

## Case studies: Statistical analysis of contributions of vitamins and phytochemicals to antioxidant activities in plant-based multivitamins through generalized partially double-index model

Jae Keun Yoo<sup>1,a</sup>, Oran Kwon<sup>b</sup>

<sup>a</sup>Department of Statistics, Ewha Womans University, Korea;

<sup>b</sup>Department of Nutritional Science and Food Management, Ewha Womans University, Korea

---

### Abstract

It is important to verify the identity of plant-based multivitamins prepared with a natural-concept and popular for daily consumption because they are easily purchased in markets with imperfect information. For this study, a generalized partially double-index model (GPDIM) was employed as a main statistical method to identify the contribution of vitamins and phytochemicals to antioxidant potentials using data on antioxidant capacities and chemical fingerprinting. A bootstrapping approach via sufficient dimension reduction is adopted to estimate the two unknown coefficient vectors in the GPDIM. Fifth order polynomial regressions are fitted to measure the contributions of vitamins and phytochemicals after estimating the coefficient vectors with the two double indices.

**Keywords:** antioxidant capacity, bootstrapping, generalized partially double-index model, plant-based multivitamins, phytochemicals, sufficient dimension reduction, vitamins

---

### 1. Introduction

Multivitamins are a fast-growing market (Radimer *et al.*, 2004). Natural-concept and/or plant-based products are easily found in grocery stores and lead consumers to anticipate additional health benefits. There are a lack of labeling requirements for natural claims to health benefits that make natural source multivitamins a target of fraud. Therefore, it is important to develop a model that applies a combined approach of a set of *in vitro* bioassay for antioxidant capacity and discriminator chemical fingerprinting to verify the identity of plant-based multivitamins.

‘Phytochemical’ broadly refers to all plant-derived chemicals (Bolling *et al.*, 2011). They are synthesized to protect the plants from adversities including outside attack, physical stress, and oxidation. Phytochemicals are expected to be physiologically beneficial to the human body vis-à-vis numerous mechanisms. A number of studies have revealed that phytochemicals are important for at least their antioxidant role (if not for other properties), to protect tissues from activities that lead to chronic disease (Pietta, 2000). If that is indeed the case with the phytochemicals that coexist in plant-based multivitamins, one must consider a strategy to verify their identity by establishing rigorous criteria that includes chemical standardization and biological assays. So far, the question of whether phytochemical existed in accompany with vitamins in plant-based multivitamins may provide extra

---

<sup>1</sup> Corresponding author: Department of Statistics, Ewha Womans University, 52, Ewhayeodae-gil, Seodaemun-gu, Seoul 03760, Korea. E-mail: peter.yoo@ewha.ac.kr

health benefits remained unsolved until now due to the immense complexity. Moreover, a lack of labeling requirements for natural claims makes fraud possible and diminishes the trustworthiness of plant-based multivitamins. A new strategy to identify extra health benefits from plant-based multivitamins can contribute to protect and maintain the sustainability of the long-term production of natural ingredients. This would assist regulatory bodies and increase consumer confidence towards natural ingredient products.

Computational models have a wide range of applications in food and nutrition science. A mathematical approach was employed to develop a reliable and feasible model to verify the identity of plant-based multivitamins, using antioxidant capacities as a model function. Our investigative focus is on the contributions of vitamins and phytochemicals to antioxidant activities; therefore, it naturally induces a consideration of a regression problem for antioxidant activities ( $Y$ ; responses) given two sets of predictors, one of which is vitamins ( $\mathbf{U}$ ), and the other is phytochemicals ( $\mathbf{W}$ ). For this, one popular statistical model should follow multiple linear regression:

$$Y|\{\mathbf{X} \in \mathbb{R}^p = (\mathbf{U} \in \mathbb{R}^{p_u}, \mathbf{W} \in \mathbb{R}^{p_w})\} = \phi + \boldsymbol{\alpha}^T \mathbf{U} + \boldsymbol{\beta}^T \mathbf{W} + \varepsilon, \quad (1.1)$$

$\phi$  is an unknown intercept,  $p = p_u + p_w$ ,  $\boldsymbol{\alpha} \in \mathbb{R}^{p_u}$  and  $\boldsymbol{\beta} \in \mathbb{R}^{p_w}$  are unknown regression coefficients, and  $\varepsilon$  is a random error assumed with zero mean and variance  $\sigma^2$ , which is independent of  $(\mathbf{U}, \mathbf{W})$ . Unknown coefficients of  $\gamma$ ,  $\boldsymbol{\alpha}$  and  $\boldsymbol{\beta}$  in (1.1) are estimated conventionally through least squares.

