



Variance estimation of change for partially overlapping data

Presentation of a new R package

Daniel Hoop

8 October 2019



Motivation

Last year at the Pacioli workshop, Swetlana Renner presented methodological aspects of the package that is being presented here.

A short summary is given on the next slides.



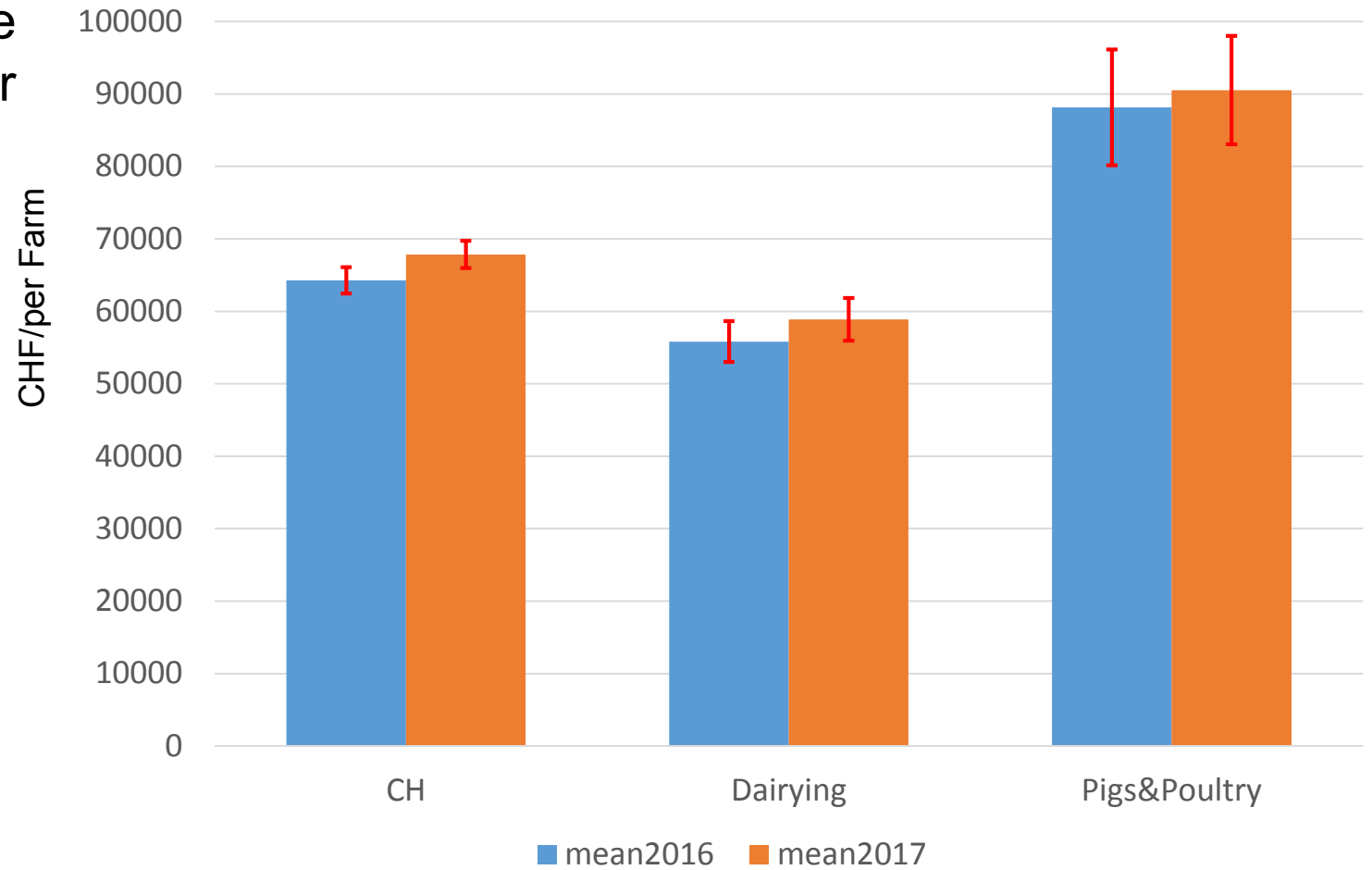
Motivation

- Monitoring changes (or trends) in agricultural performance indicators is of key importance for agricultural policy makers
- However, because FADN performance indicators are derived from a sample, not from the whole population, these estimators are not exact.
- Good news:
For random samples, the imprecision of an estimator can be estimated as well! 😊



