HUNGARIAN JOURNAL
OF INDUSTRIAL CHEMISTRY
VESZPRÉM
Vol. 29. pp. 123 – 127 (2001)

OBTAINING QUANTITATIVE INFORMATION ON THE FLUCTUATION OF THE ACTIVE INGREDIENT CONTENT IN DRUGS – WHAT WOULD THE CUSTOMER FIND

Á. DRÉGELYI-KISS and S. KEMÉNY

(Department of Chemical Engineering, Budapest University of Technology and Economics,
H-1521 Budapest, HUNGARY)

Received: October 8, 2001

This paper was presented at the 7th International Workshop on Chemical Engineering Mathematics, Bad Honnef, Germany, August 12-17, 2001

The active ingredient content of tablets is not uniform due to inhomogeneity and the fluctuation of the process circumstances. Moreover, the measured data are subject to measurement (analytical) error. Both the consumer and the producer should be aware of the possible range of active ingredient content of the tablets. The analysis of variance technique was used in the context of nested designs. Several variance components and their confidence ranges were calculated utilising the Satterthwaite-approximation.
The customers may also control the product quality. Our purpose is to study the measurement process, as the customer would perform it, raising the question on the range in which the customer finds the amount of the key compound in a tablet purchased at a pharmacy. Various cases are compared concerning the measurement precision and way of chemical analysis performed by the customer, calculating the ranges in which the active ingredient content could be found with 95% probability. The width of these ranges may be affected by the bias of the Satterthwaite-approximation.

Keywords: nested design, Satterthwaite-approximation, confidence intervals, variance components, drug analysis

Introduction

In pharmaceutical industries there are strict guidelines to check the manufacturing processes in order to assure the steadiness of quality. The companies have to elaborate their own specifications related to the processes, chemical analysis, etc. These guidelines contain the appropriate design of experiments, where it can be seen how to perform the measures and statistical methods to appraise the results, for instance giving the confidence interval for the expected value in a 3x3 design. In this paper we examine the relevant guidelines and ask some questions from the customer's point of view.

Data source

Table 1 contains data obtained from a real manufacturing process in the course of the current guideline and process validation of the factory. During the batch-wise production of drugs these tablets were collected in lose-boxes. Tablets of one batch of the finished products are collected to 13 or 14 lose-box. The first five boxes are called the beginning of the batch (first fraction), the 6-9/10th boxes are the middle and the 10/11-13/14th boxes are the end of the batch. In order to control the process the active ingredient contents have to be measured in drugs. During the sampling 2-3 tablets were taken from the three different fractions of three batches, they were pulverized and powder fractions (altogether 9) were analysed three times each. The analytical procedure was high-temperature HPLC with low-wavelength detection. These data are shown in *Table 1*. The declared active ingredient content of the tablets calculated for the average mass of tablet is 2.5 mg ±5%, i.e. 2.375 mg - 2.625 mg, thus these samples have met this requirement.

Contact information: E-mail: dregelyi.vmt@chem.bme.hu, kemeny.vmt@chem.bme.hu

Table 1 The active ingredient content of drugs in several batches, from different fractions of batches and with repeated chemical analysis

Batch	*Sampling Fraction*	*Mass [mg]*	*Batch*	*Sampling Fraction*	*Mass [mg]*	*Batch*	*Sampling Fraction*	*Mass [mg]*
1	first	2.60	2	first	2.58	3	first	2.55
1	first	2.59	2	first	2.57	3	first	2.56
1	first	2.60	2	first	2.56	3	first	2.58
1	middle	2.62	2	middle	2.58	3	middle	2.56
1	middle	2.60	2	middle	2.58	3	middle	2.60
1	middle	2.62	2	middle	2.58	3	middle	2.57
1	end	2.57	2	end	2.59	3	end	2.57
1	end	2.57	2	end	2.59	3	end	2.56
1	end	2.58	2	end	2.57	3	end	2.57

ANOVA and variance components

Model

The measured data were processed using analysis of variance technique (ANOVA) and the Statistica for Windows software was used for calculations.

The experimental design contains batch as random factor with 3 levels (1, 2, 3), sampling fraction as random factor with 3 levels (first, middle, end), and analysis repeated three times as repetition. The sampling fraction factor is nested within batches.

