Resarch Laboratory for Vegetable and Fruit Industry,
State Research Station, Blangstedgaard,
5000 Odense, Denmark

Fruit content determination in one-fruit jams and preserves

(Frugtindholdsbestemmelse i een-frugt marmelader og syltetøj)

Poul Emil Christensen

Summary

The failure of commonly used methods of fruit content determination in jams and preserves is ascribed to nonnormal frequency distribution of index-constituents in reference material.

An equation is set up which makes it possible to discriminate between index-constituent combinations which in all respects are equally well suited to the purpose.

Applying the method to two commercially produced strawberry jams indicates the possibility of determining the fruit content with reasonable accuracy and small security limits.

Introduction

Many countries have set up regulations governing the quality and composition of foods. Having established such regulations, it is also necessary to lay down procedures by which they can be policed, usually by one of two methods, by inspection at the point of manufacture, or by taking samples of the final product at the point of sale and subjecting them to examination and analysis.

This investigation only deals with the latter method, and is restricted to a quantitative analysis of the amount of fruit used for that particular kind of product which is supposed to have been manufactured from only one kind of fruit.

Before World War 1914 *Beythien* (3), *Baier* and *Hase* (2), *Beythien* and *Simmich* (4), *Härtel* and *Sölling* (8) and Ludwig (13) published their methods of fruit content determination. These methods were founded on the assumption that fixed relationsships exist among certain fruit compounds (index-constituents).

Since then, beginning with (13) many publications have appeared giving analytical data for fruit used in the preserving industry. An almost complete bibliography covering this matter up to 1969 has recently been published by *Goodall* (7).

Macara (14, 15), besides giving analytical data for the fruits in question, also treated his data statistically, being the first investigator to state not only maxima and minima as well as mean values, but also the frequency at which a certain value was to bo expected.

Since then, statistical treatment of analytical data has been further extended, and is now common practise in nearly all publications.

While, before 1940 all investigators tried to find constant relations between several constituents in the fruit, *Steiner* (19,20) as one of the first discussed the possibility of using mathematical-statistical analysis in order to further increase the analytical statement. Steiner showed that by using an optimal linear-combinations of several analytically determined index-constituents, it was possible to arrive at a more precise statement about the fruit content. This method is now generally used to solve analytical problems concerning food composition. Thus, Rolle and *Vandercook* (18) and *Vandercook* et al (21, 22, 23) have characterized lemon juice by use of multiple regression analysis. Nehring and *Klinger* (16) and *Klinger* and *Nehring* (12), on a purely theoretical basis treated the problem of fruit content determination using modern statistical-analytical methods. Recently *Kefford* (9) and *Acker* (1) also discussed the whole matter.

Index-constituents

As stated by *Kefford* (9), the ideal constituent would be one that was specific to a particular fruit and had the same concentration in that fruit at all times and places, was stable to processing, was amended to convenient and accurate determination, and so rare or expensive that it was unlikely to be added as an adulterant.

The ideal index-constituent does not exist. Fruits are in fact very variable in composition and the concentration of all their known constituents is influenced by many factors, variety, maturity, growing area, climate, fertilization and irrigation, as shown by *Kenworthy* and *Harris* (11), *Rahman* et al. (17) *Choureitah* and *Lenz* (6) and several others.

Thus, the composition of a fruit can not be defined in any absolute way, but only in terms of ranges and averages with the usual statistical specification to indicate the variability.

Major fruit constituents, such as sugars and acids are poor index compounds, because sugars and acids are commonly and legitimately added ingredients of jams and preserves. Many minor constituents have been investigated as index-constituents: inorganic elements, polyphenolic compounds, pigments, vitamins, amino-acids etc.

Index-constituents most commonly used have been inorganic elements, mainly because these elements are stable to processing and can easily be determined with great accuracy. Among these especially ash, potassium and phosphate have been widely used. It has been common to combine two or three of the index-constituent in a linear equation to give the fruit content (regression equation).

In spite af the fact that these regression-equations normally fit very well to the results of referense-material analysis, it is also recognized that these equations very often give poor results when applied to commercially manufactured fruit products.

In remedy of these shortcomings this investigations was started in 1969.

Theory

A relevant formula for a jam or preserve could be as follows (Acc. to *Nehring* and *Klinger* (16)).

Formula

content of index-constituent x_i

Indgredient	Kilogram	x_0	x_1	x_2	–	x_r
Fruit.........	S^0	x_0^0	x_1^0	x_2^0	»	x_r^0
Sucrose	S^1	x_0^1	x_1^1	x_2^1	»	x_r^1
Other sugar...	S^2	x_0^2	x_1^2	x_2^2	»	x_r^2
Pectin........	S^3	x_0^3	x_1^3	x_2^3	»	x_r^3
Acid.........	S^4	x_0^4	x_1^4	x_2^4	»	x_r^4
Preservative...	S^5	x_0^5	x_1^5	x_2^5	»	x_r^5
» ...	»	»	»	»	»	»
Ingredient m..	S^m	x_0^m	x_1^m	x_2^m	»	x_r^m

We standardize this formula so that

$$\sum_{i=0}^{m} S^i = 100$$

then we have if x_0 is soluble solid as measured by refractometry

$$\bar{x}_0 = \frac{1}{100}(S^0x_0^0 + S^1x_0^1 + ----- + S^mx_0^m) \qquad 1)$$

the mean soluble solid in the formula mix and likewise

$$\bar{x}_i = \frac{1}{100}(S^0x_i^0 + S^1x_i^1 + ----- + S^mx_i^m) \qquad 2)$$

the content of the i'the index-constituent in the mix.

The mean soluble solid $\bar{x}_0$ with given recipe is almost constant because the soluble solid mainly consists of sugar which has been added. Variations in the soluble solid in the fruit are normally left aside when setting up the formula.

