Modelagem-01

Panosso AR & Oliveira JA

2023-05-16

### Análise de modelagem 01

#### Carregando a base de dados

#sisvan\_estab <- read\_rds("../data/sisvan\_estab.rds")  
#sisvan\_estab$files\_way %>% unique()  
#sisvan\_estab$ano %>% unique()  
sisvan\_estab <- read\_rds("../data/sisvan\_estab.rds")  
glimpse(sisvan\_estab)

## Rows: 12,622,130  
## Columns: 13  
## $ files\_way <chr> "1", "1", "1", "1", "1", "1…  
## $ ano <dbl> 2006, 2006, 2006, 2006, 200…  
## $ mes <int> 2, 2, 2, 2, 2, 2, 2, 2, 2, …  
## $ sigla\_uf <chr> "SP", "SP", "SP", "SP", "SP…  
## $ id\_municipio <chr> "3501608", "3501608", "3504…  
## $ id\_municipio\_6 <chr> "350160", "350160", "350410…  
## $ indicador\_servico\_nutricao\_proprio <int> 0, 1, 0, 0, 0, 1, 1, 0, 0, …  
## $ indicador\_servico\_nutricao\_terceirizado <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, …  
## $ indicador\_servico\_lactario\_proprio <int> 0, 0, 0, 0, 0, 1, 1, 0, 0, …  
## $ indicador\_servico\_lactario\_terceirizado <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, …  
## $ indicador\_servico\_banco\_leite\_proprio <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, …  
## $ indicador\_servico\_banco\_leite\_terceirizado <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, …  
## $ municipio <chr> "Americana", "Americana", "…

#### Resumo estatístico da base

sisvan\_estab %>%   
 group\_by(ano, municipio) %>%   
 summarise(  
 servico\_nutricao\_proprio= sum(indicador\_servico\_nutricao\_proprio),  
 servico\_nutricao\_terceirizado = sum(indicador\_servico\_nutricao\_terceirizado),  
 servico\_lactario\_proprio= sum(indicador\_servico\_lactario\_proprio),   
 servico\_lactario\_terceirizado= sum(indicador\_servico\_lactario\_terceirizado), servico\_banco\_leite\_proprio= sum(indicador\_servico\_banco\_leite\_proprio), servico\_banco\_leite\_terceirizado= sum(indicador\_servico\_banco\_leite\_terceirizado)  
 )

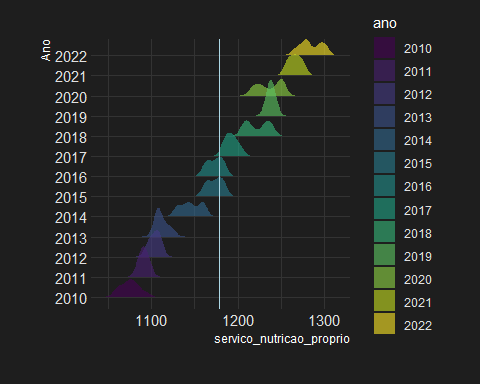
## # A tibble: 10,965 × 8  
## # Groups: ano [17]  
## ano municipio servico\_nutricao\_proprio servico\_nutricao\_terceiriz…¹  
## <dbl> <chr> <int> <int>  
## 1 2006 Adamantina 24 0  
## 2 2006 Adolfo 0 0  
## 3 2006 Aguaí 12 0  
## 4 2006 Agudos 12 0  
## 5 2006 Alambari 0 0  
## 6 2006 Alfredo Marcondes 0 0  
## 7 2006 Altair 0 0  
## 8 2006 Altinópolis 11 0  
## 9 2006 Alto Alegre 12 0  
## 10 2006 Alumínio 0 0  
## # ℹ 10,955 more rows  
## # ℹ abbreviated name: ¹​servico\_nutricao\_terceirizado  
## # ℹ 4 more variables: servico\_lactario\_proprio <int>,  
## # servico\_lactario\_terceirizado <int>, servico\_banco\_leite\_proprio <int>,  
## # servico\_banco\_leite\_terceirizado <int>

