Cluster Sample Size Calculations

```
## -- Attaching packages ---------------------------------------------------------------------------------
## v ggplot2 3.1.1       v purrr   0.3.2
## v tibble  2.1.1       v dplyr   0.8.0.1
## v tidyr   0.8.3       v stringr 1.4.0
## v readr   1.3.1       v forcats 0.4.0
## -- Conflicts ------------------------------------------------------------------------------------------
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
##
## Attaching package: 'data.table'
## The following objects are masked from 'package:dplyr':
##
##     between, first, last
## The following object is masked from 'package:purrr':
##
##     transpose
```

Plain implementations

Function implementation from Hayes & Bennet (Int. J. Epidem. 1999;28:319-326), eqs. 1&2

```
hayes_rates <- function( leff, # rate in exp. group
                         lcont, # rate in contr. group
                         alpha, beta, # significance levels
                         CV, # kappa value
                         clustersize, # average number of subjects per cluster
                         followup # follow-up time, same unit as rates
                         ) {
  zalpha <- qnorm(alpha/2, lower.tail=F)
  zbeta <- qnorm(beta)
  #y <- (zalpha+zbeta)^2*(leff+lcont)/(leff-lcont)^2
  y <- clustersize*followup
  return( 1+(zalpha+zbeta)^2 * ((leff+lcont)/y + CV^2*(leff^2+lcont^2))/(lcont-leff)^2 )
}

hayes_rates( leff=0.01, lcont=0.005, alpha=0.05, beta=0.8, CV=0.55, clustersize=200, followup=2)
```

```
## [1] 24.64475
```

Re-calculation of number of clusters per arm for an unmatched design based on Sample_size_calculations_BG-HMS.xlsx::“Rates - Unmatched”

```
hayes_rates( leff=0.01036, lcont=0.0148, alpha=0.05, beta=0.8, CV=0.29, clustersize=424, followup=1)
```