In the case that the concentration of all index-constituents in all formula ingredients (except the fruit) is small and has small standard deviations, it is justified to make further simplifications.

In that case, we can in the equation for the content of index-constituent Y_i in the jam as a function of the recipe content x_i

$$Y_i = \frac{Y_0}{x_0} \bar{x}_i \qquad 3)$$

replace the terms x_i^1 - - - - x_i^m with the means m_i and we get

$$Y_i = Y_0 \left[\frac{S^0 x_i^0 + S^1 m_i^1 + --- + S^m x_i^m}{S^0 m + S^1 m_0^1 + --- + S^m m_0^m} + f_i\right] \qquad 4)$$

In the error-term f_i are included all deviations caused by the above mentioned approximations. Setting

$$k_i = \frac{S^1 m_i^1 + S^2 m_i^2 + --- + S^m m_i^m}{S^0 m_0^0 + S^1 m_0^1 + \quad + S^m m_0^m} \qquad 5)$$

and substitute in 4) we get

$$Y_i = Y_0 \left[\frac{S^0 x_i^0}{S^0 m_0^0 + S^1 m_0^1 + --- S^m m_0^m}\right] + Y_0 k_i + Y_0 f_1 \qquad 6)$$

choosing the proper index-constituents we can interpret $Y_0 k_i$ as a correction term and $Y_0 f_i$ as an approximation term which as such can be neglected.

With the notations used above, the fruit content t in the jam can be equated as follow.

$$t = 100\, Y_0 \times \frac{S^0}{S^0 x_0^0 + S^1 x_0^1 + --- + S^m x_0^m} \qquad 7)$$

and then using 6) and 7) we finally get

$$Y_i - Y_0 k_i = \frac{t}{100} x_i. \qquad 8)$$

Now we will proceed in two different ways:

1. Method

Multiplying 8) with *a* factor b_i we get

$$b_i (Y_i - Y_0 k_i) = \frac{t}{100} b_i x_i \qquad 9)$$

and then summing for all i finally

$$\Sigma b_i (Y_i - Y_0 k_i) = \frac{t}{100} \Sigma b_i x_i \qquad 10)$$

choosing the factors b_i in such a way that

$$\Sigma b_i x_i = 100$$

which is possible always, we get

$$t = \Sigma b_i (Y_i - Y_0 k_i) \qquad 11)$$

In equation 10) the term $\Sigma b_i x_i$ is a linear combination of all index-constituents and can be formally regarded as a regression-equation. Hence, choosing the b_i-coefficients by the method of least-squares, provided that the x_i are multivariate normally distributed, the above equation will give us the fruit content with the least variance. The coefficients b_i will be denoted as regressions-coefficients.

The sum of squared residuals from fitting the model by the least squared methods is

$$s^2_y = \sum^{f(x_i..)} b_i b_j \; s^2 x_i, x_j \qquad 12)$$

where $s^2 x_i, x_j$ are the elements of the dispersion-matrix and $f(x_i ..)$ denote that the summing must be done for all possible permutations of the x_i.

If n samples of jam are analysed, and *if* the index-constituent content in sample f is denoted by Y_{if}, the corrected content $Y_{if} - Y_{0f} k_i$ with Z_{if}, then we have the following equation for the fruit content t_f

$$t_f = \Sigma b_i \times Z_{if} \qquad 13)$$

the arithmetic mean

$$t_n = \frac{1}{n} \Sigma t_f \qquad 14)$$

is then a result where all n jam samples are taken into consideration.

The fruit content in the jam samples with its statistical security limits can finally be stated as follows.

$$\bar{t} = \frac{\sum^{n} b_i Z_i}{n} \pm t\alpha/2;0{,}05 \qquad 15)$$

$$\sqrt{\left(s_{\bar{t}}^2 + \frac{\bar{t}^2 \sum b_i b_j s^2 x_i, x_j}{100^2} \right) \times \frac{1}{n}}$$

2. Method

Taken logarithm on both sides of 8) we get

$$ln\,(Y_i - Y_0 k_i) = ln + ln\,x_i - 4{,}6052 \qquad 16)$$

If, as above, n samples of jam are analysed and the index-constituent content in sample f is denoted by Y_{if}, then

$$1n(Y_{if} - Y_{0f} k_i) = V_{if} \qquad 17)$$

and *if* we further assume that the index-constituent Y_i in the jam are normally distributed we can regard 17) as n determinations of the normally distributed variable V_{if}, we can compute $1n\ t_f$ by the equation

$$1n\ t_{if} = V_{if} + 4{,}6052 - 1n\ x_i \qquad 18)$$

We can calculate a probable value $1n\ t_f$ where we are taking into account the index-constituent content of all compounds in sample f, by linear combination of all V_{if} values.

We get

$$1n\ t_f = \Sigma C_i V_{if} + 4{,}6052 - \Sigma C_i\ 1n\ \bar{x}_i \qquad 19)$$

with $\Sigma C_i = 1$

the arithmetic mean

$$1n\ t_n = \frac{1}{n} \Sigma\ 1n\ t_f \qquad 20)$$

is then a result where all n jam samples are taken into consideration.

The sum of squared residuals from fitting the linear model is

$$s^2{}_y = \sum^{f(x_i..)} C_i C_j s^2\ 1n\ x_i;\ 1n\ x_i \qquad 21)$$

where again $s^2\ 1n\ x_i;\ 1n\ x_j$ denotes the element in the logarithmic dispersion matrix and $f(x_i..)$ that the summing shall be performed for all possible permutations of C_i.