Precision use case no. 1

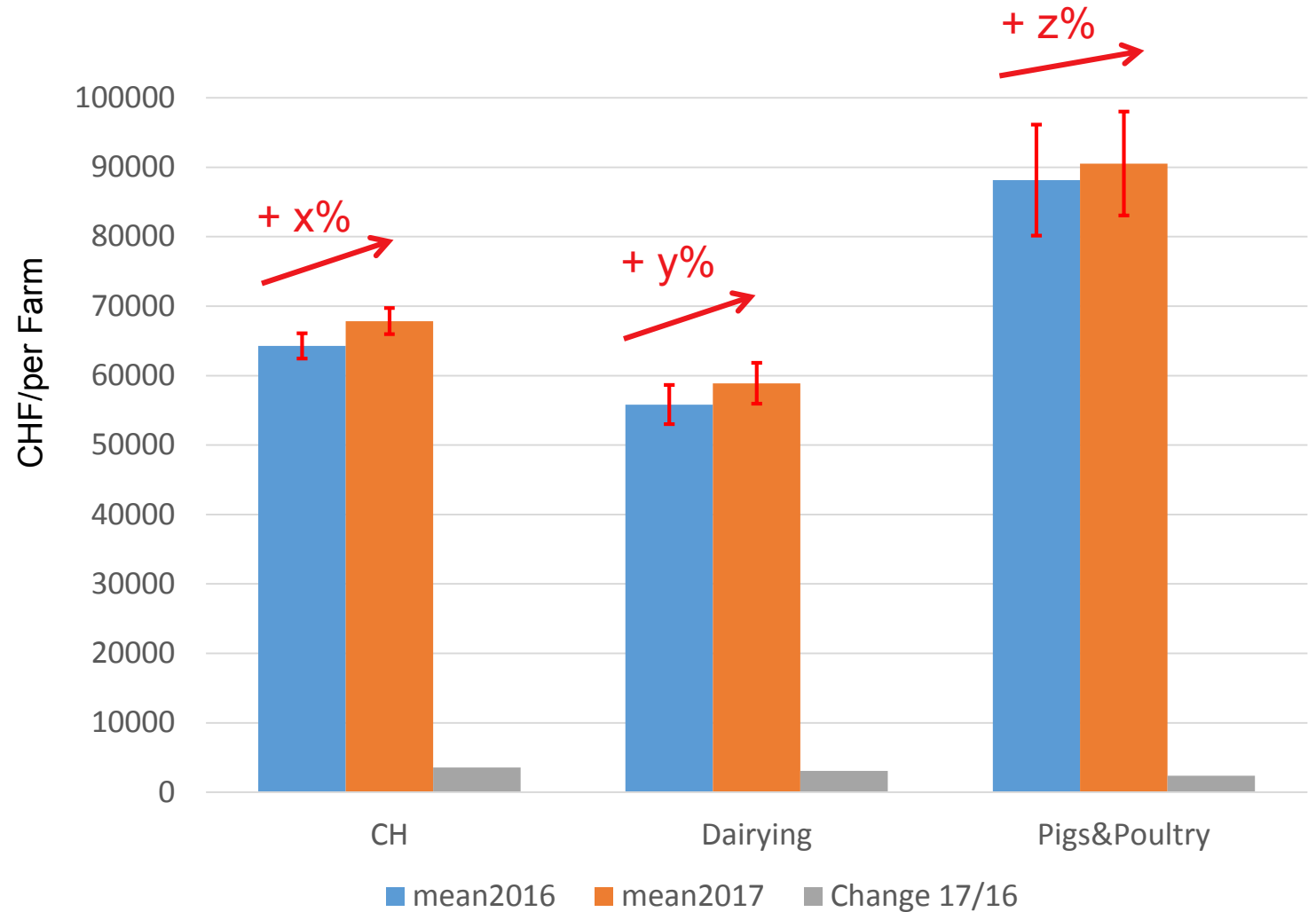
- What is the confidence interval of an estimator in a specific year?





