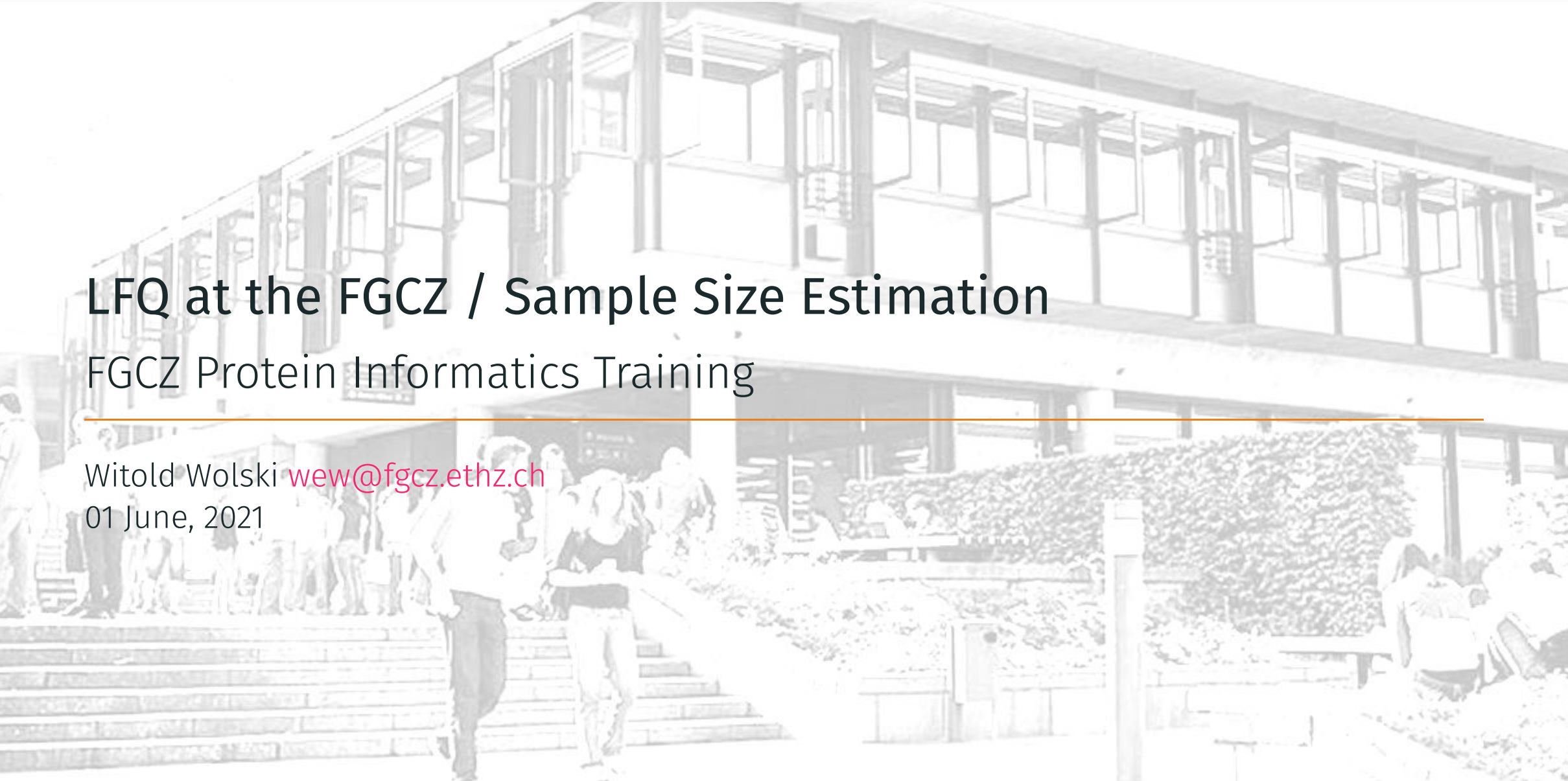


LFQ at the FGCZ / Sample Size Estimation

FGCZ Protein Informatics Training

Witold Wolski wew@fgcz.ethz.ch

01 June, 2021

Sample size estimation

Overview

- Sample size estimation
  - the `power.t.test` function
  - error of sample size estimation
  - reporting power
  - data preprocessing
- LFQ analysis at the FGCZ
  - Kickoff meeting
  - QC experiment
  - Main experiment

Types of error when testing hypothesis

A **type I error** (false positive) occurs when
the null hypothesis (H0) is true, but is rejected.
The *type I error rate* or **significance level** (p-Value)
is the probability of rejecting the
null hypothesis given that it is true.

