QUADRATIC MODEL TO ESTIMATE THE DOSES CAUSING THE HIGHEST CHOLESTEROL CONCENTRATION AND THE SAME CHOLESTEROL CONCENTRATION AS CONTROL GROUP

Wuyan Zhang

Steve Nissen

Donald Beitz

Philip Dixon

Recommended Citation

Applied Statistics in Agriculture

QUADRATIC MODEL TO ESTIMATE THE DOSES CAUSING THE HIGHEST CHOLESTEROL CONCENTRATION AND THE SAME CHOLESTEROL CONCENTRATION AS CONTROL GROUP

Wuyan Zhang[a], Steve Nissen[b], Donald Beitz[b] and Philip Dixon[a]

[a]Department of Statistics, Iowa State University, Ames, IA, 50011-1210
[b]Department of Animal Science, Iowa State University, Ames, IA, 50011-3150

ABSTRACT

High plasma cholesterol (particularly high LDL-cholesterol) is a high risk factor for coronary heart disease (CHD), which causes a high CHD morbidity and mortality. Besides clinical drugs, more and more interest is focused on finding natural components in the diet that may have hypocholesterolemic effects. Plant sterols are natural components in human diets and found to have cholesterol-lowering effects in humans. Sheanut oil has a relatively high amount of plant sterols. Therefore, the two experiments were designed to investigate the hypocholesterolemic effect of sheanut oil in hamsters. The response was not monotonic. Low doses increased plasma cholesterol, but high doses decreased plasma cholesterol. Because there was partial dose repetition between the two experiments, the two were combined together to estimate the dose leading to the highest cholesterol concentration and the dose leading to the same cholesterol concentration as the control group. A quadratic model was selected to fit the combined data after appropriate transformation of exploratory and response variable. Nonparametric smoothing method was used to justify the quadratic model. The results of point estimation and confidence interval were compared by Delta, Fieller's and bootstrapping methods.

Keyword: High plasma cholesterol, plant sterols, smoothing, Delta, Fieller's , bootstrapping.

1. Introduction

Coronary heart disease and plasma cholesterol concentration

Coronary heart disease is a condition in which blood flow is reduced in main (coronary) arteries; therefore, not enough blood and oxygen are available to heart muscles, which causes a quick death (Gurr 1992). Coronary heart disease (CHD) is the leading cause of death in industrialized countries (Smith et al. 1998). In 1998, CHD caused more deaths than cancer and accidents together in the United States (Lauber and Sheard 2001).

Many risk factors that contribute to atherosclerosis development disease have been found since the 1940s including cigarette smoking, high levels of LDL-cholesterol, low levels of high-density lipoprotein (HDL)-cholesterol, increased blood pressure, diabetes, obesity and physical inactivity (Krauss 1996). Among all these risk factors, most attention has been given to plasma cholesterol and the smoking habit (Grundy et al. 1999).

Early as the 1930s, Antischkown et al. (1933) observed that animals developed high concentration of plasma lipoprotein when fed high cholesterol diets and they also developed

atherosclerosis. After several decades of research, the lipid hypothesis was established to explain the relationship between plasma cholesterol concentration and CHD mortality (Gurr 1992). The lipid hypothesis suggested that high plasma cholesterol (particularly high LDL-cholesterol) is a high risk factor for CHD, which causes a high CHD morbidity and mortality. On the other hand, a reduction in plasma cholesterol (particularly LDL-cholesterol) can lessen the risk of CHD.

Plant sterols and sheanut oil extract

Many drugs, such as HMG-CoA reductase inhibitors, have been invented to decrease plasma cholesterol concentration in human. Besides clinical drugs, more and more interest is focused on finding natural components in the diet that may have hypocholesterolemic effects. Plant sterols, the nonsaponifiable parts of vegetable oil, are natural components in human diets. Early as the 1950s, plant sterols were found to have cholesterol-lowering effects in humans (Pollak et al. 1981). Sheanut oil, the vegetable oil from shea butter tree (Sapotaceae:*Butyrospermum paradoxum)*, has a relatively high amount of plant sterols (2-11%) (Badifu 1993). Therefore, my studies were designed to investigate the hypocholesterolemic effect of sheanut oil.

Experimental design

In the two experiments, hamsters were used as experiment animals because of their cholesterol metabolic similarity to humans. Hamsters three weeks old were randomly assigned to one of four treatment groups in both the first (60 hamsters) and second experiments (68 hamsters). Each experiment lasted for six weeks. Table 1 shows sample size and sheanut oil dose differences of the four treatment groups in each experiment.