## Nutrição

df\_aux <- sisvan\_estab %>%   
 mutate(  
 ano\_mes = make\_date(year=ano,month=mes,day=1)) %>%   
 group\_by(ano, mes, ano\_mes) %>%   
 summarise(  
 servico\_nutricao\_proprio= sum(indicador\_servico\_nutricao\_proprio),  
 servico\_nutricao\_terceirizado = sum(indicador\_servico\_nutricao\_terceirizado),  
 servico\_lactario\_proprio= sum(indicador\_servico\_lactario\_proprio),   
 servico\_lactario\_terceirizado= sum(indicador\_servico\_lactario\_terceirizado), servico\_banco\_leite\_proprio= sum(indicador\_servico\_banco\_leite\_proprio), servico\_banco\_leite\_terceirizado= sum(indicador\_servico\_banco\_leite\_terceirizado)  
 )

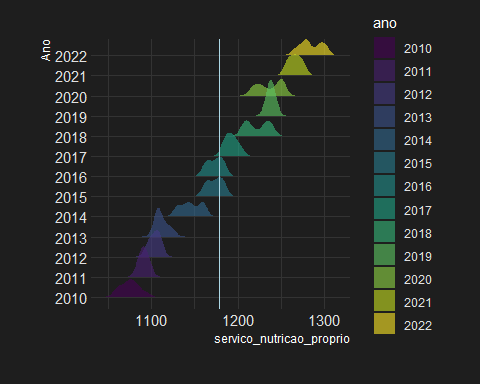
### Estatística Descritiva

df\_aux %>% filter(ano >= 2010) %>%   
 mutate(ano = as.factor(ano)) %>%   
 ggplot(aes(x = servico\_nutricao\_proprio, y = ano, fill = ano)) +  
 ggridges::geom\_density\_ridges(color = 'transparent', alpha = .6) +  
 scale\_fill\_viridis\_d() +  
 labs(  
 x = "servico\_nutricao\_proprio",  
 y = "Ano"  
 ) +  
 theme(  
 legend.position = 'none',axis.text = element\_text(size = 8)  
 ) +  
 geom\_vline(xintercept = df\_aux %>% ungroup %>%   
 filter(ano >= 2010) %>%   
 pull(servico\_nutricao\_proprio) %>% mean(),  
 color="lightblue") +  
 #theme\_bw() +  
 hrbrthemes::theme\_modern\_rc()

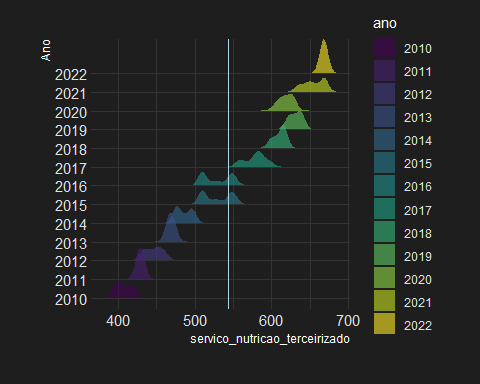


variaveis <- c("servico\_nutricao\_proprio",  
 "servico\_nutricao\_terceirizado",  
 "servico\_lactario\_proprio",  
 "servico\_lactario\_terceirizado",  
 "servico\_banco\_leite\_proprio",  
 "servico\_banco\_leite\_terceirizado")  
  
map(variaveis, ~{  
 df\_aux %>% filter(ano >= 2010) %>%   
 mutate(ano = as.factor(ano)) %>%   
 ggplot(aes\_string(x = .x, y = "ano", fill = "ano")) +  
 ggridges::geom\_density\_ridges(color = 'transparent', alpha = .6) +  
 scale\_fill\_viridis\_d() +  
 labs(  
 x = .x,  
 y = "Ano"  
 ) +  
 theme(  
 legend.position = 'none',axis.text = element\_text(size = 8)  
 ) +  
 geom\_vline(xintercept = df\_aux %>% ungroup %>%  
 filter(ano >= 2010) %>%  
 pull(.x) %>% mean(),  
 color="lightblue") +  
 #theme\_bw() +  
 hrbrthemes::theme\_modern\_rc()  
})

