.18854
```

So we can see that our stationary points are at $(-2.48, -2)$. The first and second principal axis are Equations 13 and 14, respectively:

$$y = 0.08 + 0.84x \quad (13)$$

$$y = -4.95 - 1.19x \quad (14)$$

The surface slice plots in of both principal axes in Figure 4. We can also look at the predicted surface plot using `Plotly`. We can follow the link to the `Plotly` website and look at and manipulate the predicted

surface.

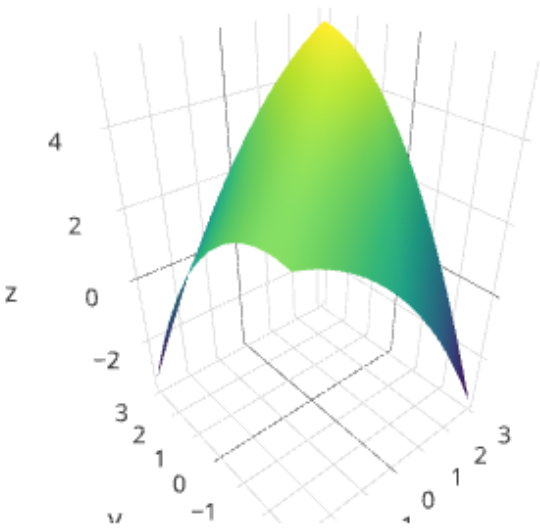


Figure 1: A screenshot of the predicted surface plot from the center perspective.

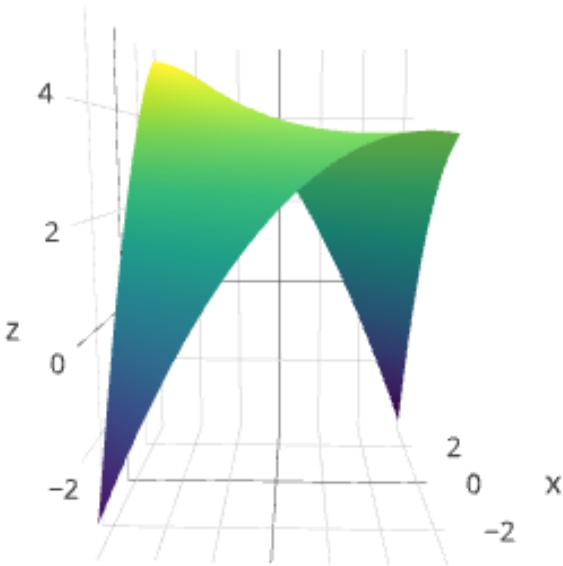


Figure 2: A screenshot of the predicted surface plot from the y axis perspective i.e. eliminating the x axis.

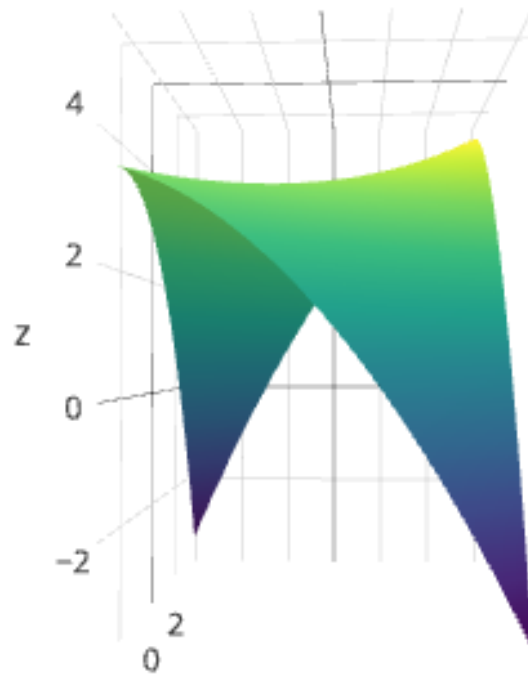


Figure 3: A screenshot of the predicted surface plot from the x axis perspective i.e. eliminating the y axis.