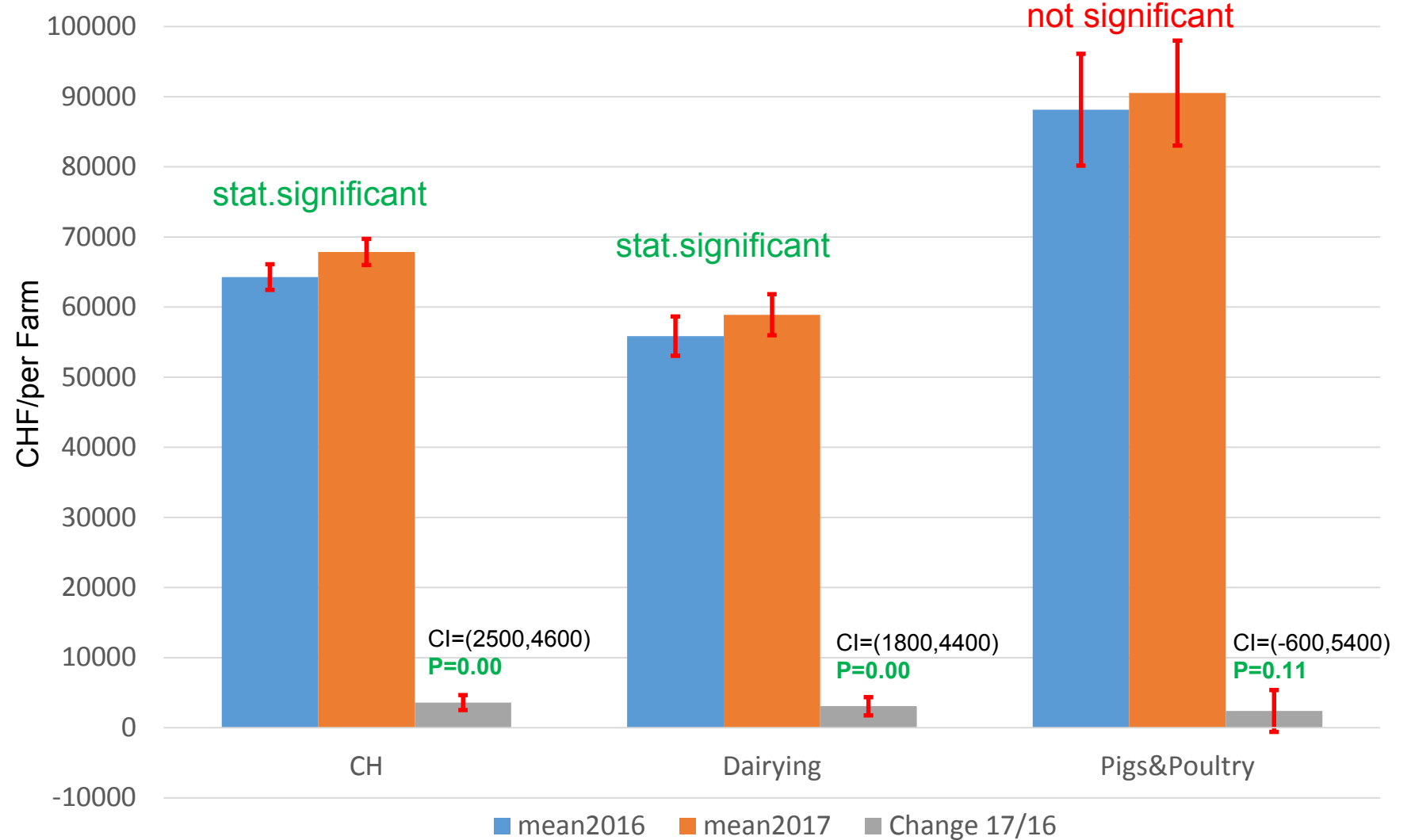
Precision use case no. 2

- Given the precision in each year, and the covariance between the two years, what is the probability that we observe change in the sample, even though there was no change in reality?





Imprecision use case no. 2





Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Federal Department of Economic Affairs,
Education and Research EAER

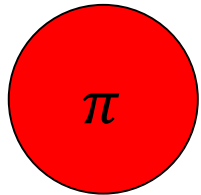
Agroscope

What's new in the package?

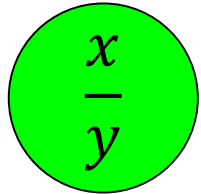


Difficulties in variance estimation

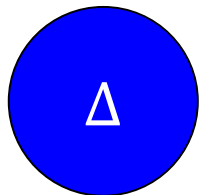
- The complexity of variance estimation increases if one of the following scenarios is given



Different inclusion probabilities (i.e. different weights) for different observations in the sample. This is the case for stratified samples.