This usual approach has several potential problems. First, there are possibilities of a relationship between  $\boldsymbol{\alpha}^T \mathbf{U}$  and  $\boldsymbol{\beta}^T \mathbf{W}$  other than the additive one. If so, the form of  $\boldsymbol{\alpha}^T \mathbf{U} + \boldsymbol{\beta}^T \mathbf{W}$  should mislead the investigation of the contribution of vitamins and phytochemicals to each antioxidant activity. Second, if the additive relation cannot be assumed to hold, then least squares are not a proper way to estimate  $\boldsymbol{\alpha}$  and  $\boldsymbol{\beta}$ . These two potential problems arise because the additive relation between  $\boldsymbol{\alpha}^T \mathbf{U}$  and  $\boldsymbol{\beta}^T \mathbf{W}$  is pre-assumed.

To overcome these deficits, in this paper, we consider the following statistical model, called GPDIM:

$$Y|\{\mathbf{X} = (\mathbf{U} \in \mathbb{R}^{p_u}, \mathbf{W} \in \mathbb{R}^{p_w})\} = f(\boldsymbol{\alpha}^T \mathbf{U}, \boldsymbol{\beta}^T \mathbf{W}) + \varepsilon.$$

We will further GPDIM in the following section. To estimate  $\boldsymbol{\alpha}$  and  $\boldsymbol{\beta}$  without knowing  $f(\cdot)$ , a method recently proposed by Yoo (2015) will be adopted. After estimating  $\boldsymbol{\alpha}$  and  $\boldsymbol{\beta}$  through the method in Yoo (2015), the original regression of  $Y|\{\mathbf{X} = (\mathbf{U} \in \mathbb{R}^{p_u}, \mathbf{W} \in \mathbb{R}^{p_w})\}$  is changed to that of  $Y|(\hat{\boldsymbol{\alpha}}^T \mathbf{U} \in \mathbb{R}^1, \hat{\boldsymbol{\beta}}^T \mathbf{W} \in \mathbb{R}^1)$  without a loss of information on the regression. However, the clear difference between the two is the number of predictors in the model. The former has  $p$  predictors, while the latter has only two predictors. A suitable model using  $Y|(\hat{\boldsymbol{\alpha}}^T \mathbf{U}, \hat{\boldsymbol{\beta}}^T \mathbf{W})$  will be constructed to investigate the contributions of vitamins and phytochemicals to antioxidant activities through polynomial regression.

## 2. Data description and statistical methods

### 2.1. Data description

This section describes how data was collected. A total of 15 multivitamins with 3 lots were purchased from a local market in Seoul, Korea. Single daily dose of each sample was ground into a fine powder and extracted with 10 mL of 1% hydrochloric methanol by sonication for 30 minutes. Extracts were centrifuged at 6,000 rpm for 15 min at 4°C, filtered through a syringe filter (0.2  $\mu\text{m}$ ), aliquoted and stored at  $-70^\circ\text{C}$  in foil covered test tubes prior to analysis.

To acquire *in vitro* antioxidant capacity measurements from commercial multivitamin products, ferric reducing antioxidant power (FRAP) was measured using a blue colored complex,  $\text{Fe}^{2+}$ /TPTZ as described by Benzie and Strain (1996). Total phenol content (TPC) was determined by the Folin-Ciocalteu colorimetric method (Singleton *et al.*, 1999). Aqueous gallic acid was used as a standard and results were expressed as mg of gallic acid equivalent (GAE) per g of daily dose of each multivitamin.