As we can see in the figures above, the first principal axes is the axis that makes the most sense. It's support lies in the plane where we have data. As we can imagine, as the two predictor variables increase the response increases as well. This makes sense because when the physician's rating of patient's health and the physician's estimate of how patient would rate their own health agree i.e. the patient and physician are in agreement about the situation, then the patient is more likely to follow the doctor's orders in the month following the visit. We can also see this in the surface plot. Notice that the lowest points in the surface are when the physician rating and the physician's perception of the patient's own rating are at odds with each other. The surface allows us to see how the two predictor variables interact and affect the response.

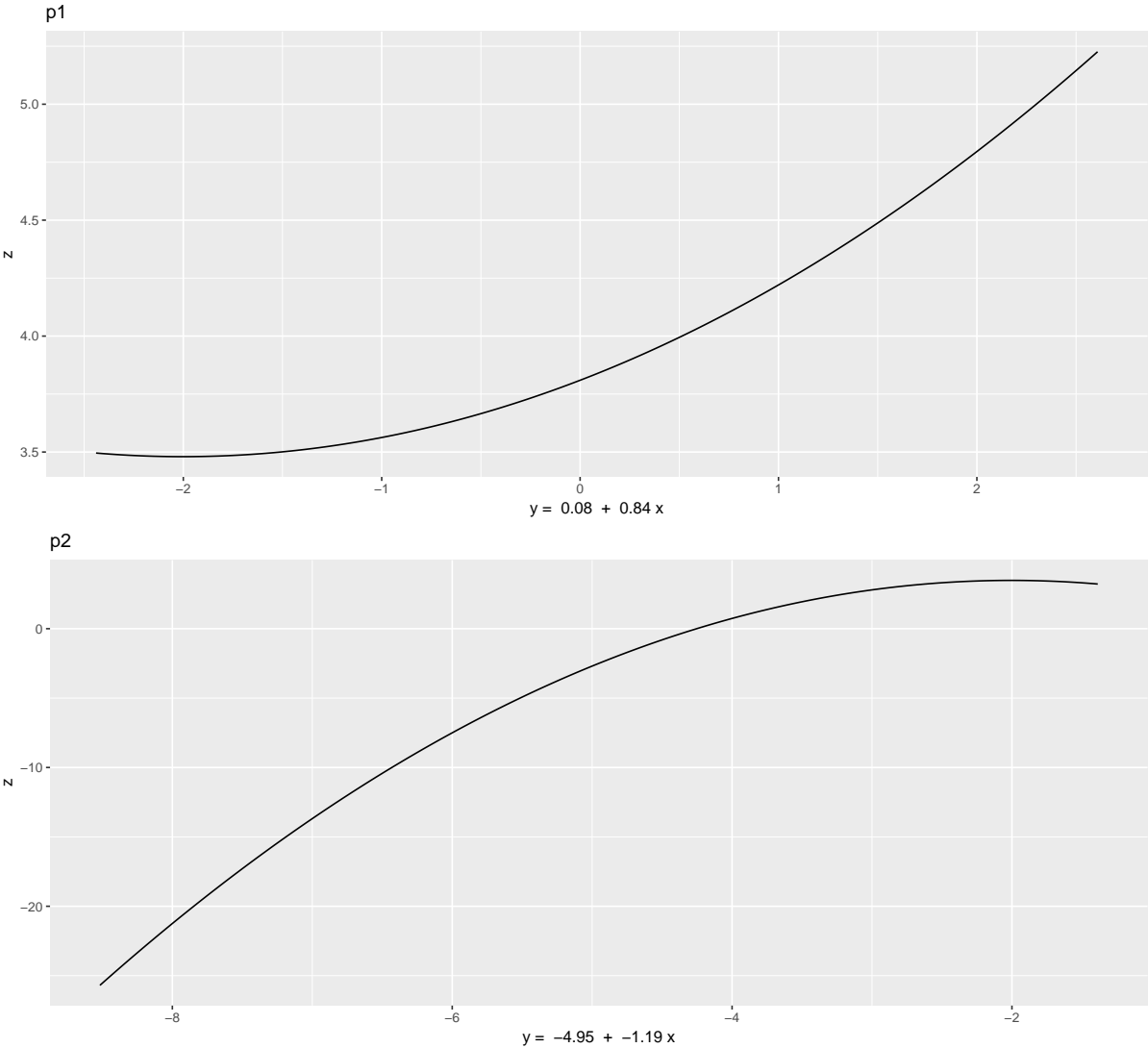


Figure 4: This Figure shows slices of the response surface taken at the two principal axes.