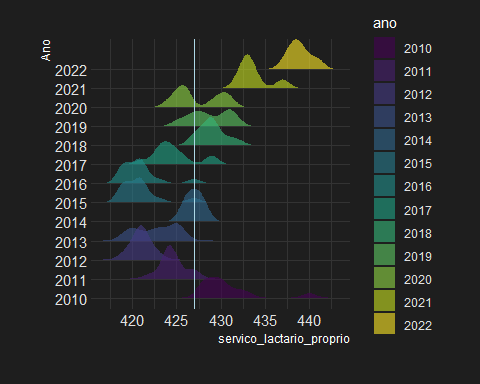
## [[1]]



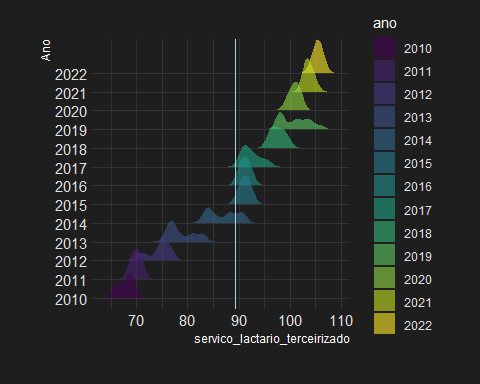
##   
## [[2]]



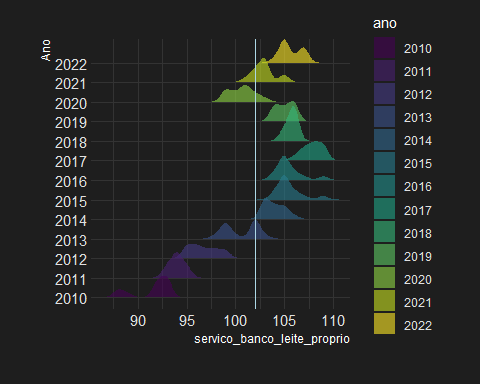
##   
## [[3]]



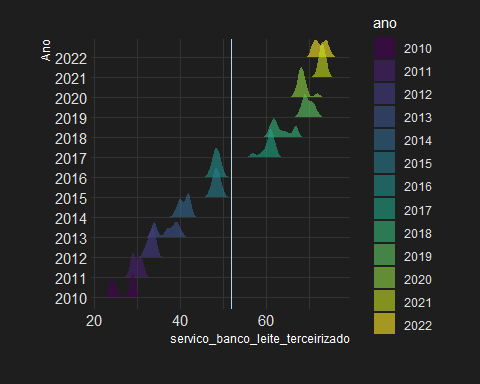
##   
## [[4]]



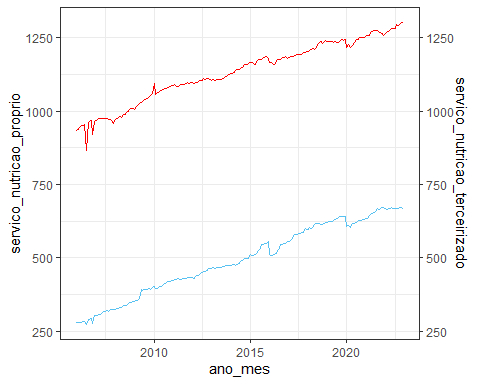
##   
## [[5]]



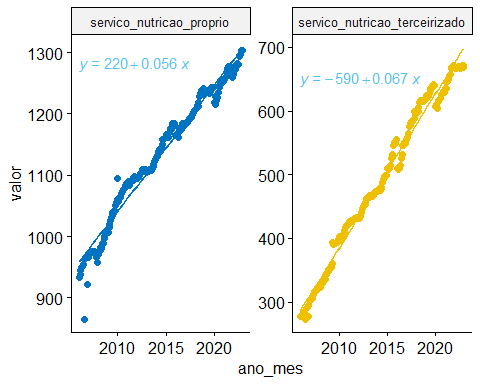
##   
## [[6]]



coeff <- 1  
df\_aux %>%   
 ggplot(aes(x = ano\_mes)) +  
 geom\_line(aes(y = servico\_nutricao\_proprio), color="red") +  
 geom\_line(aes(y = servico\_nutricao\_terceirizado/coeff)) +  
 scale\_y\_continuous(  
 name = "servico\_nutricao\_proprio",  
 sec.axis = sec\_axis(~.\*coeff, name="servico\_nutricao\_terceirizado")  
 ) +  
 theme\_bw()