## 2.2. Statistical method

Recall GPDIM defined in:

$$Y|\{\mathbf{X} = (\mathbf{U} \in \mathbb{R}^{p_u}, \mathbf{W} \in \mathbb{R}^{p_w})\} = f(\alpha^T \mathbf{U}, \beta^T \mathbf{W}) + \varepsilon. \quad (2.1)$$

In the GPDIM, the unknown function  $f(\cdot)$  describes the functional relationship between  $\alpha^T \mathbf{U}$  and  $\beta^T \mathbf{W}$ . The key characteristic of the GPDIM is that the predictors  $\mathbf{X}$  are divided into two sets of variables,  $\mathbf{U}$  and  $\mathbf{W}$ . In the model, the response of  $Y$  depends on  $\mathbf{X}$  only through two linear combinations of  $\alpha^T \mathbf{U}$  and  $\beta^T \mathbf{W}$ . The conditional relationship of  $Y$  given  $\mathbf{X}$  is fully characterized by  $\alpha^T \mathbf{U}$  and  $\beta^T \mathbf{W}$ . Therefore, the primary interest in the model is placed onto the estimation of  $\alpha$  and  $\beta$  without knowing  $f(\cdot)$ . A proper model for  $f(\cdot)$  can be built to characterize how  $Y$  depends on  $\alpha^T \mathbf{U}$  and  $\beta^T \mathbf{W}$  after estimating  $\alpha$  and  $\beta$ . For example, if  $f(\alpha^T \mathbf{U}, \beta^T \mathbf{W}) = \phi + \alpha^T \mathbf{U} + \beta^T \mathbf{W}$ , the GPDIM in (2.1) is equivalent to the multiple linear regression model in (1.1). That is, the latter can be expressed as a special case of the former, so the GPDIM generalizes the multiple linear regression.

In the GPDIM, if setting  $\eta \in \mathbb{R}^{p \times 2} = \{\alpha_0 = (\alpha, \mathbf{0})^T, \beta_0 = (\mathbf{0}, \beta)^T\}$ , we can have  $\eta^T \mathbf{X} = (\alpha^T \mathbf{U}, \beta^T \mathbf{W})$ . The least squares may be unsuitable to estimate  $\alpha_0$  and  $\beta_0$  since the specific forms of  $f(\cdot)$  are assumed as unknown. So, we adopt an approach recently developed by Yoo (2015) to estimate  $\alpha_0$  and  $\beta_0$ .

In Yoo (2015), first, the following four element matrices are constructed:

$$\begin{aligned} \mathbf{M}_{\text{OLS}} &= \Sigma^{-1} \text{cov}(Y, \mathbf{X}) \text{cov}(Y, \mathbf{X})^T \Sigma^{-1}; \\ \mathbf{M}_{\text{OLS2}} &= \Sigma^{-1} \text{cov}(Y^2, \mathbf{X}) \text{cov}(Y^2, \mathbf{X})^T \Sigma^{-1}; \\ \mathbf{M}_{\text{SAVE}} &= \Sigma^{-\frac{1}{2}} E(\mathbf{I} - \text{cov}(\mathbf{Z}|\mathbf{Y}))^2; \\ \mathbf{M}_{\text{pHd}} &= \Sigma^{-\frac{1}{2}} E((Y - E(Y))^2 \mathbf{Z} \mathbf{Z}^T) \Sigma^{-\frac{1}{2}}, \end{aligned}$$

where  $\mathbf{I}$  stands for the identity matrix,  $\Sigma = \text{cov}(\mathbf{X})$  and  $\mathbf{Z} = \Sigma^{-1/2}(\mathbf{X} - E(\mathbf{X}))$ .

Next the following weighted sums of the two among the four element matrices:

$$\mathbf{M}_C = \omega \mathbf{M}_{\square} + (1 - \omega) \mathbf{M}_{\circ},$$

where  $0 \leq \omega \leq 1$  and  $\mathbf{M}_{\square}$  and  $\mathbf{M}_{\circ}$  are two different element matrices among the four. If  $\omega$  is equal to 0 or 1,  $\mathbf{M}_C$  is reduced to one element matrix.

Then  $\mathbf{M}_C$  is spectral-decomposed such that  $\mathbf{M}_C = \sum_{i=1}^p \lambda_i \gamma_i \gamma_i^T$  with  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$  and  $\gamma_i^T \gamma_j$  is equal to 1 with  $i = j$  and zero otherwise.

According to Yoo (2015), the eigenvectors of  $\gamma = (\gamma_1, \gamma_2)$  corresponding to the two largest and non-zero eigenvalues of  $\lambda_1$  and  $\lambda_2$  are equal to  $(\alpha_0, \beta_0)$ .