A **type II error** (false negative) occurs when
the null hypothesis is false,
but erroneously fails to be rejected.
The *the type II error rate* is denoted by the Greek letter β
and is related to the **power of a test** (which equals $1 - \beta$).

For a given test, the only way to reduce both error rates
is to **increase the sample size**, and this may not be feasible.

		reality	
		H_0 = true	H_0 = false
conclusion	H_0 is not rejected	OK	type II error
	H_0 is rejected	type I error	OK

Two sample t-test

Test statistic

$$T = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

Null distribution

$$T|H_{null} \sim T(0, 1, df = N - 1)$$

where delta = $\bar{X}_2 - \bar{X}_1$, sd - standard deviation, df - depends on sample size

`power.t.test` - estimating power

Compute the power of the one- or two- sample t-test, or determine parameters to obtain a target power.

```
power.t.test(n = NULL, delta = NULL, sd = 1, sig.level = 0.05,
             power = NULL,
             type = c("two.sample", "one.sample", "paired"),
             alternative = c("two.sided", "one.sided"),
             strict = FALSE, tol = .Machine$double.eps^0.25)
```

Relates the type-2 error (power) to the other variables: type-1 error (sig.level), standard deviation (sd), effect size (delta), sample size (n)

`power.t.test` - estimating power

What is **the power** of the two-sample t-test using a significance level of 0.05 if we want to detect a biologically relevant difference of 50% given a standard deviation of 0.5 and group size of 10 samples?

```
power.t.test(delta = 0.59, n = 10, sd = 0.5, sig.level = 0.05)
```

```
##
##      Two-sample t test power calculation
##
##               n = 10
##           delta = 0.59
##              sd = 0.5
##       sig.level = 0.05
##           power = 0.7039889
##     alternative = two.sided
##
## NOTE: n is number in *each* group
```

Why `delta = 0.59`? `(50%)` difference, means FC of `(1.5)` or greater, and `(log2(1.5) = 0.59)`

`power.t.test` - esitmating difference d

What **biologically relevant difference** (or greater) can be detected with a two-sample t-test using a significance level of 0.05 and a power of 0.8 given a standard deviation of 0.5 and group size of 10?

```
res ← power.t.test(n = 10, sd = 0.5, power = 0.8, sig.level = 0.05)
res$delta
```

```
## [1] 0.6624728
```

```
2^res$delta
```

```
## [1] 1.582793
```

$$log_2(x) = y \quad \text{exactly if} \quad 2^y = x$$

$$2^d = I_1/I_2$$

$$log_2(I_1/I_2) = log2(I_1) - log2(I_2) = d$$

`power.t.test` - Sample size estimation

What **sample size** do we need, given a significance level of 0.05 if we want to detect a biologically relevant difference of 50% given a standard deviation of 0.5 and a power of 0.8?

```
res ← power.t.test(delta = 0.59, sd = 0.5, sig.level = 0.05, power = 0.8)
res$n
```

```
## [1] 12.31238
```

```
ceiling(res$n)
```