Further, it can be shown that s^2_{yi} is minimized by choosing the weights C_i according to the following equation:

$$C_i = \frac{\sum_{j=1}^{m} s^2\ 1n\ x_i;\ 1n\ x_j^{\,-1}}{\sum_{i,j=1}^{m} s^2\ 1n\ x_i;\ 1n\ x^{j\,-1}} \qquad 22)$$

where $s^2\ 1n\ x_i;\ 1n\ x_j{}^{-1}$ are the elements in the inverted logarithmic dispersion matrix.

Both methods mentioned above require a calculation of the correction term

$$Y_{0f} \times k_{if} = Y_{0f} \times \frac{S^1 m_i^1 + S^2 m_i^2 + \text{---} S^m m_i^m}{S^0 m_0^0 + S^1 m_0^1 + \text{---} S^m m_0^m} \qquad 23)$$

We can split the right side into two terms and get

$$Y_{0f} \times k_{if} = Y_{0f} \times \frac{S^1 m_i^1 + S^2 m_i^2}{S^0 m_0^0 + \text{--} S_0^m}$$

$$+ Y_{0f} \times \frac{S^3 m_i^3 + \text{--} S^m m_i^m}{S^0 m_0^0 + S^1 m_0^1 + \text{--} S^m m_0^m} \qquad 24)$$

setting

$$\frac{Y_{0f}}{S^0 m_0^0 + \text{--} S^m m_0^m} = J_{0f} \quad \text{and}$$

$$S^1 m_i^1 + S^2 m_i^2 = KS_{if}$$

$$S^3 m^3 + \text{--} S^m m_i^m = KU_{if}$$

we get

$$Y_{0f} \times k_{if} = J_{0f} \times KS_{if} + J_{0f} \times KU_{if} \qquad 25)$$

the common factor J_{0f} is a measure of the concentrating of soluble solids from the recipe-mix to the final jam. It will be possible to give a max. – and a min. – value for this factor. We get a high value when the fruit content is high and sugar solutions are used for instance glucose-sirup or invert-sugar solutions. We get a low value when the fruit content is low and solid sugars are used. Using the results of various analyses of the jam, (e.g. water insoluble solids, glucose-sirup content and invert-sugar content) it is possible to set up preliminary model formulae for the jam in question, thus giving a high, a low and the most probable value for J_{0f}.

Likewise by analysing various samples of sugars, pectins, acids etc. it will be possible (taking into proper consideration factors like firmness, acid content etc. of the jam) to state a high, a low and the most probable value for KS_{if} and KU_{if}. Having these values, a highest, a lowest and the most probable value for t can be computed, using either equation 11) or 19).

Experimental

In the strawberry season 1970, samples of washed, deep frozen strawberries were collected from the main growing areas in Denmark. All samples were collected at random at the end of the freezing-band at four freezing plants. Samples were collected morning and afternoon. All samples, totalling 84, were stored at — 30° C until analysed.

From two dactories manufacturing strawberry jam, samples of the finished product were taken during the cooking season, autumn 1970. Samples were taken directly from the production-line at the end of the cooling band.

Samples of raw material, other than strawberry, used according to recipe were also collected. New samples were taken everytime new deliveries arrived at the factory. These latter samples were mixed, so that only 2 to 5 samples of every raw material were analysed.

The strawberry samples (reference material) were analysed for the following constituents, using well known and proven methods.

x^1: Water insoluble substance (extraction in Soxhlet-apparatus with water).

x_2: Nitrate (Ion selective electrode).

x_3: Formol value.

x_4: Ash. (Dry ashing at 500° C).

x_5: Sodium. (Flame – phometry).

x_6: Potassium. (Flame – photometry).

x_8: Calcium. (Atomic – absorption spectrometry).

x_0: Magnesium. (Atomic – absorption spectrometry).

x_0: Phosphor. (Colorimetric as phosphovanadomolybdate complex).

The strawberry jam samples and the raw material samples were analysed for the same 9 constituents mentioned above. Further, the jams were analysed for:

a: Total soluble solide (refractometry).

b: Invert sugar content.

c: Total acid.

d: Benzoic acid.

e: Firmness (Breaking strength by Tarr-Baker).

Results

The results of the analyses of the reference material are given in table 1. Here also are given the means $\bar{x}$, the variance s^2_{xi}, the standard deviation s_{xi}, and the coefficient of variablility

$$\frac{s_{xi} \times 100}{\bar{x}_i}.$$

In table 2, 3, 4 and 5 the calculated dispersion matrix, the correlation matrix, the logarithmic dispersion matrix and the logarithmic correlation matrix are tabulated.

In table 6 and 7, the results of the analyses of the strawberry jam and the raw materials are given.

In table 8, three model-formulae for the two jams in question based upon these analyses are stated. Using these formulae and the results from the raw material analyses, the corrected mean $\bar{z}_i$ values are calculated. These values are also given in table 6 and 7 as $\bar{z}^H_i$ (the high value), $\bar{z}^L_i$ (the low value) and $\bar{z}^P_i$ (the most probable preliminary value).