In practice, the sample versions of the four element matrices are constructed by replacing them with their usual moment estimators. For notational conveniences, we will denote the estimator of the four elementary matrices as  $\hat{\mathbf{M}}_{\text{OLS}}$ ,  $\hat{\mathbf{M}}_{\text{OLS2}}$ ,  $\hat{\mathbf{M}}_{\text{SAVE}}$  and  $\hat{\mathbf{M}}_{\text{pHd}}$ , respectively. Then,  $\hat{\mathbf{M}}_C$  is constructed as:  $\hat{\mathbf{M}}_C = \omega \hat{\mathbf{M}}_{\square} + (1 - \omega) \hat{\mathbf{M}}_{\circ}$ . The following six kinds of combinations are considered and summarized in

Table 1: Six candidates of  $\hat{\mathbf{M}}_C$  constructed from combinations from the four sample elementary matrices of  $\hat{\mathbf{M}}_{OLS}$ ,  $\hat{\mathbf{M}}_{OLS2}$ ,  $\hat{\mathbf{M}}_{SAVE}$  and  $\hat{\mathbf{M}}_{pHd}$  with  $\omega = 0.0, 0.25, 0.5, 0.75$

Case 1:	$\hat{\mathbf{M}}_{OLS}$	Case 2:	$\omega\hat{\mathbf{M}}_{OLS} + (1 - \omega)\hat{\mathbf{M}}_{OLS2}$
Case 3:	$\omega\hat{\mathbf{M}}_{OLS} + (1 - \omega)\hat{\mathbf{M}}_{pHd}$	Case 4:	$\omega\hat{\mathbf{M}}_{OLS2} + (1 - \omega)\hat{\mathbf{M}}_{pHd}$
Case 5:	$\omega\hat{\mathbf{M}}_{OLS} + (1 - \omega)\hat{\mathbf{M}}_{SAVE}$	Case 6:	$\omega\hat{\mathbf{M}}_{OLS2} + (1 - \omega)\hat{\mathbf{M}}_{SAVE}$

Table 1. We consider  $\omega = 0.25, 0.5, 0.75$  for case 2 and  $\omega = 0.0, 0.25, 0.5, 0.75$  for cases 3–6. It should be noted that case 1 is the same as the multiple linear regression in (1.1). Therefore, the GPDIM via Yoo (2015) does not rule out multiple linear regression, but includes it as a possible case. For cases 3–6 with  $\omega = 0$ ,  $\hat{\mathbf{M}}_C$  is equal to one of  $\hat{\mathbf{M}}_{SAVE}$  or  $\hat{\mathbf{M}}_{pHd}$ . Once  $\hat{\mathbf{M}}_C$  is constructed, the eigenvectors of  $\hat{\gamma} = (\hat{\gamma}_1, \hat{\gamma}_2)$  corresponding to its two largest eigenvalues of  $\hat{\mathbf{M}}_C$  becomes the estimators  $(\hat{\alpha}_0, \hat{\beta}_0)$  of  $(\alpha_0, \beta_0)$ .

Two natural questions arise in estimating  $\alpha$  and  $\beta$ : (1) how do we choose a best combinations among all possible choices of  $\hat{\mathbf{M}}_C$  and  $\omega$  to induce the best estimation of  $\hat{\gamma}$ ?; (2) how should  $\hat{\gamma}_1$  and  $\hat{\gamma}_2$  be matched with  $\hat{\alpha}_0$  and  $\hat{\beta}_0$  and how should  $\hat{\alpha}$  and  $\hat{\beta}$  be selected from  $\hat{\alpha}_0$  and  $\hat{\beta}_0$ , once they are matched?

To handle the first question, a bootstrapping procedure is suggested by Yoo (2015). First, construct  $\hat{\eta}_{ref}$  from the original sample for each case. Next, obtain  $\hat{\eta}_b$  from a bootstrap sample of  $Y$  and  $\mathbf{X}$ , and iterate this total  $B$  times. Then compute a trace correlation  $r$  in Hooper (1959) to measure how close  $\hat{\eta}_{ref}$  and  $\hat{\eta}_b$ :

$$r_b = \sqrt{\frac{1}{2} \text{trace} \left( \left\{ \hat{\eta}_{ref} (\hat{\eta}_{ref}^T \hat{\eta}_{ref})^{-1} \hat{\eta}_{ref}^T \right\} \left\{ \hat{\eta}_b (\hat{\eta}_b^T \hat{\eta}_b)^{-1} \hat{\eta}_b^T \right\} \right)}, \quad b = 1, \dots, B.$$