4 Our Example

For our example, we were interested in understanding the congruence between survey reports and observer reports. Our data comes from the Iowa Midlife Transitions Project (MTP); this longitudinal study conducted between 1991 and 2001 on families from eight counties in rural Iowa. The study was designed to look at the effects of the 1960's farm crisis on rural Midwestern families. Families in MTP either participated in the Iowa Youth and Families Project (IFYP) or the Iowa Single Parent Project (ISPP). The IFYP started in 1989 to follow families with at least 2 children and married parents; one of the requirements was the one of the children be in seventh grade in '89 while the other sibling was in a four year range in age of the child. The ISPP started in '91. It focused on recently divorced mothers with a minimum of two children, one of which was a ninth grader in the starting year and the other child was in a 4 year age range of the other child just like before. They had a 78% participation rate of all possible participants in the IFYP; similarly the ISSP had a 99% participation rate. While there were a myriad of economic variables available, we focused on family dynamic variables. Specifically we focused on the husband and wife relationship variables. Our variables of interest are divided into two sets.

Set 1:

Z : Observed Relationship Quility of couple i at Time Wave 0

Y : Wife's Report of Husband's Hostility for couple i at Time Wave 0

X : Husband's Report of Husband's Hostility for couple i at Time Wave 0

Set 2:

Z : Observed Relationship Quility of couple i at Time Wave 0

Y : Wife's Report of Relationship Instability of couple i at Time Wave 0

X : Husband's Report of Relationship Instability of couple i at Time Wave 0

Just like in Phillips et al. (2012), we also centered our variables to ease in interpretation and analysis. We will also check to see if the use of polynomial regression is justified in our case. To do this, we will check the same criterion as we did earlier, namely:

1. Whether the unconstrained model explained a significant amount of the variance in the data.

2. Whether the regression coefficients for the unconstrained model are in expected directions and whether they are significant to the model.
3. Whether if the unconstrained model is significantly better than the constrained model.
4. Whether if a higher order model i.e. polynomial regression is significantly better than the constrained model.

We also checked that our assumptions of normality and constant variance were not violated using the same methods that Dr. Phillips used: high leverage and influential points via Cook's D. We do not find any obvious violations in assumptions for either of the two examples.

We created a table similar to Table 1 in Table 2 to help us collect our information about the confirmatory process which will in turn help us examine the criterion.

Looking at criterion 1, algebraic difference score compared to the unconstrained model, we confirmed that in both outcomes, the unconstrained model explained a significant amount of variation in the data (pval of 0.3614216 and $2.0001838 \times 10^{-12}$, respectively). For criterion 2, the regression coefficients for both models were consistent with each other which is what we would expect. The difference in direction between the two models has to deal with the fact that one would be minimizing and the other is maximizing. Additionally, for both of the unconstrained algebraic models, the regression coefficients are significant. However, the regression coefficients for the unconstrained absolute difference models were not all significant leading us to believe that it would not be a good fit for the data. Criterion 3 and 4 were both met by both examples/models. They both suggest using something greater than a simple difference model although that should not be an unconstrained absolute difference model.

Table 2: This table follows the same format as Table 1, in that it summarizes the confirmatory approach to polynomial regression. Here we see that the where $*p < 0.05$; $**p < 0.01$; $***p < .001$

Outcome	Constrained Model		Unconstrained Model				R^2	F_c	F_h
	$(X - Y)$	R^2	X	Y					
Relationship Quality w/ Hostility	0.0272793	1.5927722×10^{-4}	-0.5368065	-0.6636227			0.2184228	112.5419536	28.4712973
Relationship Quality w/ Instability	-0.0869683	9.1038721×10^{-4}	0.3462168	0.3462168			0.1117448	50.4101937	14.476915
	$ X - Y $	R^2	X	Y	W	WX	WY	R^2	
Patient Adherence	-0.4432221	0.0201095	-0.4223464	-0.7443191	-0.0898305	-0.233025	0.0425533	0.2211262	25.8086281
Physician perception of patient agreement	-0.749572	0.0463969	-0.101314	0.7383982	-0.3960084	0.5853185	-0.2433491	0.116462	7.9498871

After checking our criterion, we find that it is appropriate to use more than difference scores on both examples. If we examine closer, our first model suggests using a linear model between the two predictors and the response (p-val 0.636177). In our second model which examines the relationship between observed relationship quality of the marriage and the perceived relationship instability has moderate evidence of using a quadratic model for the relationship (p-val 0.073613) and has very weak evidence for using a cubic model on the relationship (p-val 0.0945018). Given this information, we will proceed with using

the quadratic polynomial regression model on our second example. While we could reach for using a cubic model, in this case given the complexity of the cubic polynomial regression it is best to stick with quadratic regression which already has enough challenges.