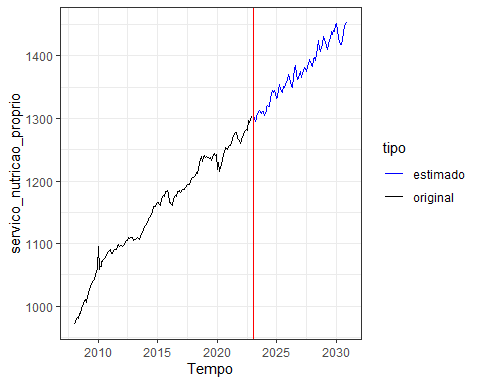


df\_aux %>%   
 select(ano\_mes,servico\_nutricao\_proprio,servico\_nutricao\_terceirizado) %>%   
 pivot\_longer(cols = c("servico\_nutricao\_proprio","servico\_nutricao\_terceirizado"),  
 names\_to = "tipo",  
 values\_to = "valor") %>%   
 ggscatter(  
 x = "ano\_mes", y = "valor",  
 color = "tipo", palette = "jco",  
 add = "reg.line"  
 ) +  
 facet\_wrap(~as.factor(tipo),scales = "free") +  
 #stat\_cor() +   
 stat\_regline\_equation()+  
 theme( legend.position = "none")

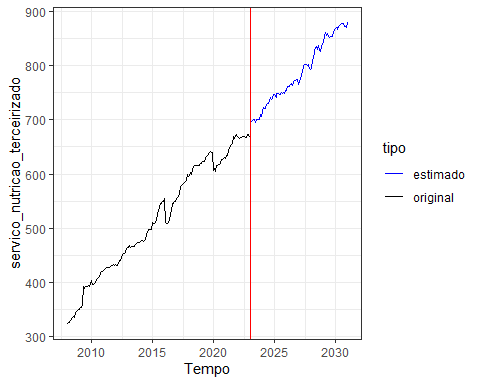


### Projeção

previsao\_sisvan\_estab(servico\_nutricao\_proprio,ano\_corte=2007,ano\_final=2030, janela = 3)



previsao\_sisvan\_estab(servico\_nutricao\_terceirizado,ano\_corte=2007,ano\_final=2030, janela = 5)

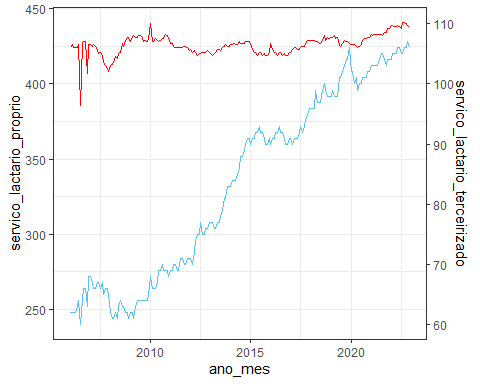


sisvan\_estab %>% mutate\_if(is.integer64, as.integer) %>%   
 mutate(  
 ano\_mes = make\_date(year=ano,month=mes,day=1)) %>%   
 group\_by(ano, mes) %>%   
 summarise(  
 servico\_nutricao\_proprio= sum(indicador\_servico\_nutricao\_proprio),  
 servico\_nutricao\_terceirizado = sum(indicador\_servico\_nutricao\_terceirizado),  
 servico\_lactario\_proprio= sum(indicador\_servico\_lactario\_proprio),   
 servico\_lactario\_terceirizado= sum(indicador\_servico\_lactario\_terceirizado),   
 servico\_banco\_leite\_proprio= sum(indicador\_servico\_banco\_leite\_proprio),   
 servico\_banco\_leite\_terceirizado= sum(indicador\_servico\_banco\_leite\_terceirizado)  
 ) %>%   
 group\_by(ano) %>%   
 summarise(  
 servico\_nutricao\_proprio = mean(servico\_nutricao\_proprio),  
 servico\_nutricao\_terceirizado = mean(servico\_nutricao\_terceirizado),  
 # servico\_lactario\_proprio= mean(servico\_lactario\_proprio),  
 # servico\_lactario\_terceirizado= mean(servico\_lactario\_terceirizado),  
 # servico\_banco\_leite\_proprio= mean(servico\_banco\_leite\_proprio),   
 # servico\_banco\_leite\_terceirizado= mean(servico\_banco\_leite\_terceirizado)  
 )