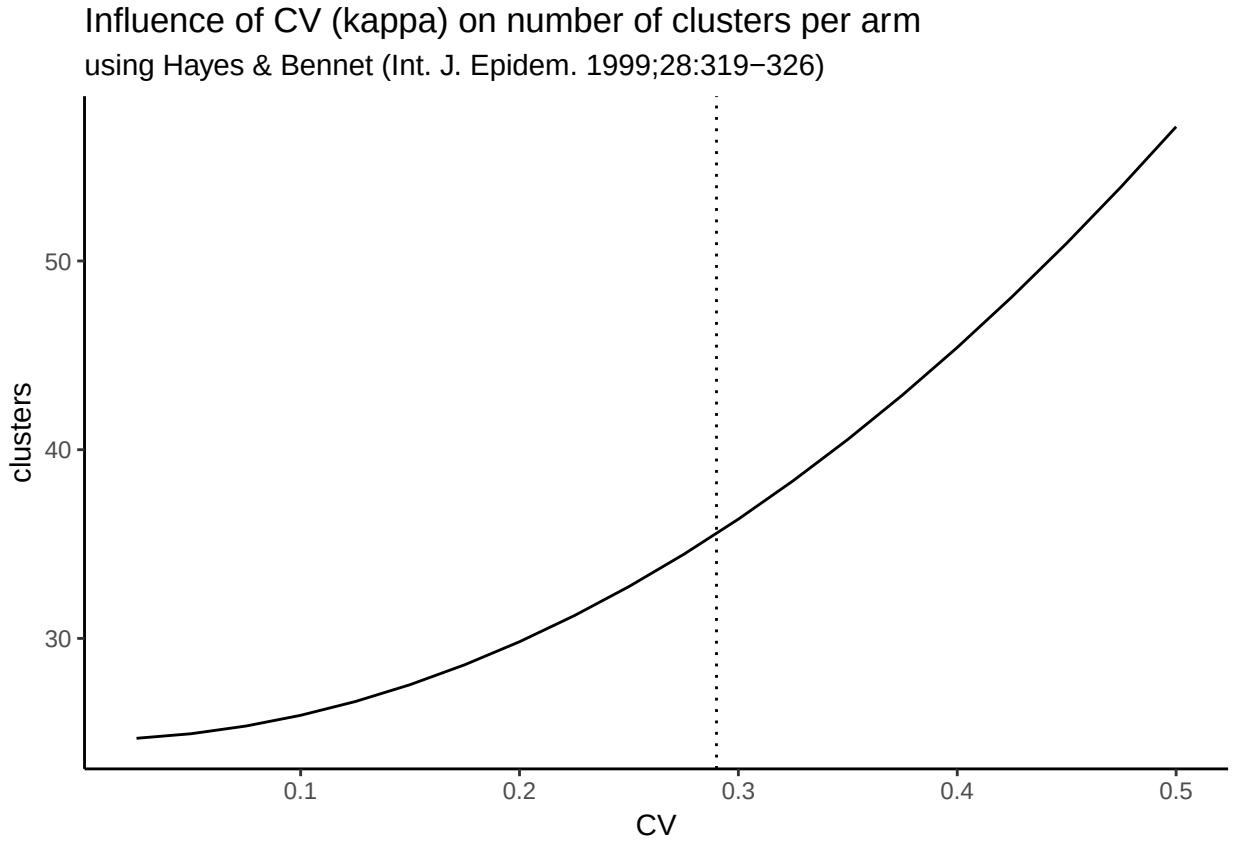
```
## [1] 35.55397
```

Assess sensitivity to CV (kappa) estimates by plotting number of clusters per arm against CV

```
CVs <- seq(0.025, 0.5, by=0.025)

df <- data.frame(
  CV = CVs,
  clusters =sapply( CVs,
          function(x) hayes_rates( leff=0.01036, lcont=0.0148, alpha=0.05, beta=0.8, CV=x, clustersize=4
  )
)

ggplot( df, aes(CV, clusters) ) +
  geom_line() +
  labs( title="Influence of CV (kappa) on number of clusters per arm",
        subtitle="using Hayes & Bennet (Int. J. Epidem. 1999;28:319-326)" ) +
  geom_vline( xintercept = 0.29, linetype="dotted" ) +
  theme_classic()
```



Alternative calculation using CRTSize (CRAN):

"Sample size estimation in cluster (group) randomized trials. Contains traditional power-based methods, empirical smoothing (Rotondi and Donner, 2009), and updated meta-analysis techniques (Rotondi and

Donner, 2012)."

```
n4incidence(le=0.01036, lc=0.0148, m=424, t=1, CV=0.29)
```

```
## The required sample size is a minimum of 35 clusters of size 424 in the Experimental Group
##  and a minimum of 35 clusters (size 424) in the Control Group, followed for time period of length 1
```

Power analysis using clusterPower (CRAN)

The within-cluster variation not given. Calculate from proposed design (i.e. 35 clusters with mean=428 subjects), assuming icc==0.05.

```
varw <- crtpwr.2mean(n=428, cv=0.29, d=0.0148-0.01036, icc=0.05, m=35, method = c("taylor", "weighted")
  tol = .Machine$double.eps^0.25 )

varw
```

```
##         varw
## 0.0008113807
```

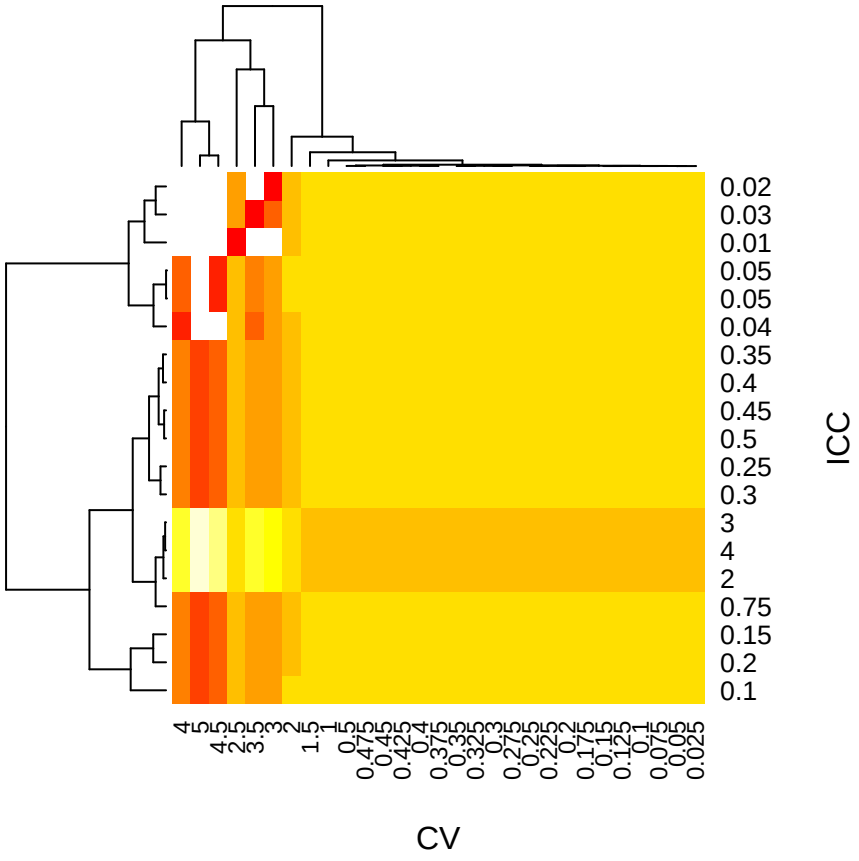
power by ICC and CV (kappa)