Recall the fitted quadratic polynomial for our example using perceived relationship instability as a predictor for relationship quality is:

$$\hat{z} = -0.52 + 0.25x + 0.23y + 0.02x^2 + 0.09xy + 0.37y^2 \quad (15)$$

Using `Plotly`, we can look at and manipulate the predicted surface by following this [link](#). I have included some screenshots here so you can get a feel for the convexity of the surface which will help in the interpretations of the key components in the surface.

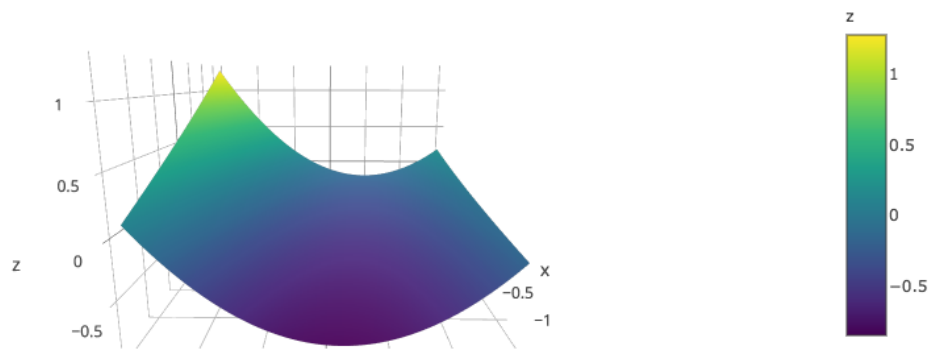


Figure 5: A screenshot of the predicted surface plot from the center perspective.

Just as we did earlier, we will use our functions to get the stationary points and the principal axes.

```
statpts(lm2Square)

##          x0          y0
## 1 -10.18558  0.920306

paxis(lm2Square)

##          p10          p11          p20          p21
```

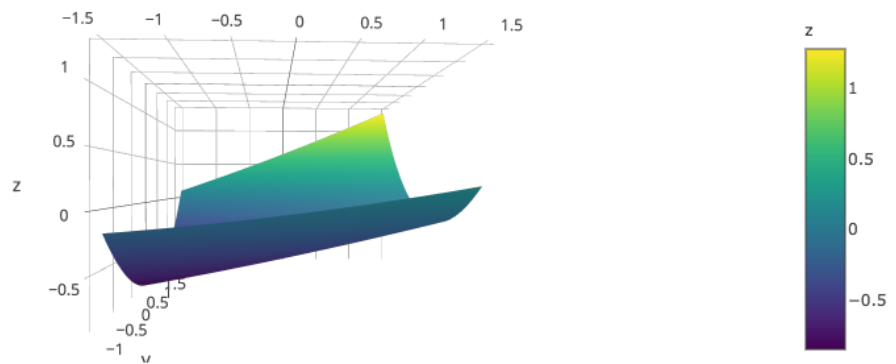


Figure 6: A screenshot of the predicted surface plot from the y axis perspective i.e. eliminating the x axis.