The factors are:

α: batch (r=3 levels)

$\beta(\alpha)$: fraction within a batch (q=3 levels)

Thus the measurements are assumed to follow the nested-random-effects model:

$$y_{ijk} = \mu + \alpha_i + \beta_{j(i)} + \varepsilon_{ijk}$$
$$i = 1,...,r; j = 1,...,q; k = 1,...,p \qquad (1)$$

where μ is the expected value, α_i is the random effect of the i^{th} batch, $\beta_{j(i)}$ is the random effect of the j^{th} fraction within the i^{th} batch, and ε_{ijk} is the random noise for the k^{th} measurement taken from the j^{th} fraction of the i^{th} batch.

Certain assumptions have to be fulfilled when calculating ANOVA. Assume that α_i, $\beta_{j(i)}$ and ε_{ijk} are independent and identically distributed variables with normal distribution, mean 0 and variance σ_A^2, $\sigma_{B(A)}^2$ and σ_e^2, respectively.

The null hypotheses:

$H_0^A: \sigma_A^2 = 0$, i.e. there is no batch effect.

$H_0^B: \sigma_{B(A)}^2 = 0$, i.e. the sampling fractions are not different (there is no inhomogeneity).

The theoretical ANOVA table is found as *Table 2* with the calculated and expected mean squares. the terms used for F-tests to check the null-hypotheses.

Results of ANOVA calculations

The homoscedasticity and normality requirements are checked with positive results. The analysis of variance results are shown in *Table 3*. There is no significant difference between batches, but the inhomogeneity is significant at 0.05 level.

As in the F test the batch mean square is compared with the mean square of the sampling fraction, the large value of the latter may cover the otherwise important effect of batches. This was checked by calculating the probability of the error of second kind (β) for a fixed probability of the error of first kind, α=0.05 .

The alternative hypothesis considered for the calculation is the value of variance found as point estimate:

$$H_1: \sigma_A^2 = \hat{\sigma}_A^2 \qquad (2)$$

This means that the question is the probability of not detecting a variance of the size really estimated ($\hat{\sigma}_A^2 = 1.13 \cdot 10^{-4}$, see later).

The probability of not detecting is:

$$\beta = P\left(F < F_\alpha \middle| H_1\right) = P\left(F \frac{E\left(s_A^2\right)}{E\left(s_{B(A)}^2\right)} < F_\alpha\right) = P\left(F < F_\alpha \frac{3\sigma_{B(A)}^2 + \sigma_e^2}{9\sigma_A^2 + 3\sigma_{B(A)}^2 + \sigma_e^2}\right) \qquad (3)$$

Degrees of freedom for calculating the F_α critical value are: $\nu_{numerator}=\nu_A=2$, and $\nu_{denominator}=\nu_B=6$. The critical value itself is $F_{0.05}=5.14$. Thus the probability of the error of second kind is:

$$\beta = P\left(F < 5.14 \frac{5 \cdot 10^{-4}}{1.515 \cdot 10^{-3}}\right) = P(F < 1.696) = 0.74$$

The chance that the difference between batches remains unobserved is β=0.74 with α=0.05 . This risk is very high, thus it is advisable to keep the batch effect in the model instead of neglecting it.

Table 2 The theoretical ANOVA-table for two-way nested-random-effects model

Effect	Sum of Squares	df	Mean Squares	Expected MS	F_0
A	$S_A = qp\sum_i (\bar{y}_{i\cdot\cdot} - \bar{y}_{\cdot\cdot\cdot})^2$	r-1	$s_A^2 = \frac{S_A}{r-1}$	$qp\sigma_A^2 + p\sigma_B^2 + \sigma_e^2$	$\frac{s_A^2}{s_{B(A)}^2}$
$B(A)$	$S_{B(A)} = p\sum_i (\bar{y}_{ij\cdot} - \bar{y}_{i\cdot\cdot})^2$	$r(q$-1)	$s_{B(A)}^2 = \frac{S_{B(A)}}{r(q-1)}$	$p\sigma_B^2 + \sigma_e^2$	$\frac{s_{B(A)}^2}{s_R^2}$
Error	$S_R = \sum_i \sum_j \sum_k (y_{ijk} - \bar{y}_{ij\cdot})^2$	$rq(p$-1)	$s_R^2 = \frac{S_R}{rq(p-1)}$	σ_e^2	