At the end of the studies, individual hamster weight was recorded. The plasma cholesterol concentration was analyzed. Figure 1 shows the relationship between plasma cholesterol concentration and sheanut oil concentration. Plasma cholesterol concentration increased slightly at low doses, then decreased. Therefore, the doses of sheanut oil which can cause the highest cholesterol concentration and the same cholesterol concentration as the control group are important to know. Because there was partial dose repetition between the two experiments, the data were combined to increase the power of statistical inference.

2. Finding the appropriate model to combine the data

Quadratic model.

The plasma cholesterol concentrations first increased in low doses, then decreased at high doses. Therefore, the quadratic model may be reasonable to fit the data.

$$TC= \beta_0 + \beta_1*Experiment+ \beta_2*BW + \beta_3*Shea^2+ \beta_4*Shea + \varepsilon \quad (Equation~1)$$

TC: plasma total cholesterol concentration.
Experiment: 1 if experiment 1
            2 if experiment 2.
Shea: sheanut oil weight percentage in hamsters diet.
$Shea^2$: the square of shea.
BW: hamsters' body weight at the end of experiment.

Applied Statistics in Agriculture

The second order term is not significant (P=0.37) and its coefficient estimation is positive (0.65). So the fitting curve was concave instead of convex. Therefore, the estimation based on this model is efficient. The residual plots show the unequal variance. All this information suggests that the transformations of response and explanatory variables are necessary.

Model transformation

Since the combined data are right skewed, the log or power (<1) transformation of the sheanut oil concentration are used. For the first experiment, the control treatment dose is zero. Therefore, the log transformation of the explanatory variables is not preferred. We tried several power (<1) transformations. When ½ power transformation is used. The data is still right skewed. When using negative power transformation, the data changes into left skewed and over adjusted. After several tries, the ¼ power is appropriate for the transformation.

The error variances are not equal and they increase as the mean response increases. Box and Cox transformations can be applied. According to the previous animal studies on plasma cholesterol concentration, log transformation is used.

$LnTC = \beta_0 + \beta_1 * Experiment + \beta_2 * BW + \beta_3 * Sheasq^2 + \beta_4 * Sheasq + \varepsilon$ (Equation 2)
Ln TC: log of TC.
Sheasq: the 1/4 power of shea.
$Sheasq^2$: the square of sheasq.
TC, Experiment, Shea and BW are defined as before.

The second order term is negative and significant (Table 2). The residual plots also show an improvement. Therefore, this model is appropriate for the estimation of the two doses.

The interaction

In the previous model, the interaction term between the explanatory variables are not included. The big concern about the interaction is that the effect of sheanut oil concentrations might not be independent of hamsters' body weight. So we included this term in the existing model.

$LnTC = \beta_0 + \beta_1 * Experiment + \beta_2 * BW + \beta_3 * Sheasq^2 + \beta_4 * Sheasq + \beta_5 * Sheasq * BW + \beta_6 * Sheasq^2 * BW + \varepsilon$ (Equation 3)
Sheasq*BW: the interaction between Sheasq and BW.
$Sheasq^2 * BW$: the interaction between $Sheasq^2$ and BW.
LnTC, Experiment, Sheasq, Sheasq2 and BW are defined as before.

The interaction between sheanut oil concentration is not significant at α=0.05. Therefore, the model which does not include the interaction is appropriate for the data.

Model justification

Many techniques provide useful tools for checking the appropriateness of the proposed parametric model. Since the combined data have repeat observations at all sheanut oil concentration, F test for lack of fit is used to determine whether our specific type of regression

function adequately fits the data. The P value is bigger than 0.05. Therefore, there is no evidence of lack of fit of the transformed model.

Another interesting method is to apply the smoothing technique to the residual plots. The idea is to check the residuals from the proposed model by smoothing (Bowman and Azzalini 1997):

Model1: $R = \beta_0 + \beta_1 * \text{Experiment} + \beta_2 * \text{BW} + \varepsilon$

Model 2: $R = \beta_0 + \beta_1 * \text{Experiment} + \beta_2 * \text{BW} + \beta_3 * S(\text{sheasq}) + \varepsilon$

R are the residuals from the proposed model ($\text{LnTC} = \beta_0 + \beta_1 * \text{Experiment} + \beta_2 * \text{BW} + \beta_3 * \text{Sheasq}^2 + \beta_4 * \text{Sheasq}$)

S(sheasq) is the smoothing function of parameter sheasq.

