

## 10.1: ANCOVA with Quantitative Factor Levels

### An Extended Overview of ANCOVA

Designed experiments often contain treatment levels that have been set with increasing numerical values. For example, a chemical process may be hypothesized to vary by two factors: the Reagent type (A or B), and temperature. So the researchers conducted an experiment that investigates a response at 40, 50, 60, 70, and 80 degrees (Fahrenheit) for each of the Reagent types.

You can find the data at [QuantFactorData.csv](#).

If temperature is considered as a categorical factor, we can proceed as usual with a  $2 \times 5$  factorial ANOVA to evaluate the Null Hypotheses:

$$H_0 : \mu_A = \mu_B \quad (10.1.1)$$

$$H_0 : \mu_{40} = \mu_{50} = \mu_{60} = \mu_{70} = \mu_{80} \quad (10.1.2)$$

and

$$H_0 : \text{no interaction} \quad (10.1.3)$$

Although the above hypotheses achieve the goal of comparing response means for the process carried out at different temperatures, no conclusion can be made about the trend of the response as the temperature is increased.

In general, the trend effects of a continuous predictor are modeled using a polynomial where its non-constant terms represent the different trends such as linear, quadratic, and cubic effects. These non-constant terms in the polynomial are called trend terms. The statistical significance of these trend terms can also be tested in an ANCOVA setting by adding columns representing the trend terms and their interaction effects with the categorical factor into the design matrix (X) of the General Linear Model (see Chapter 4 for the definition of a design matrix).

Note that the design matrix representing only the categorical factor contains the column of ones representing the reference factor level and other dummy variable columns representing the remaining factor levels.

Inclusion of the trend term columns will facilitate significance testing for the overall trend effects and the columns representing the interactions can be utilized to compare differences of each trend effect among the categorical factor levels.

Getting back to the chemical process example, if the quantitative property of measured temperature is used, we can carry out an ANCOVA by fitting a polynomial regression model to express the impact of temperature on the response. If a quadratic polynomial is desired, the appropriate ANCOVA design matrix can be obtained by adding two columns representing *temp* and *temp*<sup>2</sup> along with the column of ones representing the reagent type A, the reference reagent category, and one dummy variable column representing the reagent type B.

The *temp* and *temp*<sup>2</sup> terms allow us to investigate the linear and quadratic trends respectively. Furthermore, the inclusion of columns representing the interactions between the reagent type and the two trend terms will facilitate the testing of differences between these two trends between the two reagent types. Note also that additional columns can be added appropriately to fit a polynomial of an even higher order.

#### Rule

To fit a polynomial of degree *n*, the response should be measured at least (*n*+1) distinct levels of the covariate. Preliminary graphics such as scatterplots are useful in deciding the degree of the polynomial to be fitted.

#### Suggestion

To reduce structural multicollinearity, centering the covariate by subtracting the mean is recommended. For more details see STAT 501 - Chapter 12: Multicollinearity

The necessary software code and/or commands along with outputs and conclusions are given below.

In SAS, this process would look like this:

```
/*centering the covariate creating x^2 */
data centered_quant_factor;
set quant_factor;
x = temp-60;
x2 = x**2;
run;
proc mixed data=centered_quant_factor method=type3;
class reagent;
model product=reagent x x2 reagent*x reagent*x2;
title 'Centered';
run;
```

Notice that we specify *reagent* as a class variable, but *x* and *x*<sup>2</sup> enter the model as continuous variables. The regression coefficient of *x* and *x*<sup>2</sup> can be used to test the significance of the linear and quadratic trends for reagent type A, the reference category and the interaction term coefficients can be used if these trends differ by categorical factor level. For example, testing the null hypothesis  $H_0 : \beta_{Reagent*x} = 0$  where  $\beta_{Reagent*x}$  is the regression coefficient of the *Reagent \* x* term is equivalent to testing that the linear effects are the same for reagent type A and B.

SAS output:

Type 3 Analysis of Variance

Source	DF	Sum of Squares	Mean Square	Expected Mean Square	Error Term	Error DF	F Value	Pr > F
reagent	1	3.066357	3.066357	Var(Residual) + Q(reagent)	MS(Residual)	24	2.97	F" class=" ">0.0977
x	1	97.600495	97.600495	Var(Residual) + Q(x,x*reagent)	MS(Residual)	24	94.52	F" class=" "><.0001
x2	1	88.832986	88.832986	Var(Residual) + Q(x2,x2*reagent)	MS(Residual)	24	86.03	F" class=" "><.0001
x*reagent	1	0.341215	0.341215	Var(Residual) + Q(x*reagent)	MS(Residual)	24	0.33	F" class=" ">0.5707
x2*reagent	1	0.067586	0.067586	Var(Residual) + Q(x2*reagent)	MS(Residual)	24	0.07	F" class=" ">0.8003
Residual	24	24.782417	1.032601	Var(Residual)	.	.	.	F" class=" ">.

