

Statistics and electronics - lecture 1

The ASTA team

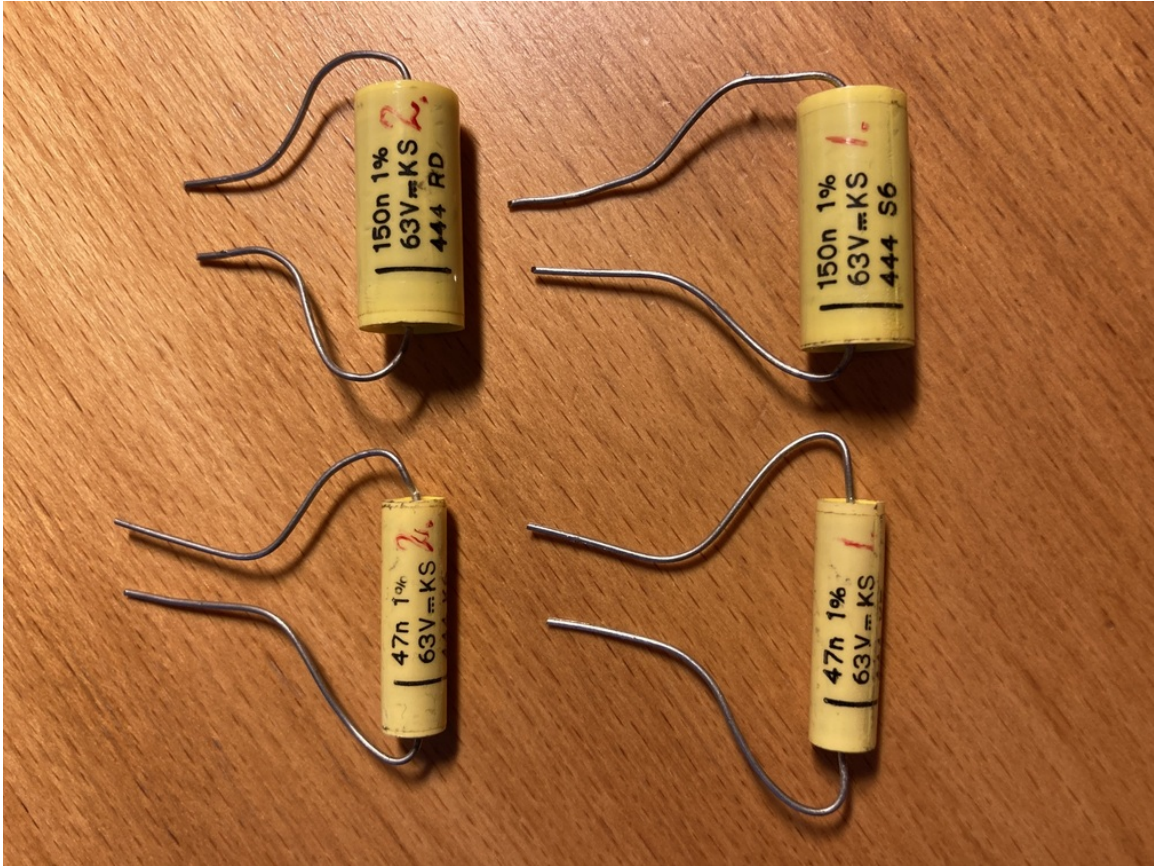
Contents

0.1	Sources of variation	1
0.2	Data from Peter Koch	2
0.3	Transformation	2
0.4	Transformation	3
0.5	Transformed data	4
0.6	Model considerations	4
0.7	Statistical model	4
0.8	Assumptions	4
0.9	Estimation of systematic error	5
0.10	Estimation of random error	5
0.11	Fit	5
0.12	Solution	5
0.13	Summing up	6
0.14	Test of no random effect	6
0.15	Lognormal variation	6
0.16	Moments of lognormal	7
0.17	CV of Lognormal	7
0.18	Linear calibration	8
0.19	Linear calibration fit	8
0.20	Calibrated values	8
0.21	Calibrated data	9

0.1 Sources of variation

We shall study 2 types of variation

- measurement variation due to random errors on a measuring device
- component variation due to random errors in the production proces



0.2 Data from Peter Koch

Peter has done 100 independent measurements of the capacity of 4 of the displayed capacitors and one additional. Nominal values are 47, 47, 100, 150, 150. All with stated tolerance of 1%.

```
load(url("https://asta.math.aau.dk/datasets?file=cap_1pct.RData"))
head(capDat, 4)
```

```
## capacity nomval sample
## 1 45.69 47 s_1_nF47
## 2 45.71 47 s_1_nF47
## 3 45.69 47 s_1_nF47
## 4 45.71 47 s_1_nF47
```

Here we see the first 4 capacity measurements of the first capacitor with nominal value 47.