Table 1. Reference-material

	Water in-soluble substance %	NO_3 mgeqv/kg	Formolvalue mgeqv/100 g	Ash %	Na mg/100g	K g/kg	Ca mg/10 g	Mg mg/10 g	P mg/10 g
	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
Source A									
n	26	26	26	26	26	26	26	26	26
$\bar{x}$	1,67	1,66	1,00	0,45	0,650	1,512	2,111	1,132	2,069
s	0,17	0,58	0,19	0,064	0,055	0,168	0,274	0,078	0,203
Source B									
n	17	17	17	17	17	17	17	17	17
$\bar{x}$	1,78	1,10	0,96	0,45	0,56	1,50	2,49	1,18	2,07
s	0,184	0,212	0,153	0,060	0,036	0,120	0,319	0,078	0,157
Soucre C									
n	17	17	17	17	17	17	17	17	17
$\bar{x}$	1,78	1,15	1,00	0,46	0,56	1,52	2,45	1,19	2,05
s	0,191	0,194	0,161	0,052	0,038	0,096	0,368	0,082	0,146
Soucre D									
n	24	24	24	24	24	24	24	24	24
$\bar{x}$	1,72	1,31	0,80	0,43	0,61	1,49	2,60	1,23	2,14
s	0,211	0,239	0,151	0,047	0,037	0,123	0,381	0,101	0,167
Total									
n	84	84	84	84	84	84	84	84	84
$\bar{x}$	1,70	1,37	0,94	0,44	0,61	1,50	2,36	1,18	2,09
s^2	0,0331	0,1886	0,0373	0,0029	0,0030	0,0189	0,1463	0,0085	0,0305
s	0,1820	0,4343	0,1931	0,0539	0,0548	0,1374	0,3825	0,0923	0,1746
$\frac{s \times 100}{\bar{x}}$	10,71	31,74	20,59	12,19	9,04	9,14	16,22	7,84	8,34

Table 2. Dispersion-matrix

X_1	X_2	X_3	X_4 å	X_5	X_6	X_7	X_8	X_9
0,0332	—0,0125	—0,0029	0,0030	—0,0007	0,0102	0,0251	0,0092	0,0153
	0,1886	0,0423	0,0088	0,0109	0,0123	—0,0191	—0,0025	0,0062
		0,0373	0,0032	0,0014	0,0054	—0,0213	—0,0022	0,0030
			0,0029	0,0001	0,0051	—0,0006	0,0010	0,0040
				0,0039	0,0008	—0,0036	—0,0003	0,0007
					0,0189	—0,0061	0,0052	0,0152
						0,1463	0,0192	0,0162
							0,0085	0,0078
								0,0305

Table 3. Correlations-matrix

X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
1	—0,1582	—0,0839	0,3108	0,0131	0,4062	0,3613	0,5451	0,4810
	1	0,5044	0,3737	0,4227	0,2064	—0,1148	—0,0622	0,0813
		1	0,3164	0,0934	0,2050	—0,2883	—0,1226	0,0871
			1	0,1219	0,6844	—0,0271	0,2041	0,4291
				1	0,1778	—0,1599	—0,0658	0,1269
					1	—0,1154	0,4080	0,6343
						1	0,5432	0,2426
							1	0,4808
								1

Table 4. Logarithmic dispersion-matrix

X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
0,0111	—0,0048	—0,0020	0,0043	—0,0006	0,0042	0,0057	0,0045	0,0042
	0,0810	0,0271	0,0111	0,0128	0,0050	—0,0038	—0,0005	0,0019
		0,0425	0,0072	0,0022	0,0036	—0,0105	—0,0024	0,0009
			0,0139	0,0008	0,0077	—0,0012	0,0020	0,0043
				0,0096	0,0008	—0,0026	—0,0005	0,0006
					0,0087	—0,0020	0,0031	0,0048
						0,0256	0,0065	0,0030
							0,0061	0,0031
								0,0070

Table 5. Logarithmic correlation-matrix

X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
1	—0,1592	—0,0940	0,3442	—0,0165	0,4212	0,3386	0,5419	0,4734
	1	0,4620	0,3312	0,4672	0,1889	—0,0826	—0,0241	0,0798
		1	0,3025	0,0850	0,1844	—0,3187	—0,1469	0,0497
			1	0,1421	0,6954	—0,0599	0,2153	0,4389
				1	0,1679	—0,1776	—0,0712	0,1329
					1	—0,1367	0,4185	0,6185
						1	0,5246	0,2214
							1	0,4760
								1

Table 6. Content of index-constituent in Strawberry-jam

Factory		n	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	RT %	Inv. %	Benzoic-acid ‰
E	$\bar{y}_i$	20	0,62	0,81	0,32	0,18	2,00	0,489	1,575	0,521	0,723	65,17	26,83	0,75
	$\bar{y}_i^H$	20	0,71	0,91	0,38	0,23	2,25	0,528	1,370	0,590	0,780	65,81	33,9	
	$\bar{y}_i^L$	20	0,55	0,69	0,26	0,14	1,75	0,440	0,990	0,490	0,605	64,42	21,95	
	$\bar{z}_i^P$	20	0,62	0,44	0,32	0,13		0,459	0,919	0,501	0,704			
	$\bar{z}_i^H$	20	0,71	0,66	0,38	0,19		0,507	1,270	0,570	0,760		Breaking	
	$\bar{z}_i^L$	20	0,55	0,32	0,26	0,08		0,410	0,810	0,465	0,585		strength:	
	s_z^2	20	0,0014	0,0031	0,0012	0,0005		0,0007	0,0080	0,0006	0,0024		59 g/cm²	
	s_z	20	0,0376	0,0556	0,0348	0,0223		0,02697	0,0892	0,0241	0,0490			
	$\frac{s_z \times 100}{\bar{z}_i}$	20	6,09	12,67	10,94	17,72		5,88	9,71	4,82	6,96			
Sucrose	m^{I}	4	0	0,27	0	0,003	0,49	0,036	0,054	0,007	0,016			
Invert-sugar	m^{II}	4	0	0,805	0	0,03	0,26	0,032	0,600	0,116	0,055	75%		
Pectin	m^{III}	3	0	31,22	0	1,74	11,38	1,948	30,73	1,620	1,128			
Acid	m^{IV}	3	0	7,58	0	0,18	0,60	0,026	0,435	0,242	0,528			
Preservatives	m^{V}	1	0	0	0	21,5	1597	0	0	0	0			