To convert the correlation (larger  $r$ , closer to each other) to the distance (smaller value, closer to each other), a trace correlation distance  $D_b^r$  is defined:

$$D_b^r = 1 - r_b.$$

Then, compute  $\bar{D}_b^r = (1/B) \sum_{b=1}^B D_b^r$ . According to Yoo (2015),  $B = 500$  is normally recommended. After computing  $\bar{D}_b^r$  for all possible cases of  $\hat{\mathbf{M}}_C$  and  $\omega$ , finally, choose the case to provide the smallest  $\bar{D}_b^r$ , and set the corresponding  $\hat{\eta}_{ref}$  as the estimate of  $\eta$ .

To answer the second question, the following distinguishing values (DV) is developed in Yoo (2015), and choose one to give the largest DV:

$$\begin{aligned} 1. \text{ DV}_{\hat{\eta}_1, \alpha_0} &= \left| \sum_{\ell=1}^{p_w} \hat{\eta}_{1\ell}^2 - \sum_{\ell=p_w+1}^p \hat{\eta}_{1\ell}^2 \right|; \\ 2. \text{ DV}_{\hat{\eta}_2, \alpha_0} &= \left| \sum_{\ell=1}^{p_w} \hat{\eta}_{2\ell}^2 - \sum_{\ell=p_w+1}^p \hat{\eta}_{2\ell}^2 \right|; \\ 3. \text{ DV}_{\hat{\eta}_1, \beta_0} &= \left| \sum_{\ell=1}^{p-p_u} \hat{\eta}_{1\ell}^2 - \sum_{\ell=p-p_u+1}^p \hat{\eta}_{1\ell}^2 \right|; \\ 4. \text{ DV}_{\hat{\eta}_2, \beta_0} &= \left| \sum_{\ell=1}^{p-p_u} \hat{\eta}_{2\ell}^2 - \sum_{\ell=p-p_u+1}^p \hat{\eta}_{2\ell}^2 \right|. \end{aligned}$$

Here is the rationale of how DV works. Suppose that  $\hat{\eta}_1$  is the correct estimate of  $\alpha_0$ . If  $\alpha_0$  is well estimated by  $\hat{\eta}_1$ , the values of  $\hat{\eta}_{1,\ell}$ ,  $\ell = (p_w + 1), \dots, p$  in  $\hat{\eta}_1$  should be close or equal to zero, so it is expected that  $\sum_{\ell=(p_w+1)}^p \hat{\eta}_{1,\ell}^2$  is close or equal to zero. Then we should have  $DV_{\hat{\eta}_1, \alpha_0} \geq DV_{\hat{\eta}_2, \alpha_0}$ . Suppose that  $\hat{\eta}_2$  is the correct estimate of  $\beta_0$ . By applying the same rationale, it is expected that  $DV_{\hat{\eta}_2, \beta_0} \geq DV_{\hat{\eta}_1, \beta_0}$ . Therefore, the largest one among the four DVs indicates that the first letter should be an estimate of the second letter in the subscripts in DV. So, the natural rule using DV is to choose the largest, and set the first subscript as the estimate of the second subscript in DV. If one of  $\hat{\eta}_1$  or  $\hat{\eta}_2$  is determined as  $\hat{\alpha}_0$ , the other is automatically done as  $\hat{\beta}_0$ . After completing this determination procedure,  $\alpha$  and  $\beta$  are estimated by the first  $p_u$  elements of  $\hat{\alpha}_0$  and the last  $p_w$  elements of  $\hat{\beta}_0$ , respectively. The codes for the statistical method are written in R and available upon request.