Model 1 is a special case of model 2. If model 1 is not significantly different from model 2, the coefficient of smoothing function is approximately zero. Therefore the residual has no obvious pattern of sheasq. The F value for these data is 0.248, which is relative small and not significant. Again, there is no evidence of lack of fit of the proposed model.

The smoothing technique also can be used for the prediction. We can compare the prediction of the quadratic transformed model with the prediction by smoothing. If the proposed model fits the data well, the two prediction should be similar. Figure 2 shows the results of the comparison. The prediction lines from the two model are almost the same. It also shows the transformed model fit the data well.

3 Estimation of doses

The two doses of interest are the sheanut oil dose that can cause the maximum plasma cholesterol concentration (M_m) and the dose that can cause the same cholesterol concentration as control group (M_c). According to transformed quadratic model, these doses can be simply expressed as:

$$m_m^{\frac{1}{4}} = -\frac{\beta_4}{\beta_3}, \quad m_c^{\frac{1}{4}} = -\frac{\beta_4}{2\beta_3} \quad \text{(Equation 4)}$$

After back transformation from quadratic regression model, they can be estimated as:

$$\hat{m}_m = \left(-\frac{b_4}{2b_3}\right)^4 = \left(-\frac{0.168}{2*0.207}\right)^4 = (0.405)^4 = 0.027$$

$$\hat{m}_c = \left(-\frac{b_4}{b_3}\right)^4 = \left(-\frac{0.168}{0.207}\right)^4 = (0.81)^4 = 0.43 \quad \text{(Equation 5)}$$

b_3 and b_4 are the point estimation of regression coefficient β_3 and β_4 in equation 2.

Both of the estimated doses are expressed as a function of the ratio of the regression coefficients. Several approaches can be applied to construct the confidence interval. One theoretical method is delta method (Everitt 2002).

Applied Statistics in Agriculture

Delta method

$$Var((\hat{m}_c)^{1/4}) = Var\left(-\frac{b_4}{b_3}\right) = \left(\frac{\partial}{\partial b_3}\left(-\frac{b_4}{b_3}\right)\right)^2 * Var(b_3) + \left(\frac{\partial}{\partial b_4}\left(-\frac{b_4}{b_3}\right)\right)^2 * Var(b_4)$$

$$+ 2\left(\frac{\partial}{\partial b_3}\left(-\frac{b_4}{b_3}\right)\right)*\left(\frac{\partial}{\partial b_4}\left(-\frac{b_4}{b_3}\right)\right)* Cov(b_3,b_4)$$

$$= \left(\frac{b_4}{b_3^2}\right)^2 * Var(b_3) + \left(-\frac{1}{b_3}\right)^2 * Var(b_4) + 2\left(\frac{b_4}{b_3^2}\right)*\left(-\frac{1}{b_3}\right)* Cov(b_3,b_4) \qquad \text{(Equation 6)}$$

$$= 0.467$$

$$Se\left((\hat{m}_c)^{1/4}\right) = \sqrt{Var((\hat{m}_c)^{1/4})} = 0.68$$

95% confidence interval for $(\hat{m}_c)^{1/4}$:

$$\left((\hat{m}_c)^{1/4} \pm Z_{0.975} * Se(\,(\hat{m}_c)^{1/4}\,)\right) = (\,-0.52, 2.14\,)$$

95% confidence interval for $\hat{m}_c$: $(0, 20.9)$

Similarly, 95% confidence interval for $(\hat{m}_m)^{1/4}$:

$$\left((\hat{m}_m)^{1/4} \pm Z_{0.975} * Se(\,(\hat{m}_m)^{1/4}\,)\right) = (\,-0.26, 1.07\,)$$

95% confidence interval for $\hat{m}_m$: $(0, 1.31)$

Fieller's theorem

   Fieller's theorem gives the theoretical formula of confidence interval for the estimators which are the function of the ratio of regression parameters. In our case, the formula for the confidence interval could be expressed as the following (Piegorsch and Bailer 1997):

$$M_l^{\frac{1}{4}}, M_u^{\frac{1}{4}} = \frac{1}{(1-g)}\left\{(\hat{m})^{1/4} - \frac{gV_{12}}{V_{22}} \pm \frac{ts}{b_3}\left[V_{11} - 2(\hat{m})^{1/4}V_{12} + (\hat{m})^{1/2}V_{22} - g(V_{11} - \frac{V_{12}^2}{V_{22}})\right]^{1/2}\right\}$$

$$g = \frac{t^2 s^2 V_{22}}{b_3^2}, t = t_{123,0.025}, s^2 = MSE, \text{var}(b_3) = s^2 * V_{11}$$

$$\text{var}(b_4) = s^2 * V_{22}, Cov(b_3,b_4) = s^2 * V_{12} \qquad \text{(Equation 7)}$$