1. The reagent effect was not significant with  $p = 0.0977$
2. Only the linear and quadratic effects were significant in describing the trend in the response, and linear and quadratic effects were the same for each of the reagent types (no interactions)

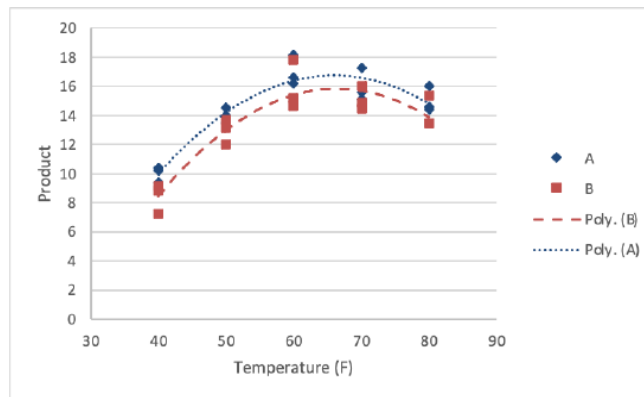


Figure 10.1.1: Graphing product vs temperature

## ? Using R

Steps:

- Load the Quant Factor Data.
- Obtain the ANOVA table after centering the covariate and creating  $x^2$ .
- Plot the data.

### Steps in R

1. Load the Quant Factor data, obtain the ANOVA table (after centering the covariate), and create  $x^2$  by using the following commands:

```
setwd("~/path-to-folder/")
QuantFactor_data <- read.table("QuantFactorData.txt",header=T)
attach(QuantFactor_data)
temp_center<-temp-60
temp_square_center<-temp_center^2
new_data<-cbind(QuantFactor_data,temp_center,temp_square_center)
ancova_model<-lm(product ~ reagent + temp_center + temp_square_center + reagent:
anova(ancova_model)
#Analysis of Variance Table
#Response: product
#
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
#reagent	1	9.239	9.239	8.9476	0.006336	**
#temp_center	1	97.600	97.600	94.5191	8.499e-10	***
#temp_square_center	1	88.833	88.833	86.0284	2.093e-09	***
#reagent:temp_center	1	0.341	0.341	0.3304	0.570749	
#reagent:temp_square_center	1	0.068	0.068	0.0655	0.800257	
#Residuals	24	24.782	1.033			
#---						

```
#Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Only the linear and quadratic effects were significant in describing the trend in the response, and linear and quadratic effects were the same for each of the reagent types (no interactions).

2. Plot the polynomial regression curve for reagent A and reagent B by using the following commands:

```
reagentA_regression <- lm(product ~ temp_center + temp_square_center, data=subset
reagentB_regression <- lm(product ~ temp_center + temp_square_center, data=subset
plot(temp, product, ylim=c(0,20), xlab="Temperature", ylab="Product", pch=23, col=if
lines(fitted(reagentA_regression) ~ temp, data=subset(new_data, reagent=="A"), co
lines(fitted(reagentB_regression) ~ temp, data=subset(new_data, reagent=="B"), co
text(locator(1), "reagent A", col="blue")
text(locator(1), "reagent B", col="red")
detach(QuantFactor_data)
```

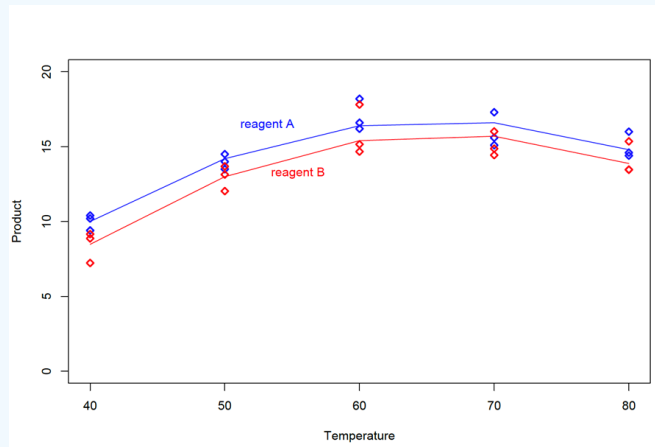


Figure 10.1.2: Graphing product vs temperature using R

This page titled [10.1: ANCOVA with Quantitative Factor Levels](#) is shared under a [CC BY-NC 4.0](#) license and was authored, remixed, and/or curated by [Penn State's Department of Statistics](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.