### 3. Statistical analysis through GPDIM and results

#### 3.1. Statistical analysis through GPDIM

Recent studies have shown that the high antioxidant activities of plants are largely attributed to phenolic compounds (Zheng and Wang, 2001) and/or carotenoids (Paiva and Russell, 1999). Based on this concept, 6 phenolics of ellagic acid (Ell), genistein (Genis), hesperidin (Hes), keracyanin (Kera), quercetin (Quer), and rosmarinic acid (Ros) and 4 carotenoids of  $\beta$ -carotene ( $\beta$ C), lutein (Lue), lycopene (Lyc), and zeaxanthin (Zeaxan) were selected and analyzed using high-performance liquid chromatography with a photodiode array detector coupled with an Agilent 6530 quadrupole time-of-flight mass spectrometer (Agilent Technologies, Santa Clara, CA, USA). Vitamins C and E and Selenium (VitC, VitE, and Sel, in order) were also analyzed along with the ten discriminator phytochemicals.

In the study, one statistical issue is placed onto investigating the contributions of vitamins and phytochemicals to FRAP and TPC, respectively, which has to be done through statistical modeling.

If considering FRAP and TPC as response variables and  $\mathbf{U} \in \mathbb{R}^3 = (\text{VitC}, \text{VitE}, \text{and Sel})$  and  $\mathbf{W} \in \mathbb{R}^{10} = (\text{Ell}, \text{Genis}, \text{Hes}, \text{Kera}, \text{Quer}, \text{Ros}, \beta\text{C}, \text{Lue}, \text{and Lyc}, \text{Zeaxan})$  as two sets of predictors, the GPDIM should be a natural choice for this study. First, the two unknown coefficient vectors of  $\alpha$  and  $\beta$  corresponding to  $\mathbf{U}$  and  $\mathbf{W}$ , respectively, were estimated by Yoo (2015). Next, proper regression models were investigated through polynomial regression of FRAP or TPC given  $\hat{\alpha}^T \mathbf{U}$  and  $\hat{\beta}^T \mathbf{W}$ . After building the models, the contributions of vitamins and phytochemicals was measured.

Let  $(\hat{\alpha}_F, \hat{\beta}_F)$  and  $(\hat{\alpha}_T, \hat{\beta}_T)$  denote the estimates of coefficients for FRAP and TPC, respectively. Then marginally standardizing each of the predictors to have a sample standard deviation of 1, the estimates of the double indices for FRAP and TPC are:

$$\begin{aligned}\hat{U}_F &:= \hat{\alpha}_F^T \mathbf{U} = 308.58\text{VitC} - 279.89\text{VitE} - 22.59\text{Sel}; \\ \hat{W}_F &:= \hat{\beta}_F^T \mathbf{W} = -1.86\text{Ell} + 10.63\text{Genis} - 5.078\text{Hes} + 14.16\text{Kera} - 27.13\text{Quer} \\ &\quad - 10.53\text{Ros} - 12.17\beta\text{C} + 163.84\text{Lue} - 246.38\text{Lyc} + 9.18\text{Zeaxin}. \\ \hat{U}_T &:= \hat{\alpha}_T^T \mathbf{U} = 378.57\text{VitC} + 1288.79\text{VitE} + 22.58\text{Sel}; \\ \hat{W}_T &:= \hat{\beta}_T^T \mathbf{W} = -0.92\text{Ell} - 1.64\text{Genis} - 1.17\text{Hes} + 80.65\text{Kera} - 7.54\text{Quer} \\ &\quad - 98.88\text{Ros} - 1.02\beta\text{C} + 55.60\text{Lue} - 53.98\text{Lyc} - 8.16\text{Zeaxin}.\end{aligned}$$

With these indices, polynomial regressions upto the fifth order were fitted:

$$\widehat{\text{FRAP}} = 7.19 + 2.61\hat{U}_F - 7.93\hat{U}_F^2 - 1.01\hat{U}_F^3 + 1.15\hat{U}_F^4 - 4.32\hat{W}_F. \quad (3.1)$$

Table 2: Measures of contributions of vitamins (**U**) and phytochemicals(**W**)

Contributions	TSS	Normalization
$C(\text{FRAP} \mathbf{U}) = 247.376$	578.152	0.136
$C(\text{FRAP} \mathbf{W}) = 78.344$	578.152	0.428
$C(\text{TPC} \mathbf{U}) = 326.588$	1737.424	0.188
$C(\text{TPC} \mathbf{W}) = 520.582$	1737.424	0.300

TSS = total sums of squares, FRAP = ferric reducing antioxidant capacity, TPC = total phenol content.