Heatmap

```
iccs <- c(seq(0.01,0.05,by=0.01), seq(0.05, 0.5, by=0.05 ), 0.75, 2:4)
CVs <- c(seq(0.025, 0.5, by=0.025), seq(1,5,by=0.5))

mat <- matrix(, nrow=length(iccs), ncol=length(CVs))
for( icc in 1:length(iccs) )
  for( cv in 1:length(CVs))
     mat[icc, cv] <- crtpwr.2mean(power=NA,
                                  n=428,
                                  cv=CVs[cv],
                                  d=0.0148-0.01036,
                                  icc=iccs[icc],
                                  m=35,
                                  varw=varw,
                                  method = c("taylor", "weighted"))
colnames(mat) <- CVs
rownames(mat) <- iccs

heatmap( mat, xlab = "CV", ylab="ICC" )
```



individual sensitivity plots

Set up data set

```
df <-
  expand.grid( ICC = c(seq(0.01,0.05,by=0.01), seq(0.05, 0.5, by=0.05 )),
               CV = c(0.1, 0.2, 0.29, seq(0.4, 3, by=0.1)),
               varw = c(0.001, seq(0.01, 1, by=0.01))) %>%
  mutate( power = crtpwr.2mean(power=NA,
                               n=428,
                               cv=CV,
                               d=0.0148-0.01036,
                               icc=ICC,
                               m=35,
                               varw=varw,
                               method = c("taylor", "weighted")) ) %>%
  filter( !is.na(power)) %>%
  as.data.table()
```

```
## Warning in sqrt(m * n/(2 * VIF)): NaNs produced
```

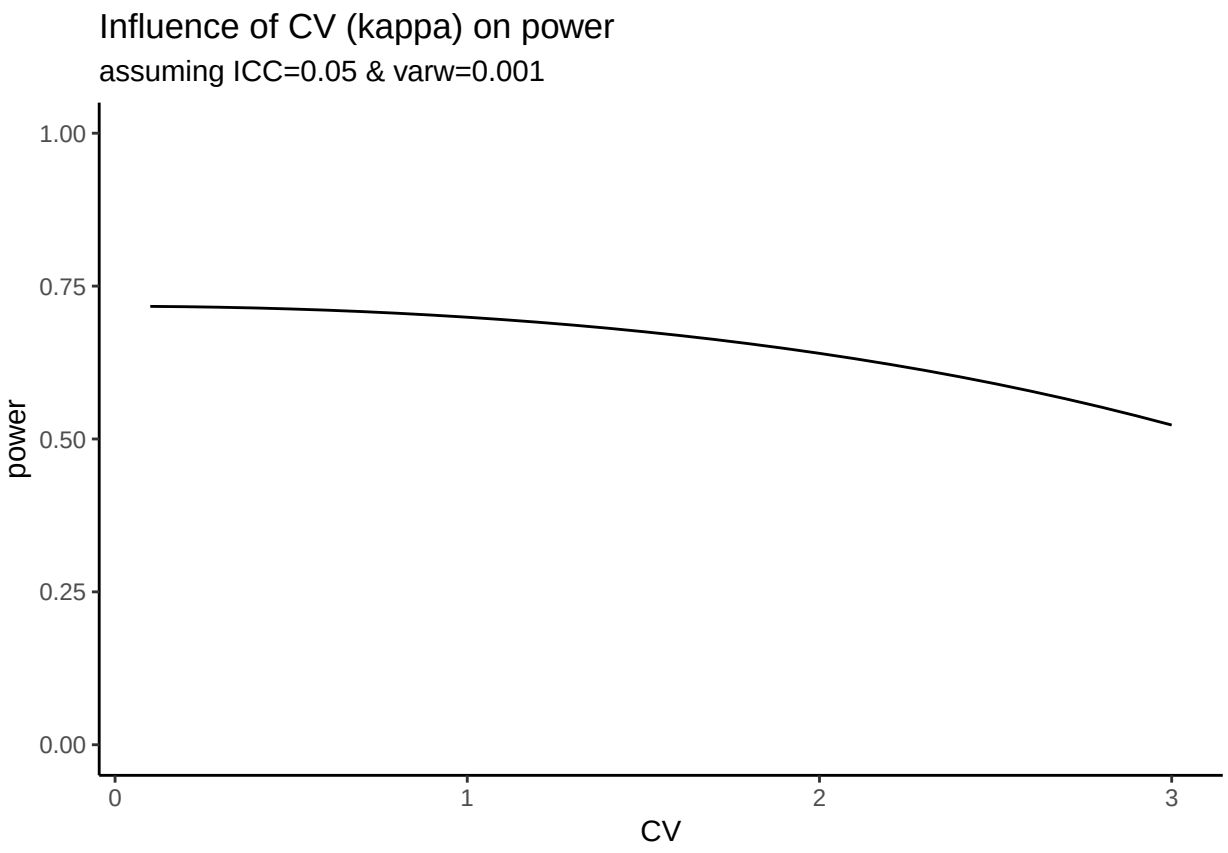
```
## Warning in if (is.na(cv)) {: the condition has length > 1 and only the
## first element will be used
```

```
## Warning in if (is.na(icc)) {: the condition has length > 1 and only the
## first element will be used
```

