# ACB spreadsheet verification: precision estimates by variance component analysis

*Ed Wilkes*

This document describes the verification of the variance component analysis spreadsheet, written by Prof Anders Kallner, that performs calculations for the assessment of assay precision (imprecision) and [optionally] trueness (bias) (July 2018 version). Calculations performed by these spreadsheets were verified in an independent statistical software (the R statistical computing environment v3.4.1) by the author of this document. The R packages required to run this code are shown below. This code can be copied and pasted into an instance of R and, given the test data as input, reproduce the analysis in this document.

### Required packages:

```
require(dplyr)
require(ggplot2)
require(knitr)
require(outliers)
require(VCA)
```

---

### Reading data into R:

```
# Read in csv file: "2018-07 ACB Precision (imprecision) and trueness (bias) - test data 1.csv"
df_1 <- read.csv(file.choose(), header = TRUE)

# Read in csv file: "2018-07 ACB Precision (imprecision) and trueness (bias) - test data 2.csv"
df_2 <- read.csv(file.choose(), header = TRUE)
```

---

### Check calculation of means, SE, SD, and CV for data set 1:

These are the data presented in cells C21:G24.

```
df_1 %>%
  group_by(day) %>%
  summarise(mean = round(mean(value), digits = 2)
            ,n = n()
            ,sd = round(sd(value), digits = 3)
            ,sem = round(sd(value) / sqrt(n), digits = 2)
            ,cv = round(sd / mean * 100, digits = 2)
  ) %>%
  kable
```

| day | mean | n | sd | sem | cv |
|---|---|---|---|---|---|

| day | mean | n | sd | sem | cv |
|---|---|---|---|---|---|
| 1 | 9.83 | 5 | 0.308 | 0.14 | 3.13 |
| 2 | 9.68 | 5 | 0.537 | 0.24 | 5.55 |
| 3 | 9.53 | 5 | 0.491 | 0.22 | 5.15 |
| 4 | 10.27 | 5 | 0.325 | 0.15 | 3.16 |
| 5 | 9.86 | 5 | 0.359 | 0.16 | 3.64 |

The calculated values match those in the spreadsheet.

---

### Check calculation of means, SE, SD, and CV for data set 2:

```
df_2 %>%
  group_by(day) %>%
  summarise(mean = round(mean(value), digits = 2)
            ,n = n()
            ,sd = round(sd(value), digits = 3)
            ,sem = round(sd(value) / sqrt(n), digits = 2)
            ,cv = round(sd / mean * 100, digits = 2)
  ) %>%
  kable
```

| day | mean | n | sd | sem | cv |
|---|---|---|---|---|---|
| 1 | 43.59 | 5 | 0.637 | 0.28 | 1.46 |
| 2 | 44.61 | 5 | 0.595 | 0.27 | 1.33 |
| 3 | 43.78 | 5 | 2.627 | 1.17 | 6.00 |
| 4 | 43.89 | 5 | 0.265 | 0.12 | 0.60 |
| 5 | 43.89 | 5 | 0.265 | 0.12 | 0.60 |

The calculated values match those in the spreadsheet.

---

### **Plot data**:

```r
ggplot(df_1, aes(x = day, y = value))+
  stat_summary(fun.data = mean_se, geom = "errorbar", width = 0.1)+
  geom_jitter(width = 0.05, alpha = 0.5)+
  stat_summary(fun.y = "mean", geom = "point", size = 2, colour = "red2")+
  theme_classic()+
  ylab("Analyte concentration")+
  xlab("Day")
```

```r
ggplot(df_2, aes(x = day, y = value))+
  stat_summary(fun.data = mean_se, geom = "errorbar", width = 0.1)+
  geom_jitter(width = 0.05, alpha = 0.5)+
  stat_summary(fun.y = "mean", geom = "point", size = 2, colour = "red2")+
  theme_classic()+
  ylab("Analyte concentration")+
  xlab("Day")
```

---

### Perform variance component analysis for data set 1:

These data are presented in cells X7:Z22.

```r
model_1 <- anovaVCA(value ~ day, df_1)

# total = intra-laboratory
# day = intermediate
# error = repeatability
# DF = degrees of freedom
# VC = variance
# SD = standard deviation

model_1
```

```
##
##
## Result Variance Component Analysis:
## -----------------------------------
##
##   Name  DF        SS       MS       VC       %Total    SD       CV[%]
## 1 total 19.028328                   0.213762 100       0.462344 4.702361
## 2 day   4         1.526951 0.381738 0.041994 19.645268 0.204924 2.084227
## 3 error 20        3.435352 0.171768 0.171768 80.354732 0.414449 4.215234
##
```

```
## Mean: 9.83216 (N = 25)
##
## Experimental Design: balanced  |  Method: ANOVA
```

All variance components match those presented in the spreadsheet.

---

### Perform Chi-squared test against claimed imprecision values:

These data are presented in cells X27:Z38.

```
model_1_tests <- VCAinference(model_1
                              ,alpha = 0.05
                              ,total.claim = 7.0
                              ,claim.type = "CV"
                              ,error.claim = 3.3)$ChiSqTest
model_1_tests
```

```
##        Name Claim ChiSq value Pr (>ChiSq)
## total total   7.0    8.586901  0.02001374
## day     day    NA          NA          NA
## error error   3.3   32.632135  0.96298894
```

Based on these results, in agreement with the spreadsheet, the precision would be deemed “Acceptable” for both intra-laboratory and repeatability.