$$\begin{aligned} \widehat{\text{TPC}} = & 6.52 + 40.83\hat{U}_T + 137.9\hat{U}_T^2 + 25.83\hat{U}_T^3 - 139.96\hat{U}_T^4 + 37.49\hat{U}_T^5 \\ & - 2.86\hat{W}_T + 2.55\hat{W}_T^2 + 0.45\hat{W}_T^3. \end{aligned} \quad (3.2)$$

In (3.1), the polynomials of  $\hat{U}_F$  upto 4 are required for the modeling, while  $\hat{W}_F$  is just the first. However, the polynomials of  $\hat{U}_T$  and  $\hat{W}_T$  upto 5 and 3, respectively, are necessary in (3.2). A direct comparison of the contributions of vitamins and polychemicals is not plausible since the numbers of variables in **U** and **W** are different and the maximum polynomial orders of  $\hat{U}$  and  $\hat{W}$  in (3.1) and (3.2) are not the same.

The contributions of vitamins (**U**) and phytochemicals (**W**) were measured by the differences in residual sums of squares between the estimated regression functions in (3.1) and (3.2) and those without either  $\hat{U}$  or  $\hat{W}$ , still keeping the polynomials in (3.1) and (3.2). For example, consider the contribution of **U** to TPC, denoting  $C(\text{TPC}|\mathbf{U})$ . Then, refit (3.2) without  $\hat{U}_T$ , that is fit a regression of  $\text{TPC}|\hat{W}_T, \hat{W}_T^2, \hat{W}_T^3$ . Let the residual sums of squares from the fit in (3.2) and  $\text{TPC}|\hat{W}_T, \hat{W}_T^2, \hat{W}_T^3$  denoted as  $\text{RSS}(\text{TPC}|\hat{U}_T, \hat{W}_T)$  and  $\text{RSS}(\text{TPC}|\hat{W}_T)$ , respectively. Then  $C(\text{TPC}|\mathbf{U})$  is measured as:

$$C(\text{TPC}|\mathbf{U}) = \text{RSS}(\text{TPC}|\hat{W}_T) - \text{RSS}(\text{TPC}|\hat{U}_T, \hat{W}_T).$$

Next, in order to measure the contribution of polychemicals  $C(\text{TPC}|\mathbf{W})$  in (3.2), a regression of  $\text{TPC}|\hat{U}_T, \hat{U}_T^2, \hat{U}_T^3, \hat{U}_T^4, \hat{U}_T^5$  is fitted, and its residual sums of square  $\text{RSS}(\text{TPC}|\hat{U}_T)$  is computed. The contribution is measured:

$$C(\text{TPC}|\mathbf{W}) = \text{RSS}(\text{TPC}|\hat{U}_T) - \text{RSS}(\text{TPC}|\hat{U}_T, \hat{W}_T).$$

Then, bigger values of  $C(\text{TPC}|\mathbf{U})$  and  $C(\text{TPC}|\mathbf{W})$  indicate more contribution to TPC. Actually, the difference of the two residual sums of squares is known for Type-III sums of squares in the context of experimental design (Oehlert, 2000), and the value of  $C(\text{TPC}|\mathbf{U})$  measures the contribution of **U** to the regression of TPC on **U** and **W** in (3.2) after adjusting **W**. In this way, the contributions of **U** and **W** to FRAP can be measured.

For the fair comparison between  $C(\text{FRAP}|\mathbf{U})$  and  $C(\text{TPC}|\mathbf{U})$ , they are normalized by dividing them by their total sums of squares from (3.1) and (3.2), respectively, because FRAP and TPC have different measurement scales. The computed contributions of **U** and **W** to FRAP and TPC are reported in Table 2.

### 3.2. Results

Table 2 shows that the phytochemicals provide three times and one and half more contributions to FRAP and TPC, respectively, than vitamins. The amount of the contribution of phytochemicals to FRAP and TPC are not the same. Phytochemicals account for about 43% and 30% of the total variation of FRAP and TPC, respectively, so phytochemicals provide more contribution to FRAP to TPC.

However, vitamins provide more contribution to TPC than to FRAP. This is an interesting finding because the changes in the amount of contributions of vitamins and phytochemicals to two antioxidant activities are opposite. The changes (0.128) between the contributions of phytochemicals to FRAP and to TPC are also larger than those (0.052) of vitamins. This implies that the changes between the contributions of phytochemicals are more sensitive to those of vitamins.