Table 3 The ANOVA table: numerical evaluation

Effect	df	Mean Squares	Expected MS	F_0	p
A: batch	2	0.001515	$9\sigma_A^2 + 3\sigma_{B(A)}^2 + \sigma_e^2$	3.030	0.123
B(A): sampling fraction	6	0.000500	$3\sigma_{B(A)}^2 + \sigma_e^2$	3.970	0.011
Error	18	0.000126	σ_e^2		

It is important to estimate variance components (σ_A^2, $\sigma_{B(A)}^2$ and σ_e^2) in order to split the variance of the process into different parts. Its usual way is the method of moments or ANOVA method, where the estimates are obtained using the terms of expected mean squares in *Table 2*:

$$\hat{\sigma}_A^2 = \frac{s_A^2 - s_{B(A)}^2}{qp} = 1.13 \cdot 10^{-4} \quad (4.a)$$

$$\hat{\sigma}_{B(A)}^2 = \frac{s_{B(A)}^2 - s_R^2}{p} = 1.25 \cdot 10^{-4} \quad (4.b)$$

$$\hat{\sigma}_e^2 = s_R^2 = 1.26 \cdot 10^{-4} \quad (4.c)$$

The estimated variances are obviously of the same order of magnitude, thus neglecting the between-batch variation is not justified.

Computation of the content range relevant for the customer

Model

Two questions arise:

- What is the range for the active ingredient content of the tablet purchased by the customer at a pharmacy?
- What is the range in which the customer would find the content analysing a tablet?

In the first case (range for the true content) the error of the analysis does not affect the result, this is achieved by assuming an infinite number of repetitions ($p' \to \infty$). The interval in which the customer at 95% probability would measure the active ingredient content of a tablet depends on the precision of her own measurement system and on the number of repetitions in chemical analysis.

The statistical treatment is common for the two cases. Student's t distribution is used to calculate the range for the content on the customer's side.

A deviation variable (d) is introduced:

$$d = \bar{y}_{\cdot} - \bar{y}_{\cdot\cdot\cdot} \quad (5)$$

where $\bar{y}_{\cdot}$ is the average value measured by the customer, $\bar{y}_{\cdot\cdot\cdot}$ is the grand average measured by the manufacturer (calculated from *Table 1*, $\bar{y}_{\cdot\cdot\cdot}$=2.580).

The expected value of this d deviation is $E(d) = 0$. Its variance is a sum of two terms:

$$Var(d) = Var(\bar{y}_{\cdot}) + Var(\bar{y}_{\cdot\cdot\cdot}) \quad (6)$$

The two variances are added, because the error of the measurements by the manufacturer is independent from that at the customer. These variances are expressed in terms of the variance components:

$$Var(\bar{y}_{\cdot}) = \sigma_A^2 + \sigma_{B(A)}^2 + \frac{\sigma_e'^2}{p'} \quad (7)$$

where $\sigma_e'^2$ is the variance of measurement error obtained by the customer, p' is the customer's number of repetition,

$$Var(\bar{y}_{\cdot\cdot\cdot}) = \frac{1}{r}\sigma_A^2 + \frac{1}{rq}\sigma_{B(A)}^2 + \frac{\sigma_e^2}{rqp} \quad (8)$$

It may well be assumed that the analytical method and the measurement apparatus of the customer is analogous to the system used by the analytical laboratory of the manufacturer, thus the uncertainty of their measurements is equal ($\sigma_e'^2 = \sigma_e^2$). The number of repetitions may not be the same, however. Upon

Table 4 The dependence of customer's 95% range on the number of repetitions

p'	s_d^2	$t_{0.975,S}$	95% interval$_S$	width of the interval$_S$	$t_{0.975,w.a.}$	95% interval$_{w.a.}$	width of the interval$_{w.a.}$
1	$4.20\cdot10^{-4}$	2.413	2.531<y<2.630	0.099	3.370	2.511<y<2.649	0.138
3	$3.36\cdot10^{-4}$	2.742	2.530<$\overline{y}$.<2.631	0.100	3.688	2.513<$\overline{y}$.<2.648	0.135
5	$3.19\cdot10^{-4}$	2.858	2.529<$\overline{y}$.<2.631	0.102	3.772	2.513<$\overline{y}$.<2.648	0.135
10	$3.06\cdot10^{-4}$	2.966	2.529<$\overline{y}$.<2.632	0.104	3.841	2.513<$\overline{y}$.<2.648	0.134
20	$3.00\cdot10^{-4}$	3.029	2.528<$\overline{y}$.<2.633	0.105	3.877	2.513<$\overline{y}$.<2.648	0.134
∞	$2.94\cdot10^{-4}$	3.098	2.527<y<2.634	0.106	3.915	2.513<y<2.647	0.134
	prescribed		2.375<y <2.625	0.250			