---

### Perform variance component analysis for data set 2:

```
## Perform variance component analysis for data set 2
model_2 <- anovaVCA(value ~ day, df_2)
model_2
```

```
##
##
## Result Variance Component Analysis:
## -----------------------------------
##
##   Name  DF        SS        MS       VC       %Total SD      CV[%]
## 1 total 23.809524                    1.559951 100    1.24898 2.841848
## 2 day   4         2.988183  0.747046 0*       0*     0*      0*
## 3 error 20        31.199013 1.559951 1.559951 100    1.24898 2.841848
##
## Mean: 43.94956 (N = 25)
##
## Experimental Design: balanced  |  Method: ANOVA | * VC set to 0 | adapted MS used for tot
al DF
```

All variance components match those presented in the spreadsheet.

---

### Perform Chi-squared test against claimed imprecision:

```
model_2_tests <- VCAinference(model_2
                              ,alpha = 0.05
                              ,total.claim = 3.4
                              ,claim.type = "CV"
                              ,error.claim = 2.5)$ChiSqTest
model_2_tests
```

```
##         Name Claim ChiSq value Pr (>ChiSq)
## total total   3.4    16.63392   0.1427094
## day     day    NA          NA          NA
## error error   2.5    25.84353   0.8289579
```

Based on these results, in agreement with the spreadsheet, the precision would be deemed “Acceptable” for both intra-laboratory and repeatability.

---

### Perform Grubb’s tests for detecting outliers:

These data are presented in cells X39:Z41.

```
grubbs.test(df_1$value)
```

```
##
##  Grubbs test for one outlier
##
## data:  df_1$value
## G = 1.87410, U = 0.84756, p-value = 0.6707
## alternative hypothesis: lowest value 8.98 is an outlier
```

```
grubbs.test(df_2$value)
```

```
##
##  Grubbs test for one outlier
##
## data:  df_2$value
## G = 3.30920, U = 0.52471, p-value = 0.001723
## alternative hypothesis: lowest value 40 is an outlier
```

Based on these results, a significant outlier was detected in data set 2 and thus the data should be checked, in agreement with the spreadsheet.

---

### Check calculations of bias, uncertainty of bias, $t$ critical value, and Z-scores:

These data are presented in cells AB7:AD18.

```
# Target values
```

```r
target_1 <- 10
target_uncertainty_1 <- 0.3 # (CV = 3.0% of 10)
target_2 <- 40
target_uncertainty_2 <- 1.0 # (CV = 2.5% of 40)

# T-score (k)
k <- round(qt(1 - (0.05/2), df = 25 - 1), 3)
cat("k =", k)
```

```
## k = 2.064
```

```r
# Bias, uncertainty of bias, and z-score 1
mean_bias_1 <- mean(df_1$value - target_1)
se_bias_1 <- sd(target_1 - df_1$value) / sqrt(25)

# Bias, uncertainty of bias, and z-score 2
mean_bias_2 <- mean(df_2$value - target_2)
se_bias_2 <- sd(target_2 - df_2$value) / sqrt(25)
```

Z-scores were computed using the following formula, where $\bar{x}$ represents the mean of the *measured* values; $\mu$ represents the *target* value; and $\sigma$ represents the *target* standard deviation.

$$Z=(\bar{x}-\mu)/\sigma$$

```r
## Calculation of Z-scores
z_score_1 <- mean_bias_1 / target_uncertainty_1
z_score_2 <- mean_bias_2 / target_uncertainty_2

# Present results
data.frame(Level = c(1, 2)
           ,Mean_bias = c(round(mean_bias_1, 2), round(mean_bias_2, 2))
           ,SE_bias = c(round(se_bias_1, 2), round(se_bias_2, 2))
           ,Z_score = c(round(z_score_1, 2), round(z_score_2, 2)))
```

```
##   Level Mean_bias SE_bias Z_score
## 1     1     -0.17    0.09   -0.56
## 2     2      3.95    0.24    3.95
```

Based on these results, the bias for level 1 would be “Acceptable” and would be “Rejected” for level 2. Note that the spreadsheet uses the following formula for Z-score calculation, where $\bar{x}$ represents the mean of the *measured* values; $\mu$ represents the *target* value; and $s$ represents the standard *error* of the *measured* data.

$$Z=(\bar{x}-\mu)/s$$

# Conclusions:

1. Calculations of of mean bias, SE, SD, and CVs matched thoses in the spreadsheet
2. Variance component analyses produced identical results to those calculated in the spreadsheet
3. Chi-squared tests produced identical results to those in the spreadsheet
4. Some discrepancies existed in the calculation of the Z-scores and biases due to differences in the formula used

- Calculated data set 1 Z-score = -0.56; spreadsheet Z-score = -1.84
- Calculated data set 2 Z-score = 3.95; spreadsheet Z-score = 3.97