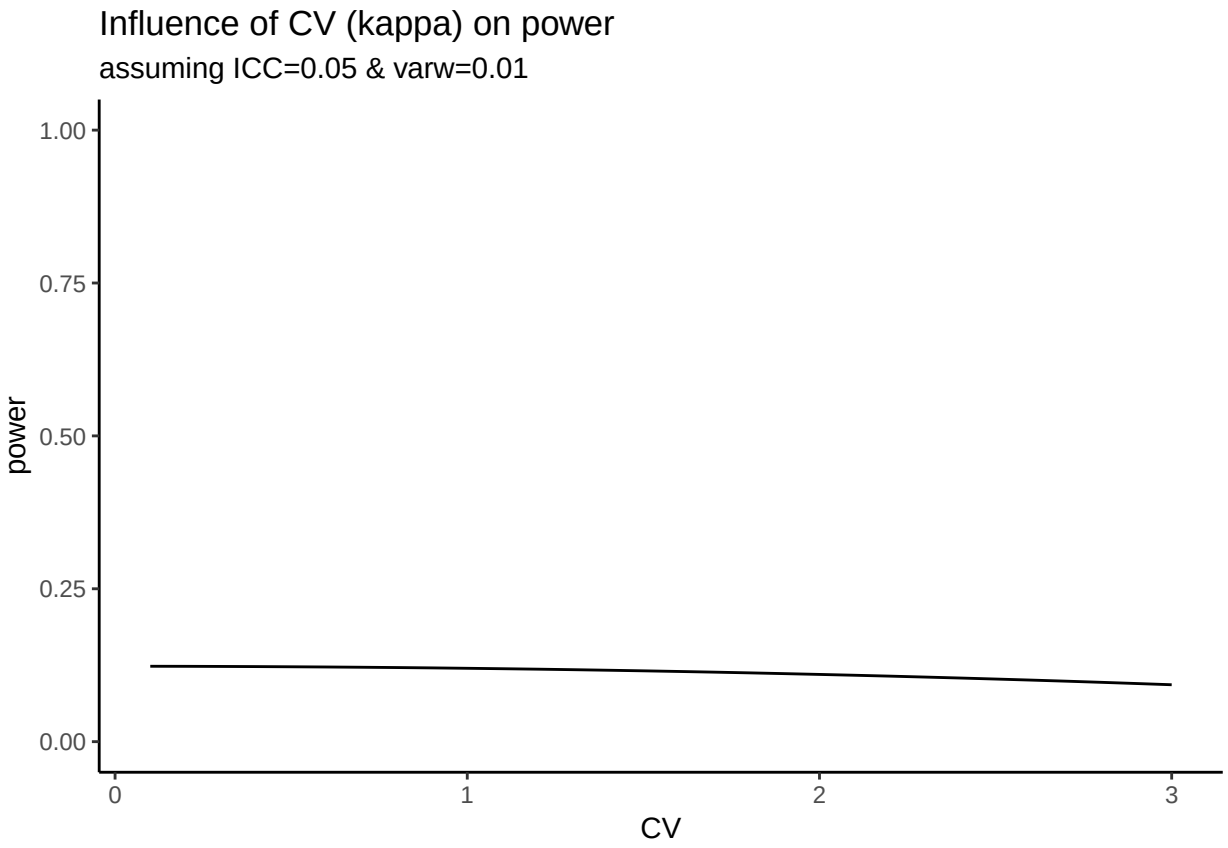
```
## Warning in if (is.na(varw)) {: the condition has length > 1 and only the
## first element will be used
```