- Remark: The measured values are consistently below the nominal value minus the 1% tolerance: $47 - 0.47 = 46.53$.

```
table(capDat$sample)
```

```
##
## s_1_nF47 s_2_nF47 s_3_nF100 s_4_nF150 s_5_nF150
## 100 100 100 100 100
```

0.3 Transformation

Linearisation:

$$f(x) \approx f(x_0) + f'(x_0)(x - x_0) \quad (1)$$

(2)

$$x_0 = 1 \quad (3)$$

$$f(x) = \log x \quad (4)$$

$$f'(x) = 1/x \quad (5)$$

(6)

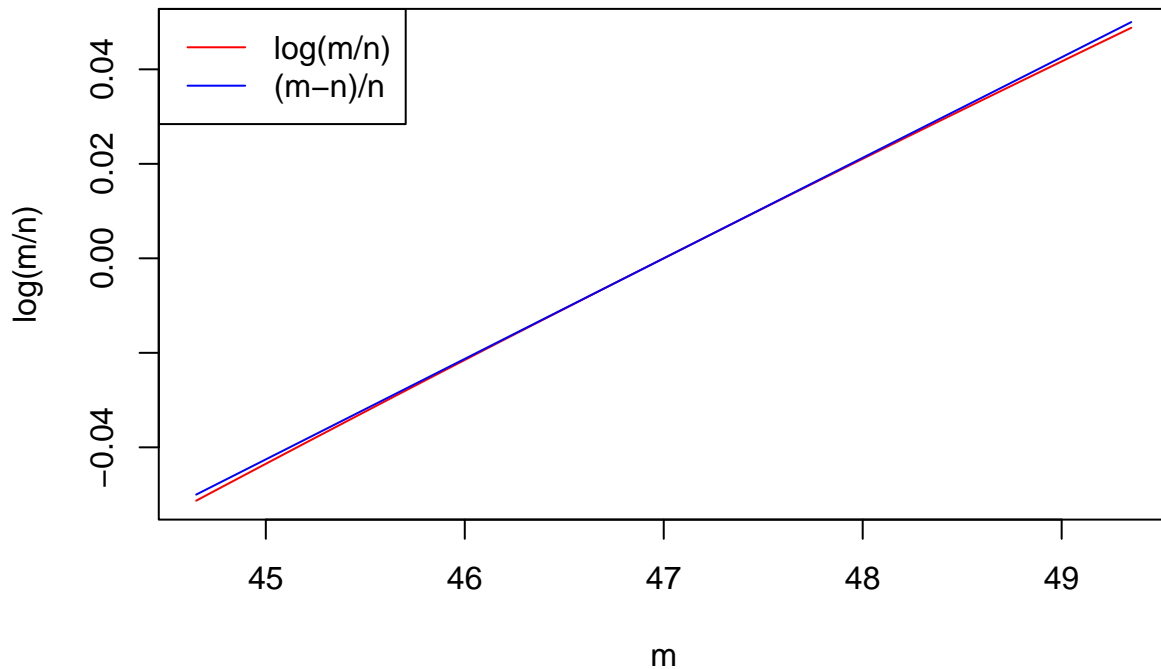
$$x = m/n \quad (7)$$

$$\log\left(\frac{m}{n}\right) \approx \log 1 + \frac{1}{1}\left(\frac{m}{n} - 1\right) \quad (8)$$

$$= \frac{m - n}{n} \quad (9)$$

$$\log\left(\frac{m}{n}\right) \approx \frac{m - n}{n}$$

```
n <- 47
m <- seq(47-5*0.01*47, 47+5*0.01*47, length.out = 100)
plot(m, log(m/n), col = "red", type = "l")
lines(m, (m - n)/n, col = "blue", type = "l")
legend("topleft", legend = c("log(m/n)", "(m-n)/n"), lty = 1, col = c("red", "blue"))
```



0.4 Transformation

$$\log\left(\frac{m}{n}\right) \approx \frac{m - n}{n}$$

Instead of the raw measurement we will consider:

$$\ln\text{Error} = \ln(\text{measuredValue}/\text{nominalValue})$$

Remark that by linear approximation:

$$\ln \text{Error} \approx \text{measuredValue} / \text{nominalValue} - 1 = (\text{measuredValue} - \text{nominalValue}) / \text{nominalValue}$$

which is the error relative to the nominal value.