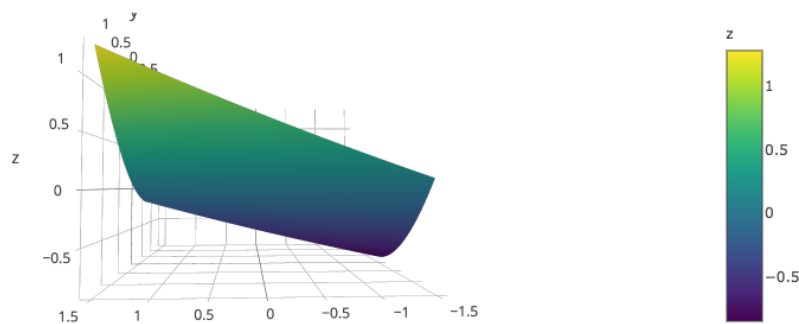


Figure 7: A screenshot of the predicted surface plot from the x axis perspective i.e. eliminating the y axis.

```
## 1 82.92166 8.050729 -0.3448691 -0.1242124
```

The stationary points are located at $(-10.19, 0.92)$; due to our convex nature of the surface, these points are the levels of relationship instability that will minimize the relationship quality in a marriage. The predicted response at the stationary point is -1.705 which is low but not too much. It is interesting that relationship quality seems to be minimized when the wife's perceived relationship instability is really low while the husband's perceived relationship instability is on the higher side.

4.1 Principal Axes (Interpretations come with the plots)

Likewise we can look at the principal axes:

$$y = 82.92 + 8.05x \quad (16)$$

$$y = -0.34 - 0.12x \quad (17)$$

We can look at surface slice plots in of both principal axes in Figure 8.

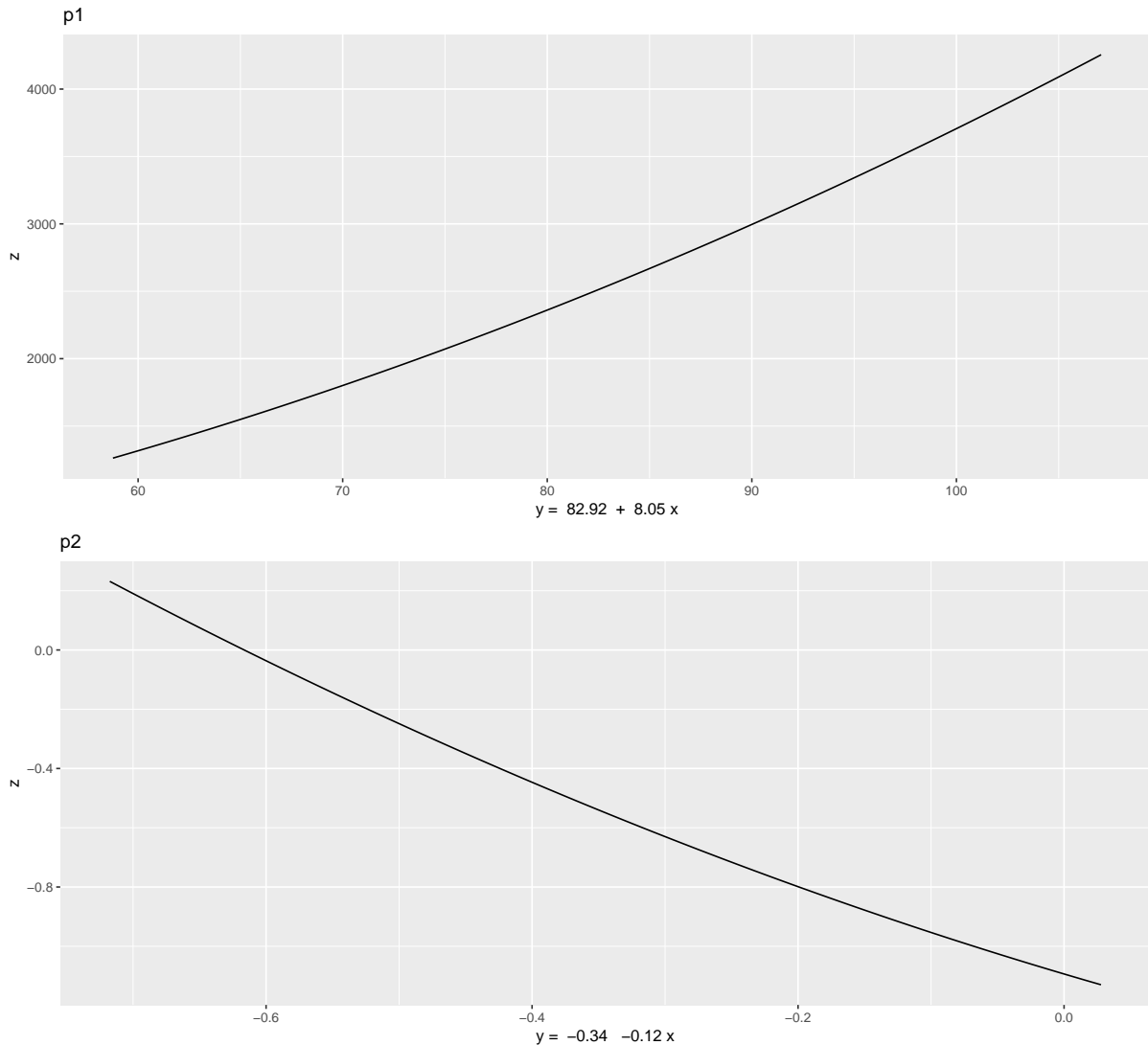


Figure 8: This Figure shows slices of the response surface taken at the two principal axes.