95% confidence interval for $(\hat{m}_c)^{1/4}$: $(-0.43, 1.45)$

95% confidence interval for $\hat{m}_c$: $(0, 4.42)$

Similarly, 95% confidence interval for $(\hat{m}_m)^{1/4}$: $(-0.215, 0.775)$

95% confidence interval for $\hat{m}_m$: $(0, 0.28)$

Bootstrapping

Bootstrapping can provide estimates of the precision of the estimates of regression coefficients, fitted value, and predication of new observation (Neter et al. 2000). Bootstrap sampling for regression can be done in two basic ways. One way is to resample the original observations by simple random sampling with replacement. The other way is to sample the residuals from the fitted regression. If the regression function being fitted is appropriate for the data, the residual term have the constant variance and the predictor variables can be regarded as fixed. Then the bootstrap sample residuals are added to the fitted values from the original regression to obtain new bootstrap Y values. The data was simulated 500 times for two different bootstrap methods. The point estimation and 95% confidence intervals were calculated by percentile method and were shown in Table 3 and Table 4. Frequency distributions were skewed as shown in Figure 3.

4. SUMMARY

Delta method, Fieller's theory and bootstrapping are not similar in estimation of M_m and M_c (Table 5). These methods also disagree considerably on the 95% confidence interval of M_m and M_c (Table 6). Among the three methods, Delta method gives the widest confidence interval. Bootstrapping method (Sampling from the residuals) has the shortest confidence interval. The calculation of 95% confidence interval by Delta method and Fieller's theorem depends on the variance-covariance matrix from the proposed model. Delta method is derived from a Taylor expansion. When the sample size is large enough, Delta method can be applied to the continuous function of known parameters. Fieller's theorem is developed especially for a function of the ratio of normal random variables. It is a more specific method. Therefore Fieller's theorem is more accurate and gives the shorter confidence interval. Bootstrapping methods use the original sample as the sampling population and draw the random sample with replacement. The increased number of simple random samples may leads to more accurate estimation. The bootstrap sampling from the residuals makes stronger assumption that residuals are exchangeable. Therefore, this bootstrap gives the shortest confidence interval. In the histogram of Mc by bootstrap, we can observe the distribution of the Mc is not very symmetric. Therefore, the bootstrap confidence interval is more appropriate than the symmetric confidence intervals from Delta.

REFERENCES

1. Anitschkow N (1933): A survey of the problem. In *Arteriosclerosis* 271-322 (Cowdry EV, ed.).
2. Badifu GIO (1993): Food potentials of some unconventional oilseeds grown in Nigeria – A brief review. *Plant Foods Hum Nutr*. 43, 211-224.
3. Bowman W, Azzalini A(1997): Applied Smoothing Techniques for Data Analysis 355-336
4. Everitt B (2002): The Cambridge dictionary of statistics. Second edition. 111 Cambridge university press.

Applied Statistics in Agriculture

5. Grundy SM, Pasternak R, Greenland P, Smith S, Fuster V (1999): Assessment of cardiovascular risk by use of multiple-risk-factor assessment equations. *Circulation* 100, 1481-1492.

6. Gurr M (1992): Dietary lipids and coronary heart disease: old evidence, new perspective. *Prog Lipid Res.* 31, 195-243.

7. Krauss RM, Richard JD, Ernst N, Fisher E, Howard BV, Knopp RH, Kotchen T, Lichtenstein AH, McGill HC, Pearson TA, Prewitt TE, Stone NJ, Van Horn L, Weinberg R (1996): Dietary guidelines for healthy American adults. *Circulation* 94, 1795-1800.

8. Lauber RP, sheard NF (2001): The American Heart Association dietary guidelines for 2000: a summary report. *Nutr Rev.* 59, 298-306.

9. Neter J, Kutner M, Nachtsheim C, Wasserman W(2000): Applied Linear Statistical Models (4th) 429-434 Mcgraw-Hill.

10. Piegorsch W, Bailer J(1997): Statistics for Environmental Biology and Toxicology. 86-89 Chapman & Hall.

11. Pollak OJ, Kritchevsky D (1981): Phytosterols-sitosterol as cholesterol depressants. Sitosterol Monographs on Atherosclerosis. 10, 60-118.