The necessity and value of this research is confirmed by statistical analysis that indicates that phytochemicals are potentially more important to antioxidant activities than vitamins.

#### 4. Conclusions

There is no labeling requirement for natural claim; therefore, determining contribution of nutrients and phytochemicals to antioxidant potentials is important to discriminate plant-based multivitamins from synthetic multivitamins, because multivitamins (which are the natural-concept and/or plant-based products) are popular in daily consumption and easily purchased in markets.

For this study, GPDIM of antioxidant activities on vitamins and various phytochemicals is employed as a main statistical tool to find potential advantages over multiple linear regression, which is a very popular statistical method. Polynomial regressions are fitted after estimating the unknown coefficient vectors in GPDIM based on the approach by Yoo (2015); in addition, the contributions of vitamins and phytochemicals are measured using residual sums of squares. The importance of phytochemicals should be noted because phytochemicals provide more contribution to antioxidant activities than vitamins.

Following the statistical analysis, an interesting further research topic would be if the amount of vitamins and phytochemicals in antioxidant capacities are measured in different ways (Huang *et al.*, 2002; Walker and Everette, 2009) has opposite and reciprocal relationship. It should also be investigated if the changes in the amount of contribution of phytochemicals to different measures of antioxidant capacities are more sensitive than those in vitamins and that the changes in vitamins seems constant.

We believe that this statistical analysis is beneficiary to prove the characteristics of plant-based multivitamins and efficacy in relevant fields of science.

#### Acknowledgements

For Jae Keun Yoo, this work was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF), funded by the Korean Ministry of Education (NRF-2014R1A2A1A11049389/2009-0093827). For Oran Kwon, this work was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by Ministry of Science, ICT and Future Planning (NRF 2013M3A9C4078153).

#### References

- Benzie IF and Strain JJ (1996). The ferric reducing ability of plasma (FRAP) as a measure of “antioxidant power”: the FRAP assay, *Analytical Biochemistry*, **239**, 70–76.
- Bolling BW, Chen CY, McKay DL, and Blumberg JB (2011). Tree nut phytochemicals: composition, antioxidant capacity, bioactivity, impact factors: a systematic review of almonds, Brazils, cashews, hazelnuts, macadamias, pecans, pine nuts, pistachios and walnuts, *Nutrition Research Reviews*, **24**, 244–275.

- Hooper J (1959). Simultaneous equations and canonical correlation theory, *Econometrica*, **27**, 245–256.
- Huang D, Ou B, Hampsch-Woodill M, Flanagan JA, and Prior RL (2002). High-throughput assay of oxygen radical absorbance capacity (ORAC) using a multichannel liquid handling system coupled with a microplate fluorescence reader in 96-well format, *Journal of Agricultural and Food Chemistry*, **50**, 4437–4444.
- Oehlert GW (2000). *A First Course in Design and Analysis of Experiments*, W. H. Freeman, New York.
- Paiva SA and Russell RM (1999).  $\beta$ -Carotene and other carotenoids as antioxidants, *Journal of the American College of Nutrition*, **18**, 426–433.
- Pietta PG (2000). Flavonoids as antioxidants, *Journal of Natural Products*, **63**, 1035–1042.
- Radimer K, Bindewald B, Hughes J, Ervin B, Swanson C, and Picciano MF (2004). Dietary supplement use by US adults: data from the National Health and Nutrition Examination Survey, 1999–2000, *American Journal of Epidemiology*, **160**, 339–349.
- Singleton VL, Orthofer R, and Lamuela-Raventos RM (1999). Analysis of total phenols and other oxidation substrates and antioxidants by means of Folin-Ciocalteu reagent, *Methods in Enzymology*, **299**, 152–178.
- Yoo JK (2015). Generalized partially double-index model: bootstrapping and distinguishing values, *Communications for Statistical Applications and Methods*, **22**, 305–312.
- Walker RB and Everette JD (2009). Comparative reaction rates of various antioxidants with ABTS radical cation, *Journal of Agricultural and Food Chemistry*, **57**, 1156–1161.
- Zheng W and Wang SY (2001). Antioxidant activity and phenolic compounds in selected herbs, *Journal of Agricultural and Food Chemistry*, **49**, 5165–5170.

Received April 28, 2016; Revised May 16, 2016; Accepted May 18, 2016