Because we have a convex surface, our interpretations of the principal axis are flipped; the first principal axis, Equation 16, represents the line in the surface where the outcomes decreases the most and the second principal axis, Equation 17, is where the surface has the lowest downward curvature. So the first principal axis shows us where relationship quality can go downhill rather fast. It shows high level of perceived

relationship instability by the wife. And likewise, the second principal axis shows that relationship quality does not decrease that much when the wife's perceived relationship instability hangs around 0 (in the middle of the observed range). There is an issue with the first principal axis, namely that the ranges we would have to make inferences on are out of the bounds of the data. This is an issue because it basically predicting way outside of the scope of our data, so we can't really trust any inference made on the first principal axis similar to how the second principal axis in Phillips(2012) study was out of the bounds of the data. From our surface, we can gather most integral to relationship quality is the wife's perception of instability. When the wife's rating of the relationship instability is held constant and one moves along the husband's rating of relationship instability the observed relationship quality does not change too much. If the opposite is considered i.e. hold husband's rating of relationship instability the observed relationship quality and move along the wife's rating of relationship instability the observed relationship quality there is a great change in predicted observed relationship quality.

There are some issues with our data that affect inferences. As we can see from Figure 9, our predictor variables are very left skewed. The inferences we can make when the relationship has a high perceived instability are fairly well rounded, but we are lacking data in the low perceived relationship instability realm as seen in Figure 10. The method does seem to allow us to easily visualize the realm of possible combinations of the predictor variables and their effect on the predicted response. Ultimately, I feel that the predictor variables need to be fairly well distributed in order to make proper inferences.

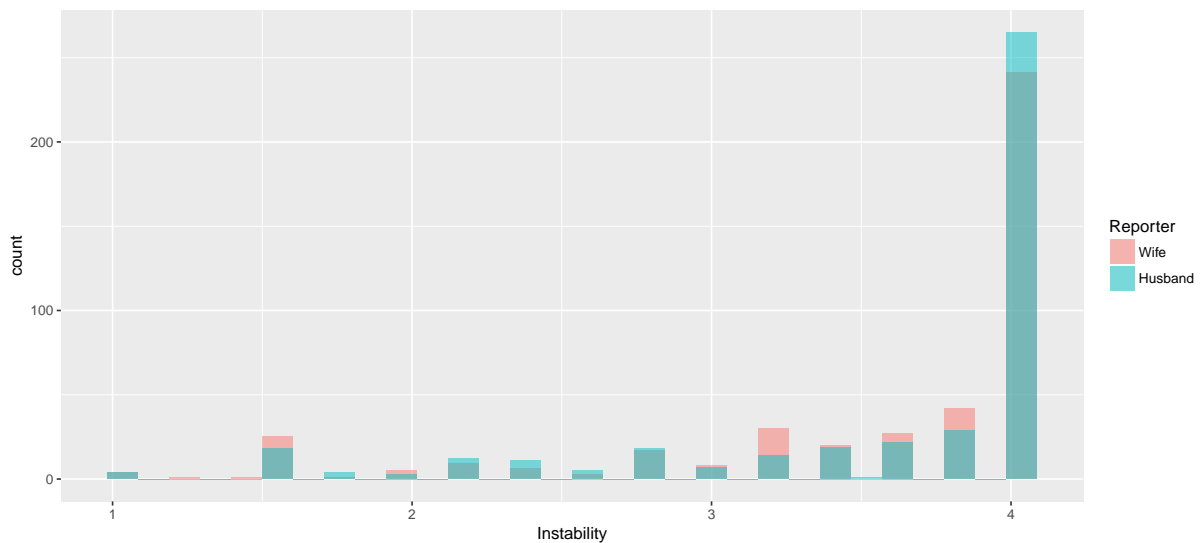


Figure 9: This Figure shows the histograms of our two predictor variables plotted on top of each other to show skewness.