12. Smith JD (1998): Mouse models of atherosclerosis. *Lab Anim Sci.* 48, 573-579.

13. SAS Institute, Inc. 1999. SAS online Doc, Versoin 8. Cary, NC: SAS Institute.

14. S-PLUS 2000 Professional Release 2. 1988-1999 Mathsoft, Inc.

Table 1. Sample sizes and sheanut oil weight percentage in the diet for Experiment 1 and 2

Sheanut oil	Diets(g/g)						
	0%	0.25%	0.5%	1%	2%	4%	8%
Experiment 1	15	15	15	15	-	-	-
Experiment 2	-	-	-	17	17	17	17

Table 2. Estimation of parameters for the transformed model

Parameter	Estimate	Standard error	T value	Pr > |t|
Intercept	5.85	0.188	31.03	<0.0001
Experiment	0.19	0.051	3.65	0.0004
Sheasq	0.17	0.099	1.69	0.094
Sheasq2	-0.21	0.057	-3.59	0.0005
Body weight	-0.0058	0.0015	-3.73	0.0003

Table 3. Bootstrap A results (sampling from the observation)

	Mean	Empirical 95% confidence interval
$\left(\hat{m}_m\right)^{1/4}$	0.75	(-0.30,0.63)
$\left(\hat{m}_c\right)^{1/4}$	0.37	(-0.59,1.25)

Table 4. Bootstrap B results (sampling from the residuals)

	Mean	Empirical 95% confidence interval
$\left(\hat{m}_c\right)^{1/4}$	0.75	(-0.33,1.21)
$\left(\hat{m}_m\right)^{1/4}$	0.37	(-0.17,0.61)

Table 5. Point estimation of M_c and $M_{m:}$

	Delta method	Bootstrapping A	Bootstrapping B
M_m	0.027	0.021	0.021
M_c	0.43	0.32	0.32

Table 6. 95% confidence interval of M_c and $M_{m:}$

	Delta method	Fieller's theory	Bootstrapping A	Bootstrapping B
M_m	(0,1.31)	(0,0.28)	(0,0.16)	(0.0.14)
M_c	(0,20.9)	(0,4.42)	(0,2.44)	(0,2.14)

Figure1. Plasma cholesterol concentration vs. sheanut oil weight percentage in diets.
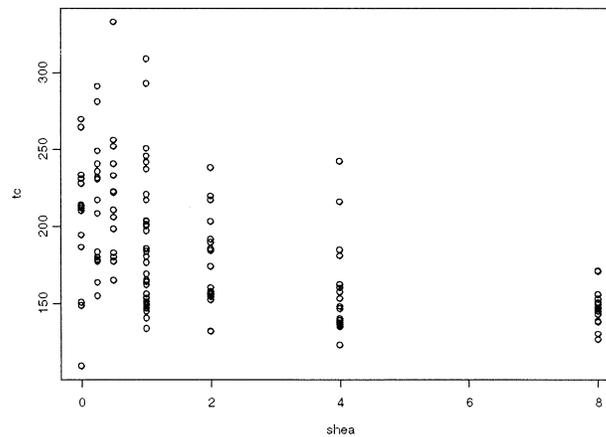
Applied Statistics in Agriculture

Figure 2. The prediction plot from transformed model and smoothing model.
Prediction from the transformed model (solid line):
$$LnTC= \beta_0 + \beta_1*Experiment + \beta_2*Sheasq + \beta_3*Sheasq^2 + \beta_4*BW + \varepsilon$$
Prediction from the smoothing model (dash line):
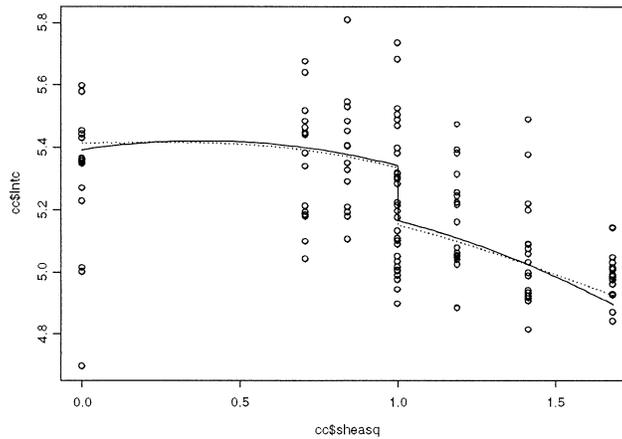$$LnTC= \beta_0 + \beta_1*S(sheasq) + \beta_2* Experiment + \beta_3*BW + \varepsilon$$

Figure 3. The histogram of 500 bootstrap A of $\left(\hat{m}_c\right)^{1/4}$.