I.e.: $\ln \text{Error}$ can be interpreted as relative error.

0.5 Transformed data

```
capDat = within(capDat, lnError <- log(capacity/nomval))
head(capDat, 2)
```

```
##   capacity nomval   sample   lnError
## 1    45.69     47 s_1_nF47 -0.02826815
## 2    45.71     47 s_1_nF47 -0.02783051
```

```
tail(capDat, 2)
```

```
##   capacity nomval   sample   lnError
## 499    145.7    150 s_5_nF150 -0.02908558
## 500    145.6    150 s_5_nF150 -0.02977216
```

- The resolution on Peters capacitance meter is with 2/1 decimal(s) in the 47/150 nF range, which means that only a limited number of different values(3-8) are observed. Meaning that box- or histogram-plots are noninformative.

0.6 Model considerations

- The measurements are more than 2.7% below the nominal value. This must be due to a systematic error on the meter.

In this case we have as earlier mentioned two further sources of error:

- $\ln(\text{measuredValue} / \text{nominalValue}) = \text{systematicError} + \text{productionError} + \text{measurementError}$

0.7 Statistical model

$$\ln(\text{measuredValue} / \text{nominalValue}) = \text{systematicError} + \text{productionError} + \text{measurementError}$$

We formulate the model:

- $Y_{ij} = \mu + A_i + \varepsilon_{ij}$

where

- Y_{ij} is the log error measurement
- μ is the systematic error on the meter
- A_i is the random production error
- ε_{ij} is the random measurement error
- $i = 1, 2, 3, 4, k = 5$ is the number of the 5 samples
- $j = 1, \dots, n = 100$ is the number of the observation in each sample

0.8 Assumptions

This is the model treated in WMM chapter 13.11, where it is assumed that

- A_i is normally distributed with mean 0 and variance σ_α^2 , which represents the production error
- ε_{ij} is normally distributed with mean 0 and variance σ^2 , which represents the measurement error

0.9 Estimation of systematic error

The systematic error is simply estimated by the mean

- $\hat{\mu} = \bar{y}_{..}$

```
muhat <- mean(capDat$lnError)
muhat
```

```
## [1] -0.0288375
```

The meter systematically reports a value, which is estimated to be 2.88% too low.

0.10 Estimation of random error

Notation from WMM chapter 13.3:

- $SSA = n \sum_i (\bar{y}_i - \bar{y}_{..})^2$ (related to production error)
- $SSE = \sum_{ij} (y_{ij} - \bar{y}_i)^2$ (related to measurement error)

Theorem 13.4 states:

- $E(SSA) = (k-1)\sigma^2 + n(k-1)\sigma_\alpha^2$
- $E(SSE) = k(n-1)\sigma^2$

0.11 Fit

```
fit <- lm(lnError ~ sample, data = capDat)
anova(fit)
```

```
## Analysis of Variance Table
##
## Response: lnError
##           Df    Sum Sq   Mean Sq F value    Pr(>F)
## sample      4 0.0046576 0.00116440  4067.4 < 2.2e-16 ***
## Residuals 495 0.0001417 0.00000029
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

where we read