Visualizations

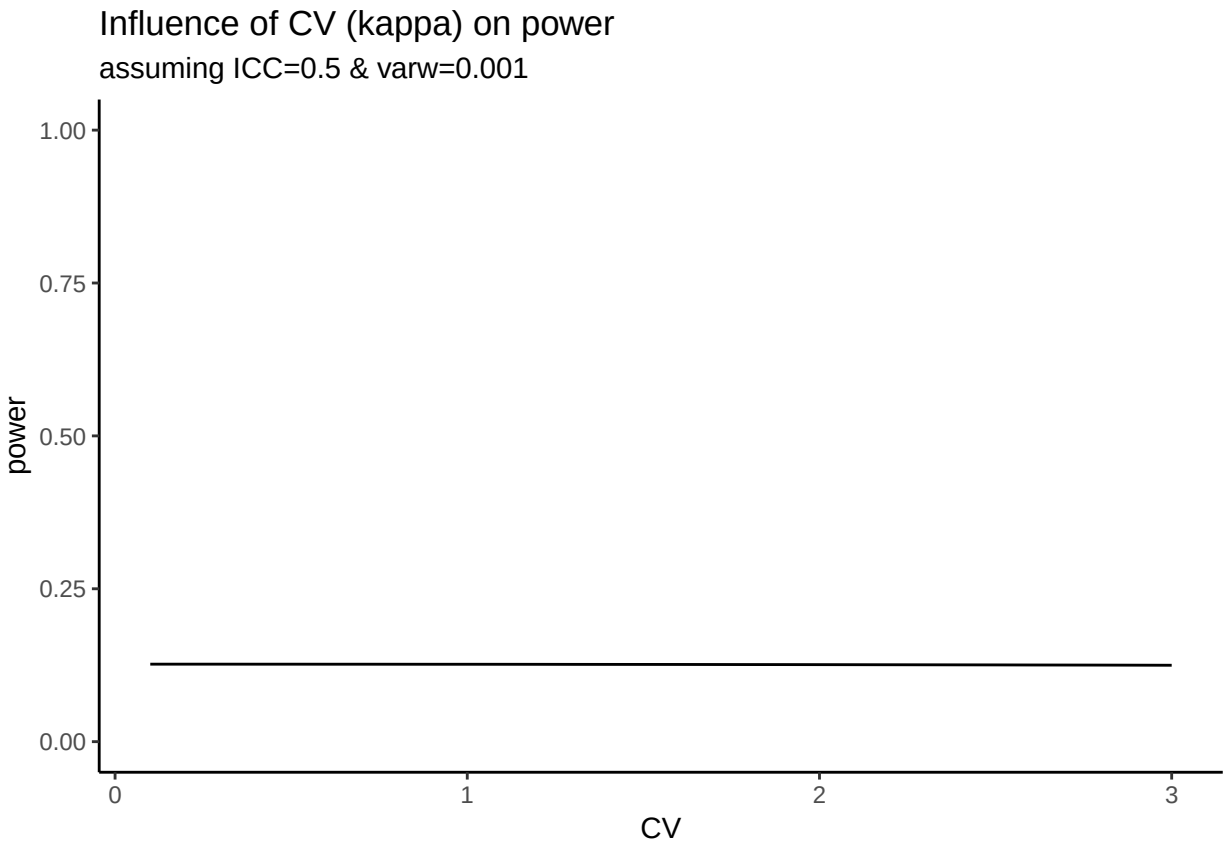
```
ggplot( df[ICC==0.05 & varw==0.001], aes(CV, power) ) +
  geom_line() +
  scale_y_continuous( limits=c(0,1))+
  labs( title="Influence of CV (kappa) on power",
        subtitle="assuming ICC=0.05 & varw=0.001") +
  theme_classic()
```