Subscripts: *S* means calculating with Satterthwaite method, *w.a.* means using weighted average method

substitution the resulting variance for the d deviation variable is

$$Var(d)=\left(1+\frac{1}{r}\right)\sigma_A^2+\left(1+\frac{1}{rq}\right)\sigma_{B(A)}^2+\left(\frac{1}{p'}+\frac{1}{rqp}\right)\sigma_e^2 \quad (9)$$

As the variance components (σ_A^2, $\sigma_{B(A)}^2$, σ_e^2) are not known, they are estimated from the experimental data:

$$s_d^2=\left(1+\frac{1}{r}\right)\hat{\sigma}_A^2+\left(1+\frac{1}{rq}\right)\hat{\sigma}_{B(A)}^2+\left(\frac{1}{p'}+\frac{1}{rqp}\right)\hat{\sigma}_e^2 \quad (10)$$

Upon substituting Eqs.(4.a)–(4.c) for the estimates of variance components the following expression is obtained:

$$s_d^2=\frac{r+1}{rqp}s_A^2+\frac{q-1}{qp}s_{B(A)}^2+\left(-\frac{1}{p}+\frac{1}{p'}\right)s_R^2 \quad (11)$$

The interval, where the customer would find the average active ingredient content of a tablet at e.g. 95% probability, is calculated as:

$$P\left(\overline{y}...-t_{0.975,\nu}s_d<\overline{y}.<\overline{y}...+t_{0.975,\nu}s_d\right)=0.95 \quad (12)$$

The main difficulty of the further calculation lies in the fact that the estimator for the resulting variance, as a linear combination of mean squares, does not follow $\frac{\chi^2\sigma_d^2}{\nu}$ distribution, thus the range above is only approximate. According to the Satterthwaite approximation [1] the $\sum_i a_i s_i^2$ linear combination of mean squares is treated as if it were $\frac{\chi^2\sigma_d^2}{\nu}$, with degrees of freedom expressed as:

$$\nu=\frac{\left(\sum_i a_i s_i^2\right)^2}{\sum_i \frac{\left(a_i s_i^2\right)^2}{\nu_i}} \quad (13)$$

where s_i^2 is the i^{th} mean square, ν_i its degrees of freedom, a_i is the coefficient of the i^{th} mean square in the linear combination.

Another method [2] suggests calculating $t_{1-\alpha/2,\nu}$ as a weighted average of the appropriate t critical values related to the calculated mean squares and degrees of freedom:

$$t_{1-\alpha/2,\nu}=\frac{\sum_i a_i s_i^2 t_{1-\alpha/2,\nu_i}}{\sum_i a_i s_i^2} \quad (14)$$

Results and discussion

Eq. (11) gives the following expression for s_d^2 if r=3, q=3 and p=3 is substituted for the number of batches, number of sampling fractions and number of repetitions, respectively:

$$s_d^2=\frac{4}{27}s_A^2+\frac{2}{9}s_{B(A)}^2+\left(-\frac{1}{3}+\frac{1}{p'}\right)s_R^2 \quad (11.a)$$

The degrees of freedom using Satterthwaite's approximation is given as

$$\nu=\frac{\left(\frac{4}{27}s_A^2+\frac{2}{9}s_B^2+\left(-\frac{1}{3}+\frac{1}{p'}\right)s_R^2\right)^2}{\frac{\left(\frac{4}{27}s_A^2\right)^2}{\nu_A}+\frac{\left(\frac{2}{9}s_B^2\right)^2}{\nu_B}+\frac{\left(\left(-\frac{1}{3}+\frac{1}{p'}\right)s_R^2\right)^2}{\nu_R}} \quad (13.a)$$