```
SS <- anova(fit)$`Sum Sq`
SSA <- SS[1]
SSE <- SS[2]
```

- $SSA = 0.00466$ and $SSE = 0.000142$

0.12 Solution

Solving the equations

- $SSA = E(SSA)$ and $SSE = E(SSE)$

yields

- $\hat{\sigma}_\alpha^2 = \frac{1}{n} \left(\frac{SSA}{k-1} - \hat{\sigma}^2 \right) = 11.64 \times 10^{-6}$
- $\hat{\sigma}^2 = \frac{SSE}{k(n-1)} = 0.29 \times 10^{-6}$

0.13 Summing up

- the meter has an estimated systematic error of -2.88%
- the estimated standard error of the meter is $\sqrt{0.29 \times 10^{-6}} = 0.054\%$
- the estimated standard error of the production is $\sqrt{11.64 \times 10^{-6}} = 0.34\%$. So the 3-sigma limit is 1.02% , which is in accordance with the tolerance of 1% . It should be noted that the estimate is insecure, as it is based on 4 degrees of freedom only.

The estimated variance on log error

- $0.29 \times 10^{-6} + 11.64 \times 10^{-6} = 11.93 \times 10^{-6}$

is clearly dominated by the production error.

0.14 Test of no random effect

We have the possibility of testing the hypothesis

- $H_0: \sigma_\alpha = 0$

This is equivalent to

- $E(SSA/(k-1)) = E(SSE/k/(n-1)) = \sigma^2$

Under H_0 the statistic

- $F = \frac{\frac{SSA}{k-1}}{\frac{SSE}{k(n-1)}}$

has an F-distribution with degrees of freedom $(k-1, k(n-1))$

In the actual case $f_{obs} = 4067.4$, which is highly significant (p-value=0).

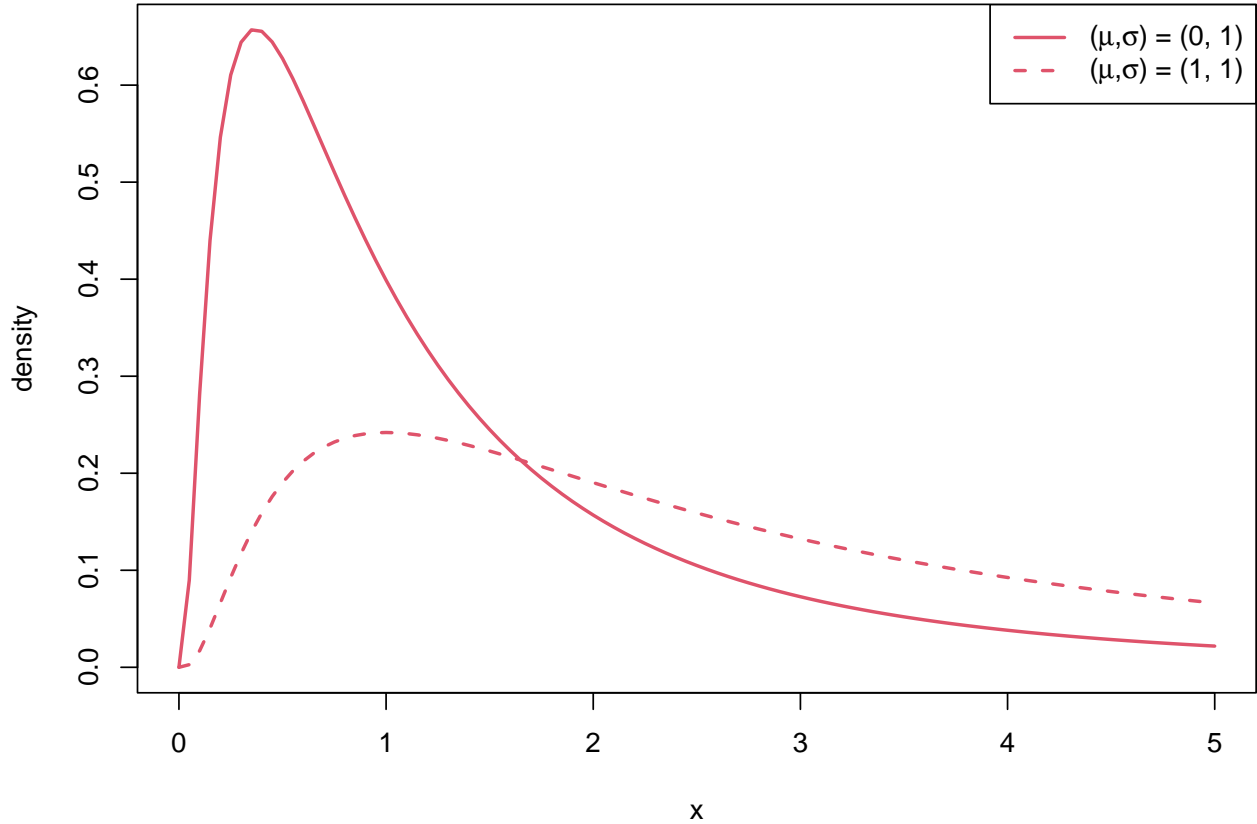
0.15 Lognormal variation

In the preceding we assumed normal errors after a log transformation.

Let X be a random variable and $Y = \ln(X)$.

We say that X has a lognormal distribution if Y has a normal distribution with - say - mean μ and standard deviation σ .

Density plots:



0.16 Moments of lognormal

If $Y = \ln(X)$ has a normal distribution with mean μ and standard deviation σ , then Theorem 6.7 of WMM states:

- $E(X) = \exp(\mu + \sigma^2/2)$
- $Var(X) = \exp(2\mu + \sigma^2)(\exp(\sigma^2) - 1)$

If we are interested in relative variation, it is common to look at the coefficient of variation

- $CV(X) = \frac{\sigma}{\mu}$

if e.g. $CV=0.05$ then 95% of our measurements are within

- $\mu \pm 2\sigma = \mu \pm 2 * 0.05\mu = \mu(1 \pm 0.1)$

i.e. most observations are within 10% of the mean.

0.17 CV of Lognormal

If $Y = \ln(X)$ has a normal distribution with mean μ and standard deviation σ , we calculate CV for X as

$$\bullet CV(X) = \frac{E(X)}{\sqrt{Var(X)}} = \sqrt{\exp(\sigma^2) - 1}$$

In Peter's data we estimated the variance of the log error to 11.64×10^{-6} , which means that the estimated CV of the capacity measurement is

- $CV = \sqrt{\exp(11.64 \times 10^{-6}) - 1} = 0.34\%$.

i.e., if we correct for the systematic error of the meter, then our measurements are extremely precise.

0.18 Linear calibration

In our previous analysis, we assumed, that the systematic error on the meter did not depend on nominal value.

To check this assumption consider the model

- $Y = \ln(\text{measuredValue})$ is a linear model of $x = \ln(\text{nominalValue})$
- $Y = \alpha + \beta x + \varepsilon$

where we have previously assumed slope(β) equal to 1.

0.19 Linear calibration fit