Calculating ratios

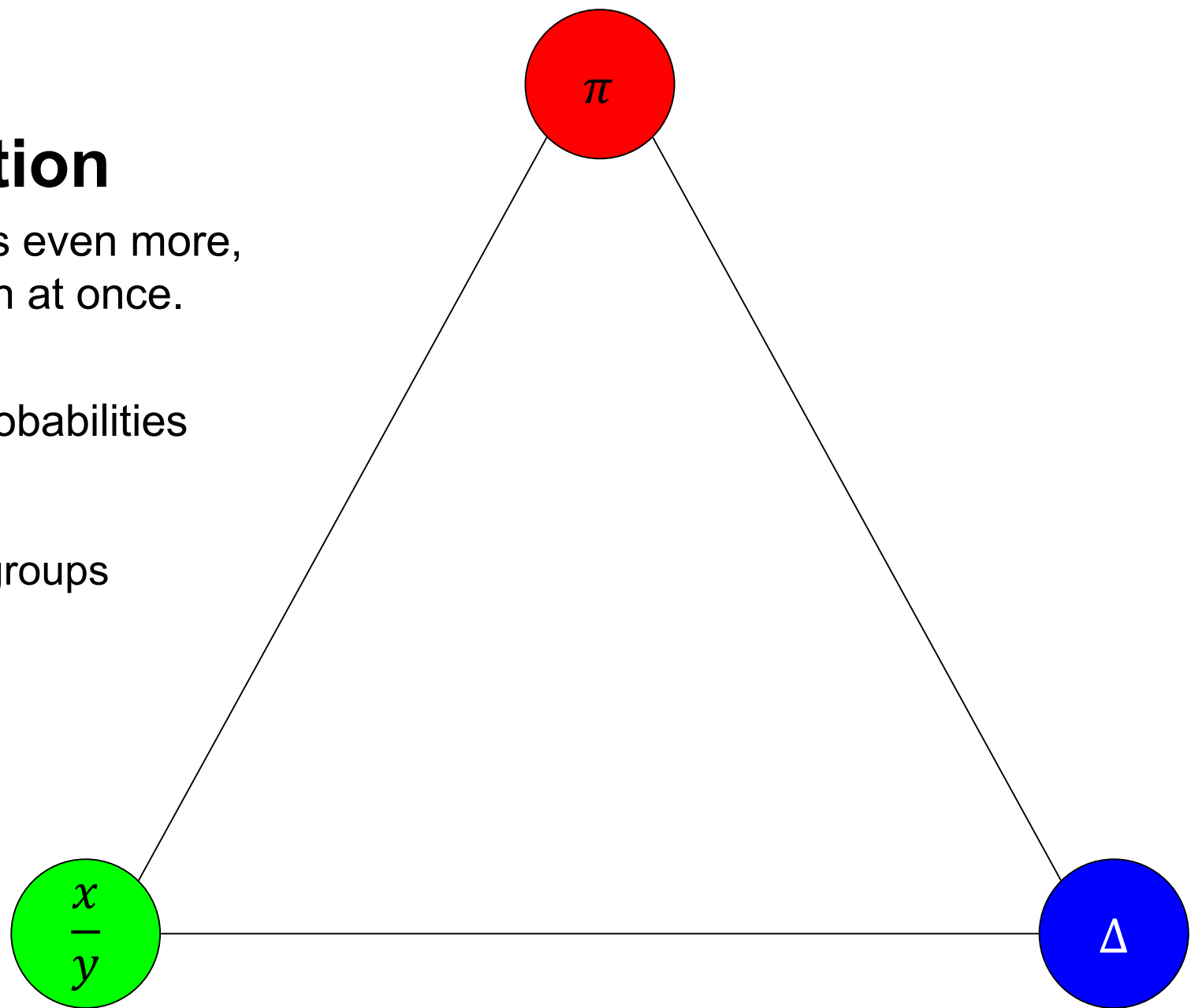


Calculating differences between groups, e.g. between year 1 and year 2.



Difficulties in variance estimation

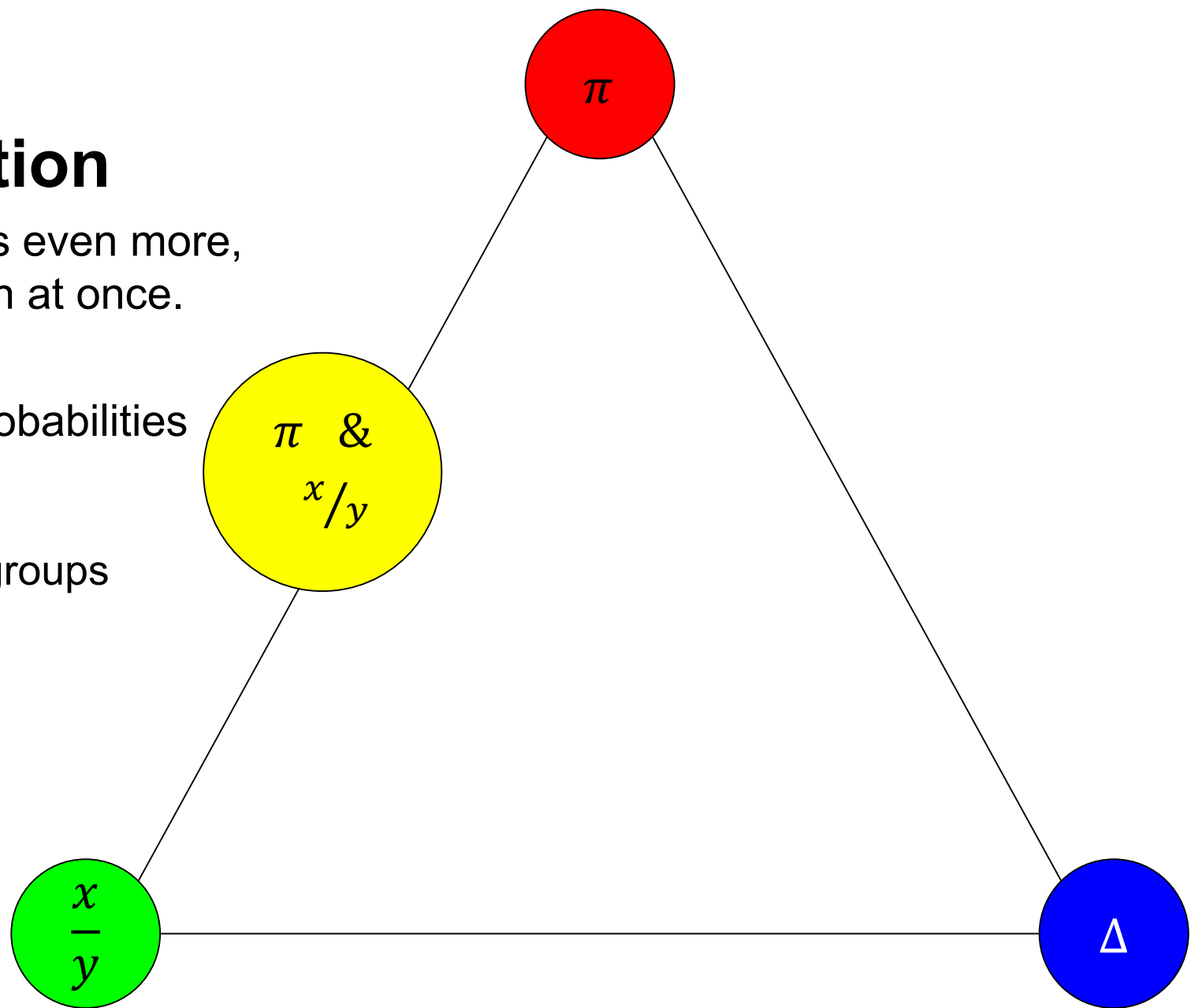
- The complexity increases even more, if these scenarios happen at once.
- π = Different inclusion probabilities
- $\frac{x}{y}$ = ratios
- Δ = Differences between groups