The other method, averaging the critical t values takes the following form:

$$t_{1-\alpha/2,\nu}=\frac{\frac{4}{27}s_A^2\cdot t_{1-\alpha/2,\nu_A}+\frac{2}{9}s_{B(A)}^2\cdot t_{1-\alpha/2,\nu_B}+\left(-\frac{1}{3}+\frac{1}{p'}\right)s_R^2\cdot t_{1-\alpha/2,\nu_R}}{\frac{4}{27}s_A^2+\frac{2}{9}s_{B(A)}^2+\left(-\frac{1}{3}+\frac{1}{p'}\right)s_R^2} \quad (14.a)$$

Several p' values were taken for calculations, including $p'\to\infty$, the latter stands for the case of no analysis on the customer's side, giving the range for the true content.

The results for the 95% intervals are given in *Table 4*.

It is well seen that neither the true content nor the values to be obtained by the customer upon chemical analysis at 95% probability are within the required range, there is a clear overage. At the same time the range of uncertainty is much narrower than it would be

allowed. The uncertainty here means not only the measurement error but also batch differences and inhomogeneity within batches, as the tablet purchased by the customer may come from any batch and from any sampling fraction of a batch.

What is surprising in *Table 4* is, that the width of the 95% range obtained using Satterthwaite approximation is increasing with p'. That means the more precise is the measurement due to more repetitions, the wider is the interval. The reason comes from Eq. (13.a): Increasing the number of repetitions (p') the numerator decreases, while the denominator is almost unchanged. This gives smaller degrees of freedom and larger critical t value at larger p'. This over-compensates the reduction of s_d^2 and slightly broadens the interval. This anomaly is the error of the approximation, but the numerical consequence is not serious.

The weighted average method results in wider intervals.

Conclusion

In pharmaceutical industry there are strict guidelines regarding e.g. active ingredient content. The problem is that the prescribed interval is related to the average measured by a laboratory near the process or the average measured by the "customer" (this could be even the next laboratory). If we take the customer's uncertainty into consideration, the interval for the average may be evaluated. There are two methods to construct this interval; the first one uses the Satterthwaite-approximation, the second one calculates the average of critical t-values weighted by mean squares. Due to the bias of the approximation, the larger is the number of repetition, the broader is the width of the interval. The second method gives broader interval for the average. In spite of the fact that all tablets analysed, individually conform to the specifications, the interval in which values may occur is partly outside the specifications.

Acknowledgement

The authors wish to express their gratitude to Dr. K. KOLLÁR-HUNEK (Department of Chemical Informatics, Budapest University of Technology and Economics BUTE). The work has been supported by the Hungarian National Research Foundation OTKA under contract number T0033005 and Varga József Foundation of the Chemical Engineering Faculty of the BUTE.

SYMBOLS

d deviation variable
F Fisher test statistic
H hypothesis
p number of repetitions in the producer's laboratory
p' number of repetitions performed by the customer
q number of levels of the fraction factor
r number of levels of the batch factor
s^2 mean square
$t_{\alpha,\nu}$ critical value of Student-distribution for α probability and ν degrees of freedom
y_{ijk} the measured value
$\bar{y}.$ the average value measured by the customer
$\bar{y}...$ the average value measured by the laboratory of the manufacturer

Greek symbols

α the probability of the error of first kind
α_i the effect of the batch
β the probability of the error of second kind
$\beta_{j(i)}$ the effect of the fraction within the batch
ε_{ijk} the random noise
ν degrees of freedom
μ the expected value
σ variance component

Subscript/Superscript

$'$ related to the customer
A related to the factor of the batch
B related to the factor of the fraction
d related to the deviation variable
r related to the error
R related to the error

REFERENCES

1. SATTERTHWAITE F. E.: Biometrics Bull., *1946*, 2(6), 110-114
2. PAARK K. and BURDICK R. K.: Commun. Statist. – Theory Meth., *1998*, 27(11), 2807-2825
3. BOX G. E. P., HUNTER W. G., HUNTER J. S.: Statistics for experimenters, J. Wiley and Sons, 1978
4. LORENZEN T. J., ANDERSON V. L.: Design of experiments. A no-name approach, Marcel Dekker, 1993
5. KEMÉNY S., DEÁK A.: Design of experiments, Műszaki Könyvkiadó, Budapest, 2000 (in Hungarian)