Table 7. Content of index-constituent in Strawberry-jam

Factory		n	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	RT %	Inv. %	Bonzoic-acid ‰
F	$\bar{y}_i$	5	0,46	0,89	0,28	0,18	2,51	0,356	0,812	0,353	0,539	64,51	18,29	0,85
	$\bar{y}_i^H$	5	0,57	0,99	0,34	0,19	2,85	0,386	0,845	0,390	0,623	64,95	22,24	
	$\bar{y}_i^L$	5	0,36	0,83	0,23	0,16	2,20	0,325	0,790	0,310	0,490	64,08	14,64	
	$\bar{z}_i^P$	5	0,46	0,52	0,28	0,13		0,33	0,58	0,33	0,52			
	$\bar{z}_i^H$	5	0,57	0,69	0,34	0,15		0,36	0,68	0,370	0,603		Breaking	
	$\bar{z}^L{}_i$	5	0,36	0,37	0,23	0,10		0,30	0,490	0,290	0,470		strength:	
	s_z^2	5	0,0064	0,0043	0,0018	0,0002		0,0006	0,0005	0,0009	0,0027		61 g/cm²	
	s_z	5	0,080	0,0659	0,0424	0,0130		0,0236	0,0214	0,0303	0,0524			
	$\frac{s_z \times 100}{\bar{z}_i}$	5	17,1	12,7	15,2	10,2		7,2	3,7	9,1	10,1			
Sucrose	m^{I}	1	0	0,25	0	0,05	0,02	0,034	0,185	0,020	0,020			
Invert-sugar	m^{II}	0	–	–	–	–	–	–	–	–	–			
Pectin	m^{III}	1	0	30,94	0	1,61	10,40	1,855	29,00	1,440	1,195			
Acid	m^{IV}	1	0	19,49	0	0,05	0,08	0,076	0,790	0,050	0,010			
Preservatives	m^{V}	1	0	0	0	21,5	1597	0	0	0	0			

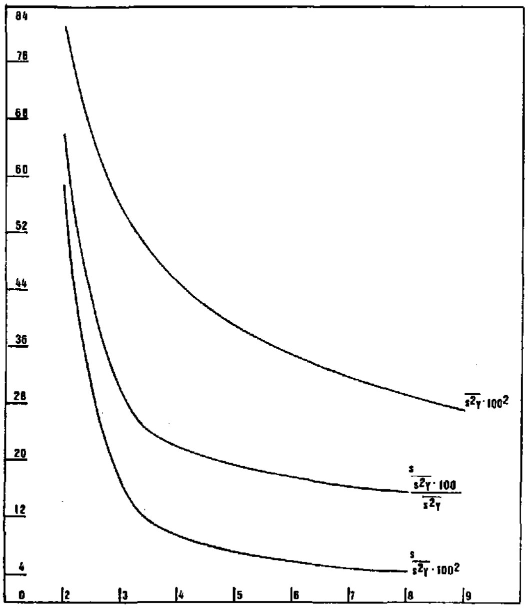
Table 8. Model recipes

	Factory E			Factory F		
	Highest value	Lowest value	Probable value	Highest value	Lowest value	Probable value
Fruit............	40%	20%	35%	40%	20%	30%
Sucrose	45%	75%	55%	60%	80%	70%
Invert-sugar	15%	5%	10%	10%	0%	0%
Pectin...........	5‰	2‰	3,5‰	5‰	2‰	3,5‰
Acid.............	10‰	6,6‰	8,3‰	5‰	3‰	4‰
Preservative.......	1,5‰	1‰	1‰	1,5‰	1‰	1‰

Based upon the analysis results of the reference material, all regression-coefficients b_i and all weights C_i were computed, together with the corresponding variance s_y^2 for all possible index-constituents combinations (totalling 502 comb.).

A reason for tabulating all these numbers can

Diagram 1.



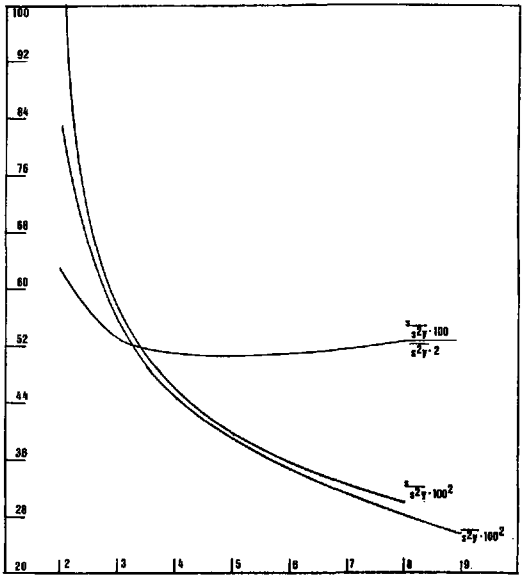
Number of index-constituents in combination.

hardly be found, since they can be displayed more easily and illustratively otherwise. Therefore, in diagram 1 (method 1) I have displayed the dependence of the mean $\bar{s}_y^2$ of all s_y^2 – numbers within each group containing either 2, 3 etc. index-constituents in relation to numbers of compounds in the combination. Also, in diagram 1 the standard deviation $s_{sy}^{\bar{2}}$ as well as the coefficient of variability

$$\frac{s_{sy}^{\bar{2}} \times 100}{s_y^2}$$

are given.

Diagram 2.