Difficulties in variance estimation

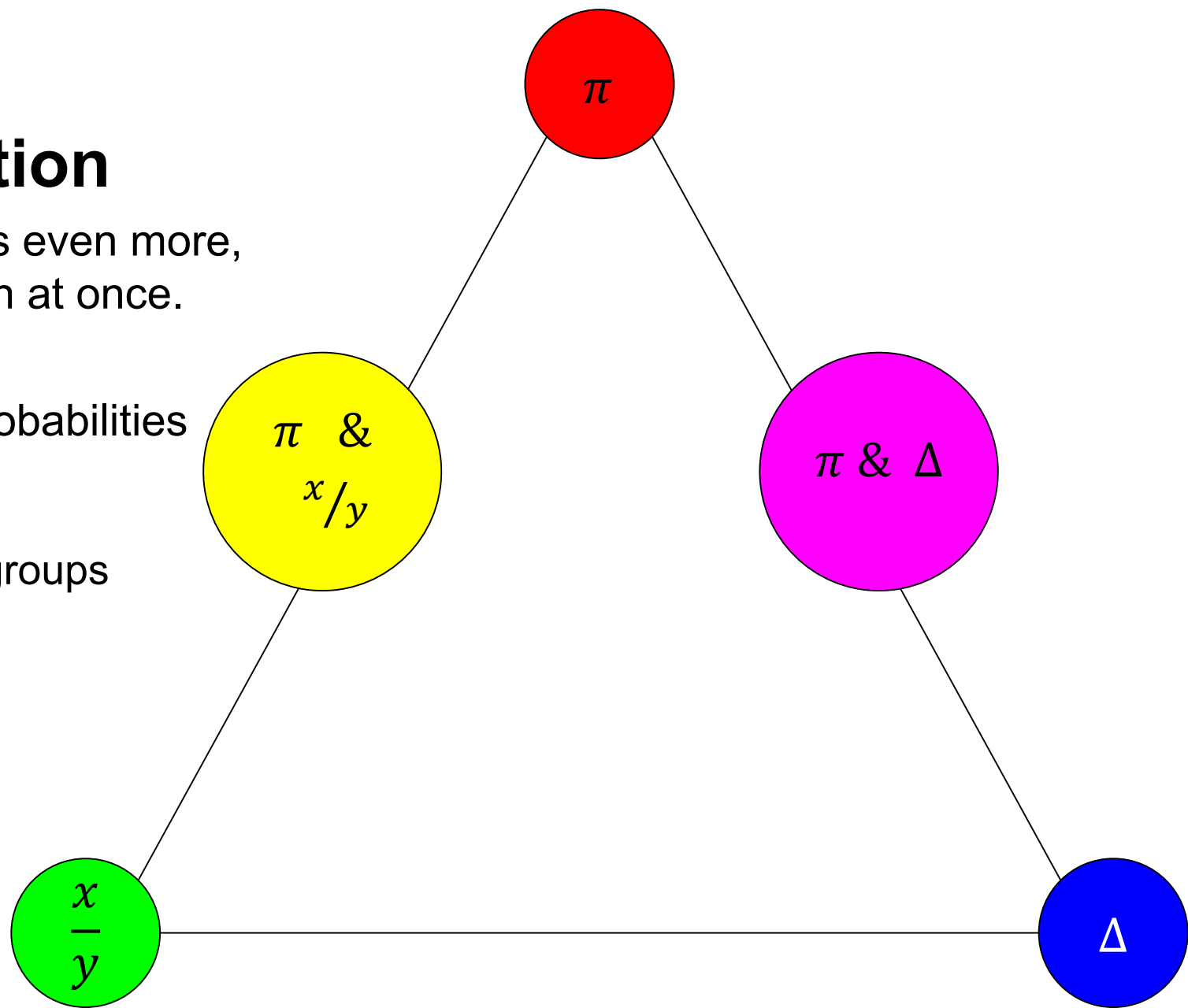
- The complexity increases even more, if these scenarios happen at once.
- π = Different inclusion probabilities
- $\frac{x}{y}$ = ratios
- Δ = Differences between groups