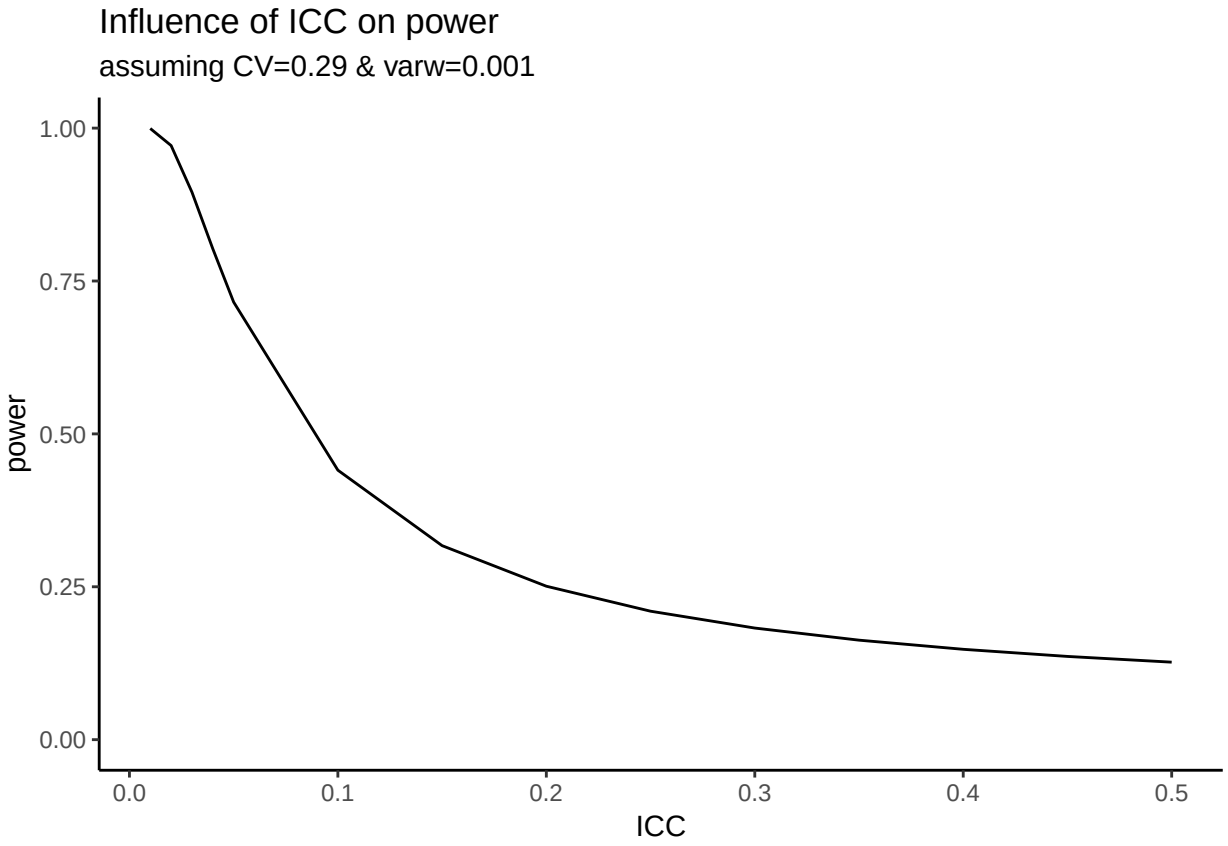

```
ggplot( df[ICC==0.05 & varw==0.01], aes(CV, power) ) +
  geom_line() +
  scale_y_continuous( limits=c(0,1))+
  labs( title="Influence of CV (kappa) on power",
        subtitle="assuming ICC=0.05 & varw=0.01") +
  theme_classic()
```




```
ggplot( df[ICC==0.5 & varw==0.001], aes(CV, power) ) +
  geom_line() +
  scale_y_continuous( limits=c(0,1))+
  labs( title="Influence of CV (kappa) on power",
        subtitle="assuming ICC=0.5 & varw=0.001") +
  theme_classic()
```




```
ggplot( df[CV==0.29 & varw==0.001], aes(ICC, power) ) +
  geom_line() +
  scale_y_continuous( limits=c(0,1))+
  labs( title="Influence of ICC on power",
        subtitle="assuming CV=0.29 & varw=0.001") +
  theme_classic()
```


Influence of ICC on power
assuming CV=0.29 & varw=0.001
power
1.00
0.75
0.50
0.25
0.00
0.0
0.1
0.2
0.3
0.4
0.5
ICC