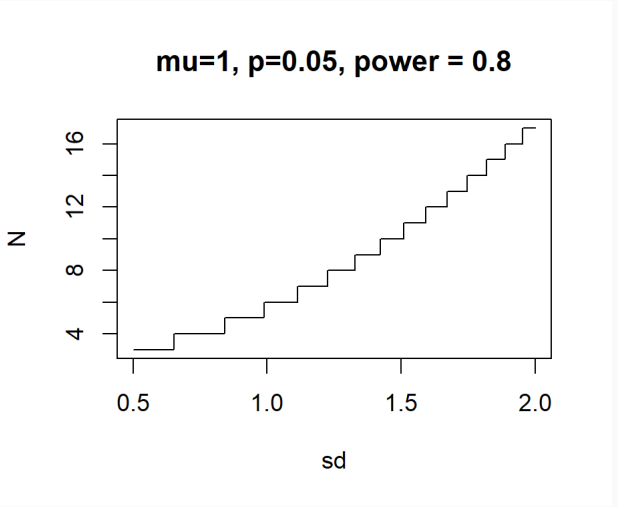
```
## [1] 13
```

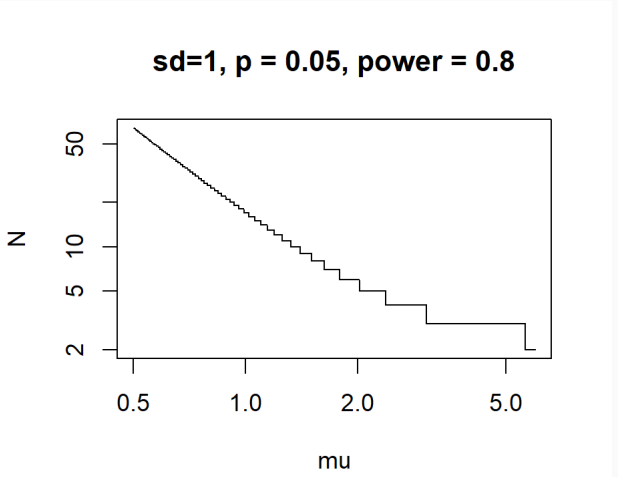
`n - number of observations (per group)`

`power.t.test` - summary

For each statistical test,
there exists a unique relation between:

- desired smallest detectable effect size μ
- sample variance σ^2
- sample size N
- critical p-value p_0
- statistical power

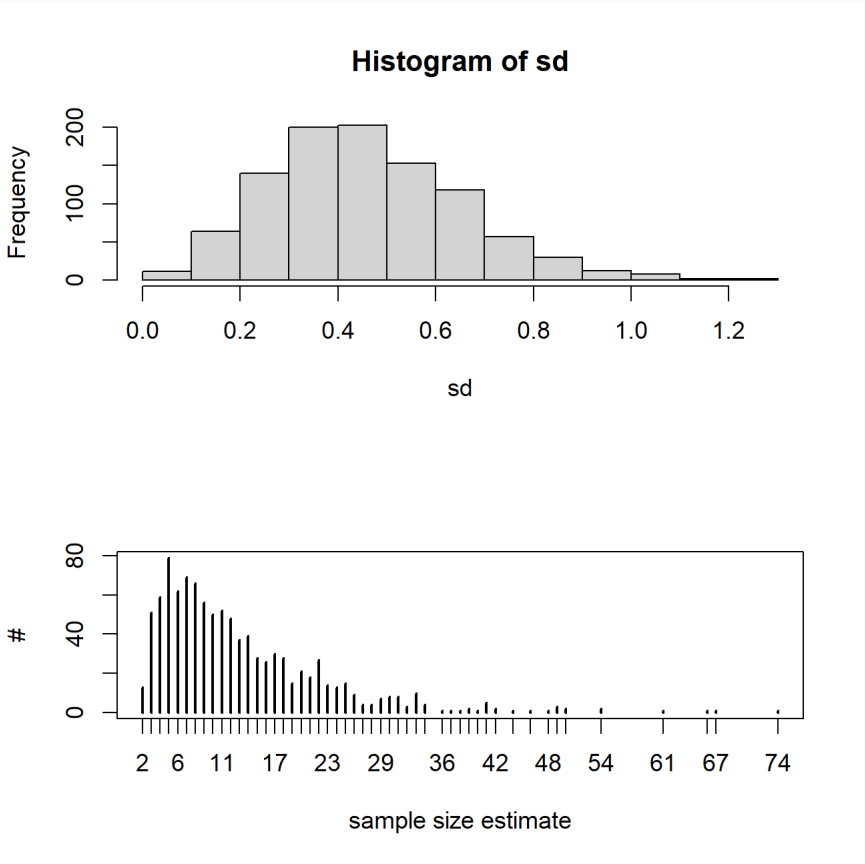




Top - greater standard deviation requires larger sample sizes, Bottom - smaller effect size requires larger sample sizes

Sample size estimation - error of

```r
sd ← sapply(1:1000,
            function(x){
              sd(rnorm(4,mean = 0, sd = 0.5))
              })
tmp ← sapply(sd,
             function(x){
               power.t.test(d = 0.59,
                            sd = x,
                            sig.level = 0.05,
                            power = 0.8)$n})
par(mfrow = c(2,1))
hist(sd)
plot(table(ceiling(tmp)),
     ylab="#", xlab="sample size estimate")
```



We simulate 1000 samples of size 4 from normal distribution and determine the standard deviation of each sample. For each standard deviation we computed the sample size using the `power.t.test` function.