Difficulties in variance estimation

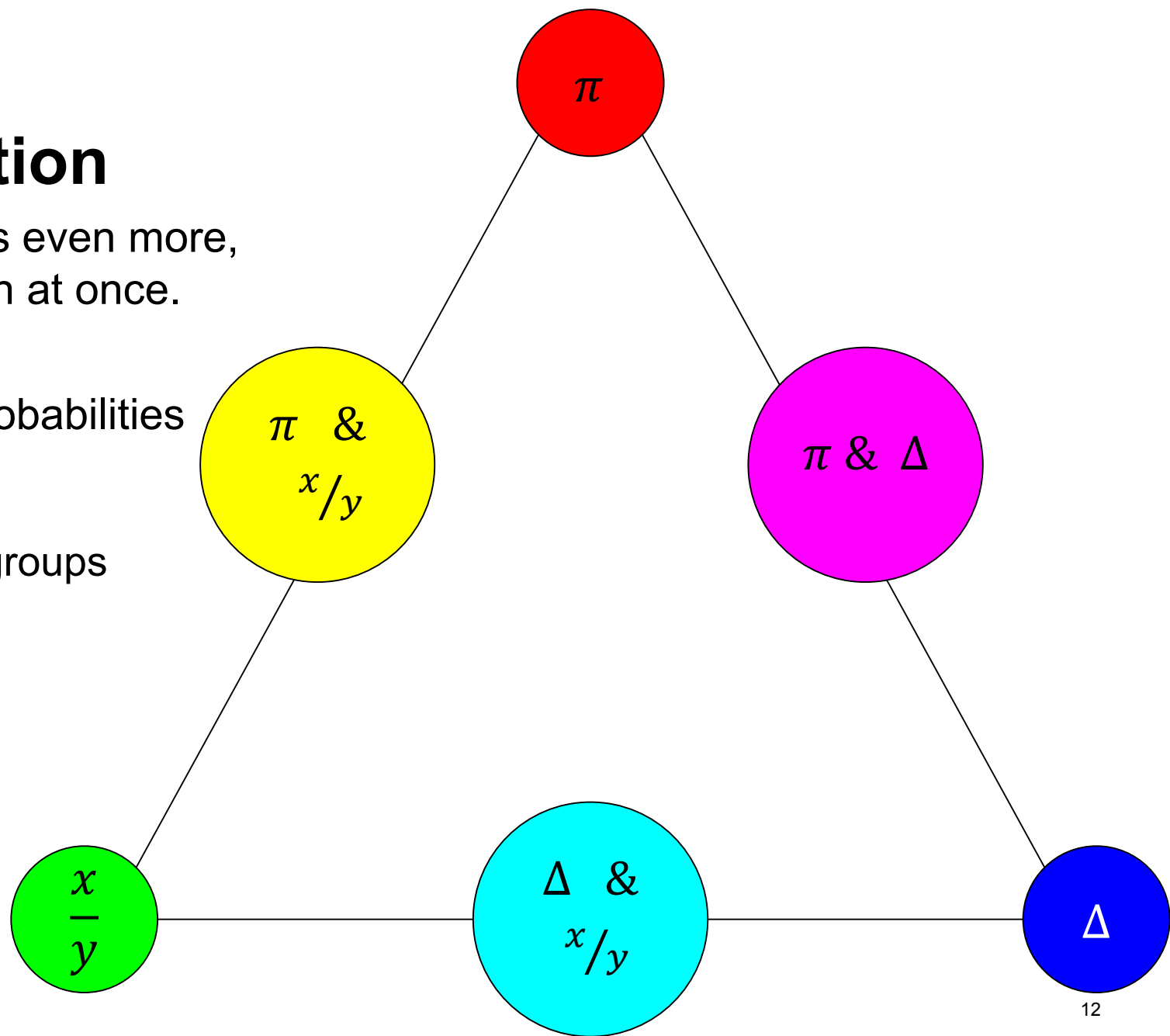
- The complexity increases even more, if these scenarios happen at once.
- π = Different inclusion probabilities
- $\frac{x}{y}$ = ratios
- Δ = Differences between groups