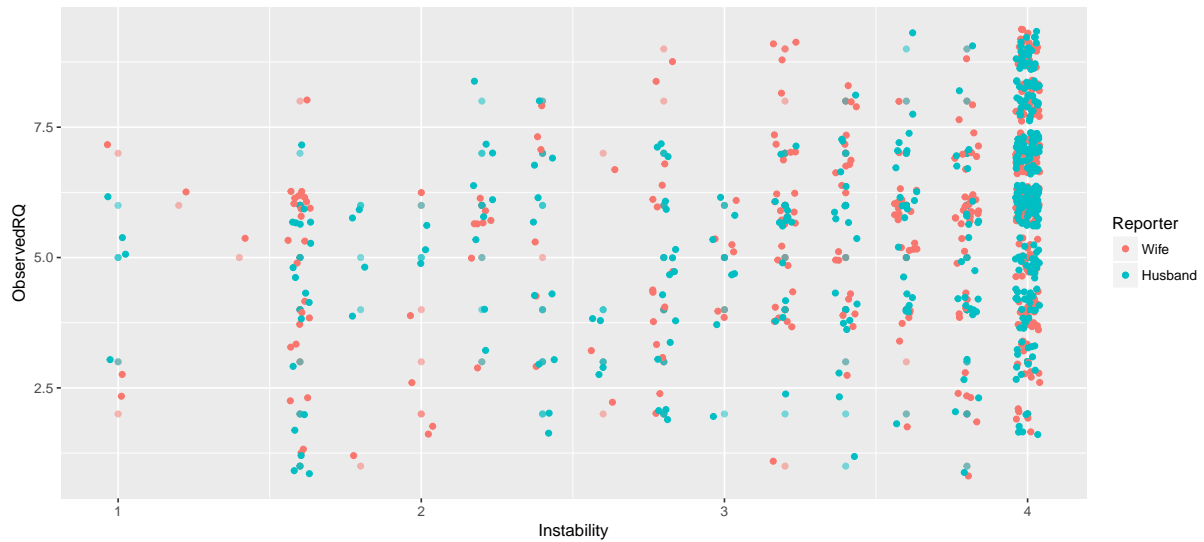


Figure 10: This Figure shows a scatterplot of Instability against Observed Relationship Quality colored by who reported the instability rating.

References

- [1] J. R. EDWARDS, *On the use of polynomial regression equations as an alternative to difference scores in organizational research*, *Academy of Management Journal*, 36 (1993), pp. 1577–1613.
- [2] ———, *Problems with the use of profile similarity indices in the study of congruence in organizational research*, *Personnel Psychology*, 46 (1993), pp. 641–665.
- [3] A. KHURI AND J. CORNELL, *Response Surfaces: Designs and Analyses: Second Edition*, Statistics: A Series of Textbooks and Monographs, Taylor & Francis, 1996.
- [4] L. A. PHILLIPS, *Congruence research in behavioral medicine: methodological review and demonstration of alternative methodology*, *Journal of Behavioral Medicine*, 36 (2012), pp. 61–74.
- [5] L. A. PHILLIPS, M. A. DIEFENBACH, I. M. KRONISH, R. M. NEGRON, AND C. R. HOROWITZ, *The necessity-concerns framework: a multidimensional theory benefits from multidimensional analysis*, *Annals of Behavioral Medicine*, 48 (2014), pp. 7–16.
- [6] R CORE TEAM, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2014.