Sample size estimation - error of

For a protein your uncertainty of the standard deviation and the required sample size is large when you measure 4 samples to estimate the sd.

In high throughput setting you measure **thousands** of proteins and determine their variances. Assuming that most of the proteins have a similar variance, the error of the median standard deviation will be small.

```
power.t.test(sd = median(sd), d = 0.59, sig.level = 0.05, power = 0.8)$n
```

```
## [1] 9.930872
```

Proteins will have **different** unknown underlying variances.

Calculating Observed Power

You run a test and you have the **estimates** of the sd, $p.value$ and fold change $delta$.
We could compute the power by taking the *estimates* obtained by the test.

```
power.t.test(delta = estimated_fc, n = nr_samples_per_group,
             sd = esitmated_sd,  sig.level = p.value)
```

Some proteomics software is doing it (e.g. Progenesis).
How good might this power **estimate** be?
Calculating Observed Power Is Just Transforming Noise

How you report the power of your test?
State the parameters of the sample size estimation for your experiment (which includes power).
You can use the standard deviation estimates from the main experiment to determine the power or sample size for a verification experiment.

Sample size estimation in practice.

- Data preprocessing
- Checking assumptions
- Sample size estimation

https://wolski.github.io/prolfqua/articles/QualityControlAndSampleSizeEstimation.html

Data preprocessing - Normalizing Intensities

The coefficient of variation is the slope of the line we when plotting sd against the μ of raw MS data.

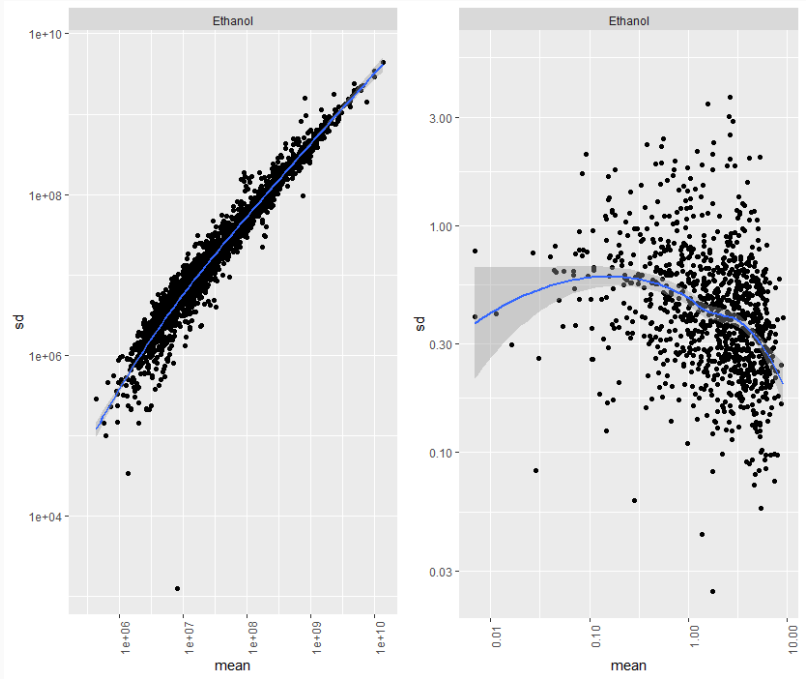
$$CV = \frac{sd(x)}{mean(x)}.$$

The sd is proportional to μ.

For testing we need iid data (identically distributed).

We can obtain constant (identical) standard deviation for all measurements by log_2 transforming the data.

Note:
$\frac{d}{dx}\ln(x) = \frac{1}{x}$.