```
fit <- lm(log(capacity) ~ log(nomval), data = capDat)
summary(fit)

##
## Call:
## lm(formula = log(capacity) ~ log(nomval), data = capDat)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.0064121 -0.0010784  0.0007315  0.0013879  0.0050839
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.0300145  0.0011907  -25.21  <2e-16 ***
## log(nomval)  1.0002636  0.0002648 3776.74  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.003101 on 498 degrees of freedom
## Multiple R-squared:  1, Adjusted R-squared:  1
## F-statistic: 1.426e+07 on 1 and 498 DF, p-value: < 2.2e-16
```

The slope is more than close to 1. But is actually extremely significantly different from 1 (tvalue=3776.74 »» 3).

Clearly, it is a bit dubious to assume a linear relationship, as we only have 3 nominal values.

0.20 Calibrated values

If we stick to the linear calibration model, it is sensible to correct our measured errors according to the calibration of the meter:

- $\text{measuredError} = \alpha + \beta * \text{correctError}$

- $\text{correctError} = (\text{measuredError} - \alpha) / \beta$

```
ab = coef(fit)
ab
```

```
## (Intercept) log(nomval)
## -0.03001454  1.00026359
```



```
capDat$lnError_c = (capDat$lnError - ab[1])/ab[2]
```

0.21 Calibrated data

```
head(capDat)
```

```
##   capacity nomval  sample    lnError  lnError_c
## 1    45.69     47 s_1_nF47 -0.02826815 0.001745930
## 2    45.71     47 s_1_nF47 -0.02783051 0.002183452
## 3    45.69     47 s_1_nF47 -0.02826815 0.001745930
## 4    45.71     47 s_1_nF47 -0.02783051 0.002183452
## 5    45.70     47 s_1_nF47 -0.02804930 0.001964715
## 6    45.69     47 s_1_nF47 -0.02826815 0.001745930
```

The calibrated data now shows that the production error on component s_1_nF47 is in the vicinity of 0.2%. Well below the tolerance 1%.