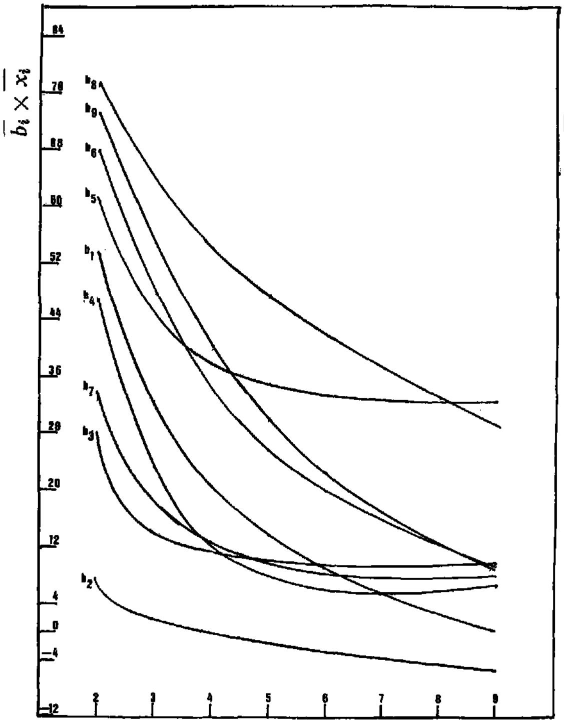


Number of index-constituents in combination.

Likewise, in diagram 2 you will find the values calculated according to method 2.

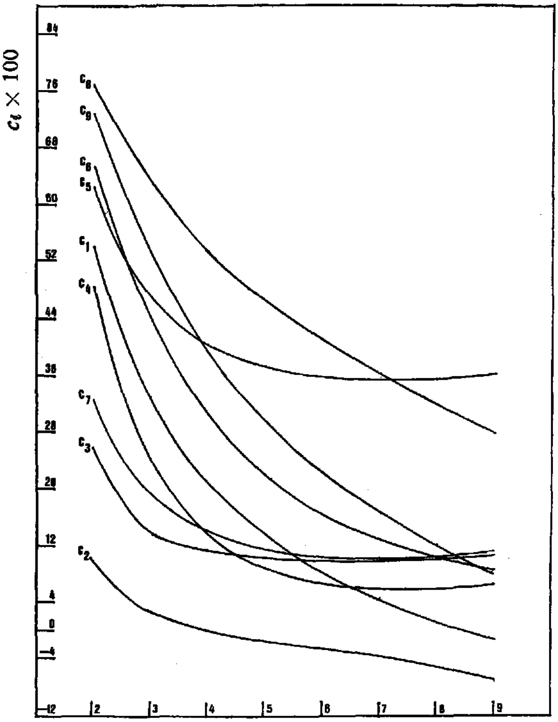
In diagram 3, the dependence of the means of the regression-coefficients b_i multiplied with $\bar{x}_i$ (reference-material) within each group containing 2, 3 etc. indexconstituents is displayed.

Diagram 3.



Number of index-constituents in combination.

Diagram 4.



Number of index-constituents in combination.

Finally, in diagram 4 the means of the weights C_i are illustrated in the same way.

Doing so, we get an idea of how much (in percentage) every index-constituent contributes on the average to 100% fruit content, as well as how much this contribution varies according to variation among the index-constituents used within each group.

Discussion

Both diagrams 1 and 2, show that the means s_y^2, after an abrupt fall in the beginning, very soon become almost constant.

This behaviour indicates that in practice it will not be advisable or economical to use more than 5 to 8 index-constituents.

We can get an impression of the main difference between method 1 and 2 by looking at the curves for the coefficient of variability of above mentioned means i.e.

$$\frac{s_{s^2y} \times 100}{s_y^2}$$

For method 1, these coefficients show the same tendency as s_y^2, i.e. an approach to an almost constant value after an abrupt fall in the beginning. In contrast with this, method 2 shows no abrupt fall but an almost constant value over the whole range.

This behaviour indicates that in our case method 1 would be preferable to method 2. Comparing the elements in the correlation matrices also indicates this.

From diagram 3 and 4 it's seen that the curves for some $\overline{b_i x_i}$ values and $\overline{C_i}$ 100 values very soon become almost parallel with the abscissae e.g. b_3, b_4, b_5, and b_7 as well as C_3, C_4, C_5, and C_7, whereas the others constantly diminish with increasing number of index-constituents in the combination. Many interesting speculations can arise from regarding this curious behaviour of certain index-constituents, but these speculations are purely mathematical-statistical in nature and as such outside the scope of this publication.

With the present material and using either method 1 or 2, we find considerable deviations between the result of fruit content determination in the two jam samples, depending on the index-constituent combination used. This is so, dispite the fact that many of these combinations give the same security limits on the calculated t-values.

The result thus confirms earlier findings, that the methods often fail when applied to commercially manufactured products.

The question now arises. Are our chemical analysis results wrong, or are the mathematical-statistical methods used not valid in our case?

As all analysis results are means of two or more determinations with small deviations and the analysis methods used have been well practised, the conclusion is that the fault must be within the assumptions.

These assumptions are mainly two.

1. The concentration of all index-constituents in all recipe-ingredients, except the fruit, is small and has small standard deviation.
2. The index-constituents x_i are multivariate – normally distributed.

Ré 1. As seen from table 6 and 7 we must conclude, especially when excluding sodium, that assumption 1 holds in our case. We are then forced to accept that assumption 2 is not justified.

In order to get an impression of the deviations from normality of the reference material the two quantities γ_1, and γ_2 given by the equations,

$$\gamma_1 = \frac{\mu_3}{\mu_2^{3/2}} \text{ and}$$

$$\gamma_2 = \frac{\mu_4}{\mu_2^2}$$

as well as Pearson's skewness factor given by the equation

$$Sk = \frac{\mu_3\mu_4 + 3\mu_2^2\,\mu_3}{(10\mu_2\mu_4 - 18\mu_2^3 - 12\mu_3^2)\sqrt{\mu_2}}$$

are calculated where μ_2, μ_3, and μ_4 denote the second, third, and fourth factorial moments. The calculated moments used are corrected for grouping using Sheppard's corrections.