Data preprocessing - Normalizing Intensities

To remove systematic differences among samples:

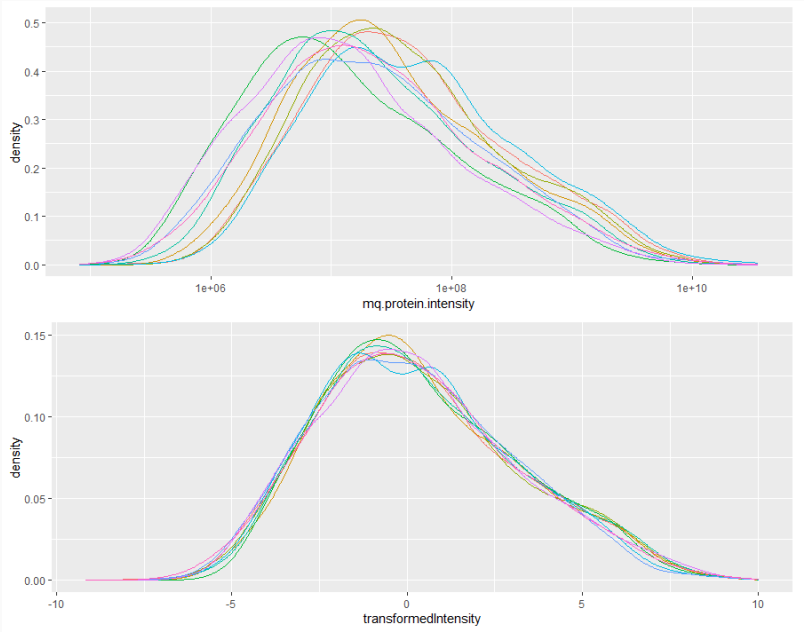
- apply robust z-transformation

$$I_i^n = \frac{I_i^t - med(I_i^t)}{mad(I_i^t)} \cdot \sum_{i=1}^{N} mad(I_i^t)/N$$

After z-transform samples have the same deviation

$$mad(I_1^t) = mad(I_2^t)$$

equal to the average **original variance** of all samples.



More methods fo data normalization are discussed here: Elo *et al.*

The LFQ workflow at the FGCZ - bioinformatics part

- **kickoff meeting** - bioinformatics part
  - Agenda
- **QC experiment**
  - to determine biological and biochemical variability:
    - nr of peptides and proteins ...
    - within group variability ...
  - Quality Control and *Sample size* estimation
- **Main experiment**
  - Data Analysis and *result delivery* including linear modelling and GSEA analysis.
  - meeting to discuss results

Kickoff meeting - bioinformatics

- Aim
  - Improve quality of services
  - Get all the information we need to perform analysis
  - Give *statistical* guidance
  - Ensure reproducibility
- Use of an Agenda
  - helps to prepare for the meeting
  - streamlines discussion
  - focuses on points relevant to the data analysis
  - Keep meeting short (15-30 minutes)
  - Ensure reproducibility

https://wolski.github.io/prolfqua/articles/Auxiliary_ExDesignSurvey.html

Kickoff meeting - bioinformatics

Document

- which protein database to use
- which and how many samples to use for the QC
- collect all the parameters for sample size estimation
- specify the design of the main experiment
  and the hypothesis to be tested
- document all the names of the factors and factor levels

https://wolski.github.io/prolfqua/articles/Auxiliary_ExDesignSurvey.html

Sample size estimation

- Hand in 4 **Bio-replicates** of the same condition (ideally of condition with highest variability).
- Estimate the variance of all measured protein.
- Compute sample sizes for main experiment given:
  - biologically relevant effect size
  - power of the test
  - observed variance
  - size of test

Sample size estimation report

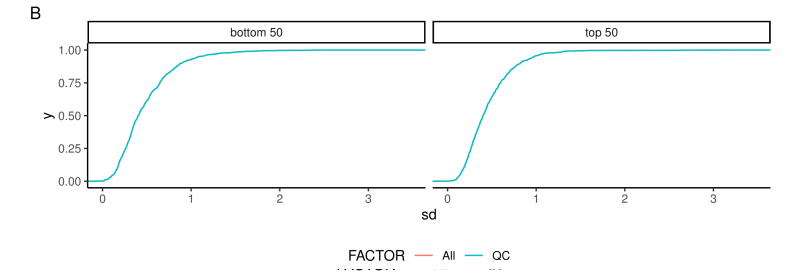
- check if there are no technical problems
- check if assumptions for data normalization are met.