Difficulties in variance estimation

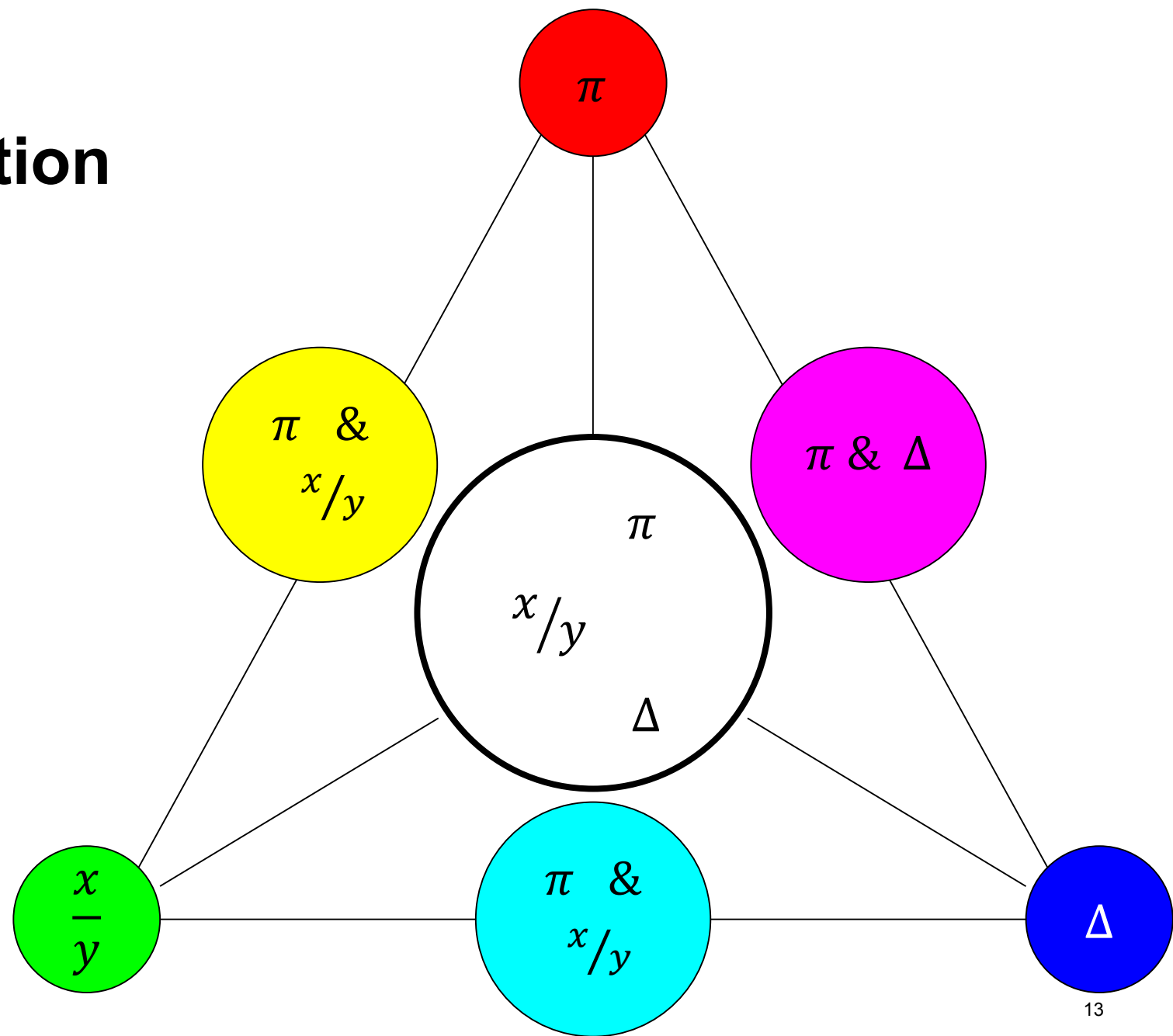
- The complexity increases even more, if these scenarios happen at once.
- π = Different inclusion probabilities
- $\frac{x}{y}$ = ratios
- Δ = Differences between groups





Difficulties in variance estimation

- *The packages does everything for you!*





What can be calculated for 1 year?

- Variance
- Standard error
- Margin of error = Half length of the confidence interval

What can be calculated for group differences?

- Variance, standard error, margin of error
- The probability value for the null hypothesis that the difference between the group is 0 (one sided test).

For all cases, it is possible to take advantage of the calibration method to reduce the imprecision of an estimator.



Installing the package

- The package is available only on a Agroscope internal git repository.
- So, you need to have a local copy and install it from there.

```
# Install the package
# If you don't have devtools installed, try the compiled version for windows
install.packages("path/to/your/file/VESP.zip", repos = NULL)

# Else try the source file
install.packages("path/to/your/file/VESP.tar.gz", repos=NULL, type="source")

# Load the package
library(VESP)
```



Tutorials on how to use the package

- The package includes two step-by-step tutorials on
 - Calibration of weights
 - Variance estimation

```
# Show a tutorial on calibration  
calibTutorial()
```

```
# Show a tutorial for variance estimation  
varEstTutorial()
```




Estimate the standard error (SE) of a figure in a given year

```
varianceEstimate(data = df,  
                 cols = c("X1", "X2"),  
                 weights = df[, "weights"],  
                 inclusProbs = df[, "inclusProbs"],  
                 figure="SE")
```

```
      X1      X2  
8.062173 16.369944
```

- Explanation of arguments
 - data = Your data.frame containing all data.
 - cols = The columns in 'data' for which you want to calculate the SE.
 - weights = The weights
 - inclusProbs = Inclusion probabilities (in most cases 1 / weights)
 - figure = The figure you want to calculate. "SE" = Standard Error



Estimate the SE of a ratio

- In argument 'cols', specify the ratio, similar to a ratio in an R formula

```
varianceEstimate(data = df,  
                 cols = c("x1", "x2", "x1_div_x2" "I(x1 / x2)"),  
                 weights = df[, "weights"],  
                 inclusProbs = df[, "inclusProbs"],  
                 figure="SE")
```

- The result will differ from the case where you first calculate the ratio for each observation and then calculate the standard error!