Although the reference material is too small to justify a precise assertion of the exact shape of the distribution curves, they nevertheless give som indication of how these curves deviate from normal ones. From table 9, where these values are tabulated, we find that γ_2 for the variables x_2, x_7 are > 0 and for all others < 0. As Pearson's skewness factor also departs from 0 to a less or greater extent, we must conclude that the distribution curves deviate considerably from normal ones.

In the case of the variable x_6, the above mentioned calculations have been omitted, because this variable is apparently bimodal-distributed.

In order to further investigate the nature of the reference material, I have tested whether the means and the variances calculated for sampling place are possibly alike. The results of these tests is specified in table 10 and 11.

The result indicates that there exist variance-homogeniety in the referencematerial between sampling places, but that the means in several cases are different.

Tests whether variance-homogeneity between the jam-samples and the reference material exist, using their dispersion-matrices for comparison has also been performed. The test method used is described by *Box* (5), see also *Kendall* (10).

The result is that homogeneity does not exist.

All these tests have convinced me that the assumption 2, presupposed by the theory is not

Table 9. Measures of kurtosis and skewness

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
γ_1	0,2722	2,5784	0,1311	0,9424	0,2660	—	0,3807	0,2348	1,2495
γ_2	—0,9541	7,5828	—2,8712	—0,0767	—0,3978	—	0,5895	—0,9586	—1,6509
SK	0,8749	4,3493	0,0242	9,7102	0,2078	—	—0,1553	—0,6755	0,2338

Table 10. Test for variance homogenity

Sources	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
A : B	x		x	x	x	x	x	x	x
A :C	x		x	x	x		x	x	x
A : D	x		x	x	x	x	x	x	x
B : C	x	x	x	x	x	x	x	x	x
B : D	x	x	x	x	x	x	x	x	x
C : D	x	x	x	x	x	x	x	x	x

x indicates that the hypothesis hold at 95% level.

Table 11. Test for equal means

Sources	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
A : B	x		x	x					x
A : C	x		x	x					x
A : D	x			x					x
B : C	x	x	x	x	x	x	x	x	x
B : D	x					x	x		x
C : D	x					x	x		x

x indicates that the hypothesis about equal means hold at 95% level.

justified in the present case. The question now arises, on what considerations our selection of the most suitable index-constituent combination should be based.

One method seems natural.

As we have found that the index-constituent is not normally distributed, we could possibly transform the x_i-values so that the new variables were more normal-distributed, that is, we could fit a type of Pearson curve to the analysis results and then transform the values.

However, the present material does not warrant this immense computational task. In fact, more than 500 samples would be necessary to justify such a transformation, without any guarantee that the final result would be better, especially considering the lack of variance-homogeneity between jam-samples and reference-material.

Therefore, considering the above mentioned deviations from normality and the lack of variance-homogenity, it seems reasonable that a combination where one index-constituent accounts for a considerable percentage of the calculated fruit content would be inferior to a combination where all index-constituents are alike. It would also seem likely, that an index-constituent with very great correction-terms would be inferior to one where these correction-terms are diminutive.

Based on these considerations I have set up two equations which allow us to discriminate between the various combinations.

These equations are:

1. Method

$$U_b = \frac{n}{n-1} \times \sum b_i\, b_j\, s^2\, x_i, x_j$$

$$\left[1 + \sum \left|\frac{1}{n} - \frac{b_i \bar{x}_i}{100}\right| \frac{\overline{Y}_i^P}{\overline{Z}_i^P}\right]$$

$$\text{with } \sum \frac{b_i \bar{x}_i}{100} = 1$$

and for

2. Method

$$U_C = \frac{n}{n-1} \times \sum C_i C_j\, s^2\, 1n\, x_i, 1n\, x_j$$

$$\left[1 + \sum \left|\frac{1}{n} - C_i\right| \frac{\overline{Y}_i^P}{\overline{Z}_i^P}\right] \text{with } \sum C_i = 1$$

where n is the number of index-constituents in the combination.

The value of U_b and U_C are esaily computed together with the regression coefficients b_i and the weights C_i.

By choosing the combination for which either U_b or U_C is at the lowest, we are assured that the influence of all deviations from the assumptions made in the theory are diminished.

For both jam-samples under consideration, the above stated equations have given that the best suited combination would be the one containing the following compounds, x_1, x_2, x_3, x_4, x_6, x_7, and x_9 and using method 1.

In table 12 the fruit contents in the jam samples are tabulated, using this combination and method 1. Also given are the figures for the fruit content as received from the two factories after this investigation had been finished.

In our case where we are assured that the two factories have not altered their recipes during

Table 12. Combination 1-2-3-4-6-7-9

Method 1

Source	Sample	Fruit-content high	Fruit-content low	Fruit-content probable
E	90	33,20	32,19	32,66
	91	33,60	32,67	33,06
	92	30,84	29,98	30,69
	93	32,93	31,74	32,39
	94	38,09	36,91	37,55
	95	33,49	32,79	33,26
	96	34,15	33,14	33,61
	97	33,11	32,10	32,59
	98	35,19	34,46	34,68
	99	36,16	35,15	36,62
	100	35,66	34,65	35,12
	101	35,55	32,54	33,01
	102	35,16	34,15	34,62
	103	35,07	34,06	34,53
	104	33,96	32,95	33,42
	105	35,42	34,40	34,87
	106	33,36	32,34	32,81
	107	35,23	34,22	34,67
	108	35,34	34,33	34,77
	109	37,38	36,29	36,73
	$\bar{t}$	34,5	33,6	34,1 ± 1,4

»True value« 32,5 % ± 0,5 %

Source	Sample	Fruit-content high	Fruit-content low	Fruit-content probable
F	85	25,84	24,61	25,15
	86	26,54	25,31	25,85
	87	22,64	21,41	21,95
	88	23,68	22,45	22,99
	89	26,82	25,59	25,13
	$\bar{t}$	25,1	23,9	24,2 ± 3,4

»True value« 23,7 % ± 0,5 %

the sampling period, the method used above for calculating the fruit content and the statistical security limits can be regarded as valid. In other cases, especially when recipe-alteration can be suspected, it will be more appropriate to perform a discriminant analysis.