Visualize the distribution of standard deviation
of all proteins using the density function and
the empirical cumulative density function (ecdf)

Sample size estimates for parameters
specified in protocol

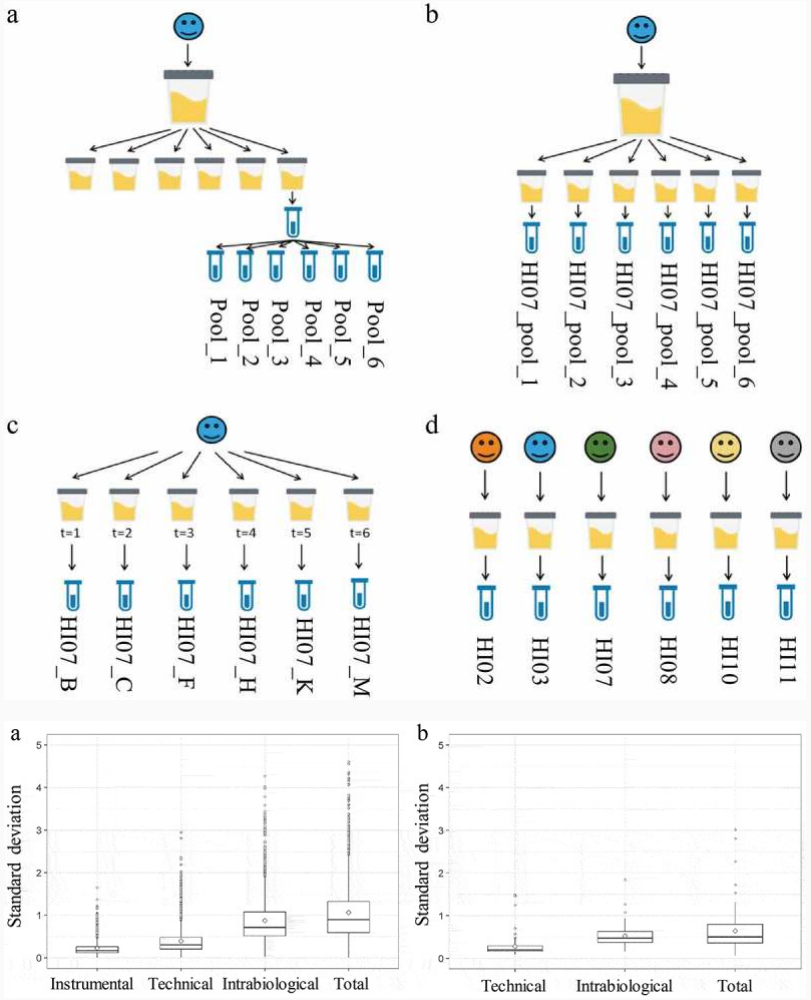
- *power* 0.8
- *significance level* 0.05
- *delta* $0.59, 1, 2$
- and standard deviation taken from QC experiment.

quantile	sd	FC=1.51	FC=2	FC=4
50%	0.40	9	4	3
60%	0.47	12	5	3
70%	0.57	16	7	3
80%	0.68	22	9	4
90%	0.85	34	13	5



Sources of variation

- (a) the instrumental variation
- (b) the total technical variation
- (c) the intrabiological variation
- (d) the interbiological variation



https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6807909/

Conclusion

- Works if **biological variability >>** *biochemical + technical variability*
- How can you know?
  - To understand sources of variance measure:
  - technical and biochemical and biological replicates and determine:
  - *technical CV*, *biochemical CV*, or *biological CV*.
- Sample size calculation is based on the standard deviation estimate.
  - $\boldsymbol{sd}$ estimate for single protein have a large error
  - small error for the median $\boldsymbol{sd}$ of all proteins
- Calculating Observed Power Is Just Transforming Noise
- Sample size calculation using `power.t.test` function ignores multiple testing problems.
- Observations need to be *independent and identically distributed* **iid**.