Make use of calibration

```
varianceEstimate(data = df,  
  cols = c("x1", "x2", "I(x1 / x2)"),  
  weights = df[, "weights"],  
  inclusProbs = df[, "inclusProbs"],  
  figure="SE",  
  method="calib",  
  Xs=Xs)
```



If weights were calibrated, additional information is needed

- Weights are calibrated using a matrix.
- Imagine that you want to calibrate your weights, such that the sum of certain variables in the sample are equal to the sum in the population, or on domain level.
 - Number of livestock units -> Calibrate to population total
 - Agricultural area -> Calibrate on regional level
 - Number of farms -> Calibrate on strata level: Region x Farm type
- The head of the calibration matrix “**Xs**” would look something like this:

region	type		LU	area.region1	area.region2	n.region1.type1	n.region1.type2	n.region2.type1	n.region2.type2
1	1		55	34	0	1	0	0	0
1	2		45	76	0	0	1	0	0
2	1		36	0	23	0	0	1	0
2	2		87	0	58	0	0	0	1



Margin of error

- Either specify the multiplier, e.g. 1.96 (for 95% probability).

```
varianceEstimate( ... ,  
                 figure="marginOfError",  
                 MEMultiplier=1.96)
```

- Or, if the sample (or some domain levels) have few observations, let the package estimate the right multiplier. Then, the index has to be given, that specifies the strata (e.g. “region” x “farm type”). Mostly, the margin will be a little higher.

```
varianceEstimate( ... ,  
                 figure="marginOfError",  
                 MEprob=0.95,  
                 indexStrata=df[, "regType"])
```



Calculating probability for differences between years

```
varianceEstimate( ... ,  
                 figure="pValue",  
                 deltaBetweenYears=TRUE,  
                 year0=2000,  
                 year1=2001,  
                 id=df[, "id"],  
                 year=df[, "year"])
```



Estimating on domain level

- The domain level “reg” (region) has 3 values: 1, 2, 3

```
varianceEstimate(...  
  figure="SE",  
  index=df[, "reg"])
```

	X1	X2	X3	X4
1	31.41916	65.32866	90.97115	122.1399
2	32.92920	67.28259	98.33701	123.9328
3	29.73349	58.01865	92.23353	124.3519



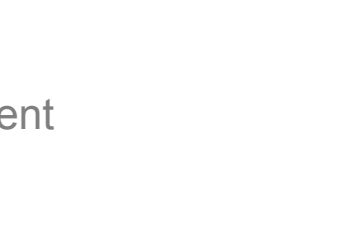
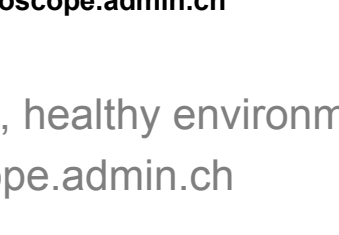
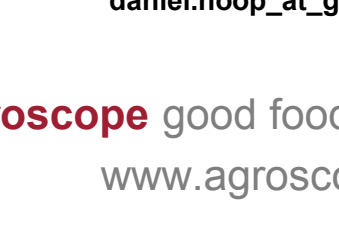
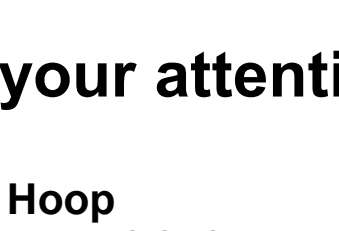
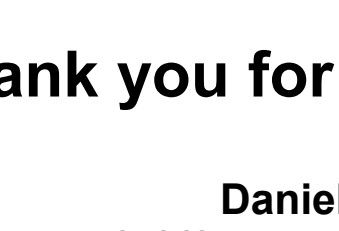
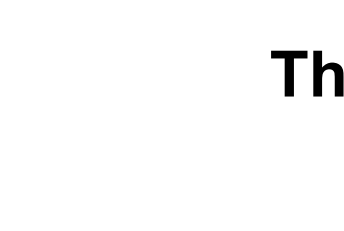
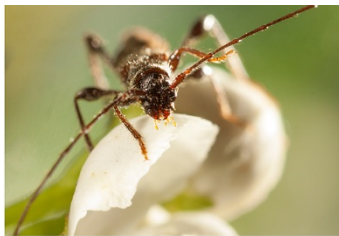
Known issues

- If variables are calibrated redundantly on different domain levels, this may cause problems in variance estimation.
- In some cases, the scale of calibrated variables influences the result of variance estimation. This issue has been reported to the package maintainer of a package we depend on.
 - They know the issue but yet they don't have a solution.



We would like to hear your feedback!

- Before submitting this package to CRAN, we would like to have it user tested.
- Do you think it could be improved?
 - If so, how?
 - Documentation?
 - Functions too complicated to use?
 - Include another use case in the tutorial?
- Please send me an email and I will provide the package archive to install the package on your local machine.
- All you need to know to start:
 - `varEstTutorial()`
 - `?varianceEstimate`



Thank you for your attention

Daniel Hoop
daniel.hoop_at_groscope.admin.ch

Agroscope good food, healthy environment
www.agroscope.admin.ch