This has been done for purely illustrative purposes, using the 5 jam-samples from factory 2.

Using the results from reference-material analysis and the combination found most suitable, two jam populations can be constructed, one population containing 20% fruit in the jam, and the other 35%. Then an discriminant equation can be set up in the usual way, which by inserting the corrected Z_{if} values assign the jam-sample to either the population containing 20 to 27,5% fruit or to the population containing 27,5–35% fruit.

The result for factory 2 samples is, that all 5 samples belong to population 20–27,5% fruit content with the probalility of misclassification in the order of 0,4%.

Conclusion

The shortcomings of many methods of fruit content determination in jams and preserves are according to this investigation, due to the fact, that the frequency-distribution of the index-constituents used, deviates considerably from a normal distribution. Hence the statistic used (regression-analysis) is not valid. A method is developed which to a certain extent remedies these difficulties. Using the methods, we have succeeded in determining the fruit content in two strawberry jams with reasonable accuracy, despite the fact, that the index-constituent content in these jams in several ways must be regarded as outliers compared with the same in the reference material.

The method developed can eliminate considerable deviations from normality as well as uncertainties regarding the cooking-recipe used and the quality of the other raw materials, besides fruit, used in the production.

Acknowledgements

The author thanks lic.agro. Poul Hansen for valuable aid concerning the many analyses, lic. agro. Karl Sandvad for the regression analyses and lic.agro. Karl Kaack for many valuable discussions throughout this work, also thanks to stud. mag. Sten Schousboe for aid with the English translation.

References

1) *Acker, L.* (1969). Über Verfälschungen von Lebensmitteln. Die Naturwissenschaften 56. 72-77.
2) *Baier, E. & Hse, P.* (1908). Zeitschrift für Lebensmittel-Untersuchung und-Forschung 15. (140-00).
3) *Beythien, A.* (1903). Zeitschrift für Lebensmittel-Untersuchung und-Forschung 6. (1095-0000).
4) *Beythien, A.* und *P. Simmich* (1910). Zeitschrift für Lebensmittel-Untersuchung und-Forschung 20. (241-000).
5) *Box, G. E. P.* (1949). A general distribution theory for a class of likelihood criteria. Biometrica. 36. 317.
6) *Choureitah, A. & Lenz, F.* (1971) Inhaltstoffe bei Erdbeeren (Senga Sengana) in abhängigkeit von Ernährung und Fruchtbetrag. Der Erwerbs Ostbau 13. 133-136.
7) *Goodall, H.* (1969) The Composition of Fruits. B.F.M.I.R.A. Scientific and Technical Surveyes 59. 1-101.
8) *Härtel, F. & Sölling, J.* (1911) Zeitschrift für Lebensmitteluntersuchung und - Forschung 21 (168, 553).
9) *Kefford, J. F.* (1969) Analytical Problems with Fruit Products. Food Preservation Quarterly 29. 65-71.
10) *Kendall, M. G. & Stuart, A.* The Advanced Theory of Statistics. London: Ch. Griffin. Vol. 1 3'ed 1969, vol. 2 2'ed 1967, vol. 3 2'ed 1968.
11) *Kenworthy, A. L. & Harris, N.* (1963) Composition of McIntosh, Red Delicious and Golden Delicious apples as related to environment and season.
12) *Klinger, H. & Nehring, P.* (1965) Zur Beurteilung von Obstkonfitüren. Zeitschrift für Lebensmittel-Untersuchung und - Forschung 129, 76-83.
13) *Ludwig, W.* (1907). Zeitschrift für Lebensmittel-Untersuchung und - Forschung 13. (5-0).
14) *Macara, T.* (1931) The composition of fruits as used for jam manufacture in Great Britain. Analyst 56. 35-43.
15) *Macara, T.* (1935) The composition of raspberries. Analyst 60. 592-595.
16) *Nehring, P. & Klinger, H.* (1965) Zur beurteilung von Obstkonfifüren. Zeitschrift für Lebensmittel-Untersuchung und - Forschung 129. 1-9.
17) *Rahman, A. A., Shalaby, A. F. & Monayeri, M.O.* El. (1971) Effect of Moisture Stress on Metabolic Products and Ion Accumulation. Plant and Soil 34. 65-90.
18) *Rolle, L. A. & Vandercook, C. E.* (1963) Lemon Juice Composition. III. Characterization of California-Arizona Lemon Juice by Use of a Multiple Regression Analysis. J. Ass.off.agric. Chem. 46. 362-365.
19) *Steiner, E. H.* (1948) Application of Statical Methods in Calculating Proportions of Ingredients in certain Food Products. Analyst 73. 15-30.
20) *Steiner, E. H.* (1949) The Statical Use of Several Analytical Constituents for Calculating Proportions of Ingredients in Certain Food Products. Analyst 74. 429-438.
21) *Vandercook, C. E., Rolle, L. A. & Ikeda, R. M.* (1963) Lemon Juice Composition. I. Characterization of California-Arizona Lemon Juice by It's Total Acid and 1-Malic Acid Content J. Ass. off. agric. Chem. 46. 353-358.
22) *Vandercook, C. E. & Rolle, L. A.* (1963) Lemon Juice Compostion. II. Characterization of California-Arizona Lemon Juice by its Polyphenolic Content. J. Ass. off. agric. Chem. 46. 359-362.
23) *Vandercook, C. E. & Guerrero, H. C.* (1969) Citrus juice characterization. Analysis of the phosphorus fraction. J. agric. Fd. Chem. 17. 626-628.