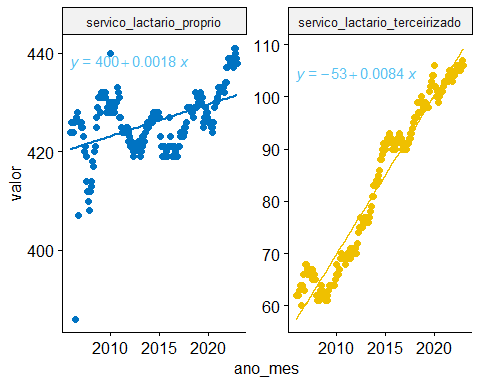
## # A tibble: 17 × 3  
## ano servico\_nutricao\_proprio servico\_nutricao\_terceirizado  
## <dbl> <dbl> <dbl>  
## 1 2006 943. 284.  
## 2 2007 972. 314.  
## 3 2008 990. 336.  
## 4 2009 1035. 380.  
## 5 2010 1075. 408.  
## 6 2011 1091. 429   
## 7 2012 1104 447.  
## 8 2013 1112. 469.  
## 9 2014 1146. 486.  
## 10 2015 1174. 530.  
## 11 2016 1174. 530.  
## 12 2017 1193. 579.  
## 13 2018 1223. 611.  
## 14 2019 1239. 632.  
## 15 2020 1237. 619.  
## 16 2021 1267. 656.  
## 17 2022 1285. 669.

### Lactário

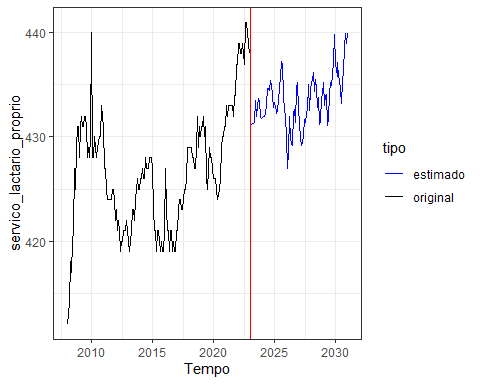
coeff <- 1/4  
df\_aux %>%   
 ggplot(aes(x = ano\_mes)) +  
 geom\_line(aes(y = servico\_lactario\_proprio), color="red") +  
 geom\_line(aes(y = servico\_lactario\_terceirizado/coeff)) +  
 scale\_y\_continuous(  
 name = "servico\_lactario\_proprio",  
 sec.axis = sec\_axis(~.\*coeff, name="servico\_lactario\_terceirizado")  
 ) +  
 theme\_bw()



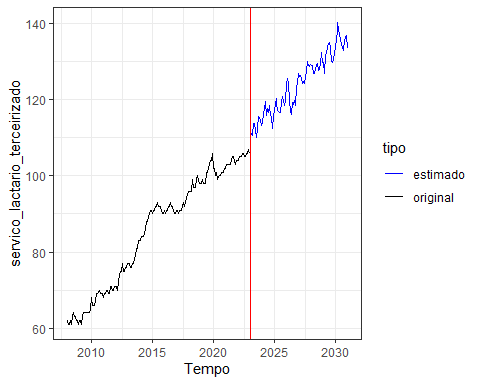
df\_aux %>%   
 select(ano\_mes,servico\_lactario\_proprio,servico\_lactario\_terceirizado) %>%   
 pivot\_longer(cols = c("servico\_lactario\_proprio","servico\_lactario\_terceirizado"),  
 names\_to = "tipo",  
 values\_to = "valor") %>%   
 ggscatter(  
 x = "ano\_mes", y = "valor",  
 color = "tipo", palette = "jco",  
 add = "reg.line"  
 ) +  
 facet\_wrap(~as.factor(tipo),scales = "free") +  
 #stat\_cor() +   
 stat\_regline\_equation()+  
 theme( legend.position = "none")



previsao\_sisvan\_estab(servico\_lactario\_proprio,ano\_corte=2007,ano\_final=2030, janela = 6)



previsao\_sisvan\_estab(servico\_lactario\_terceirizado,ano\_corte=2007,ano\_final=2030, janela = 3)

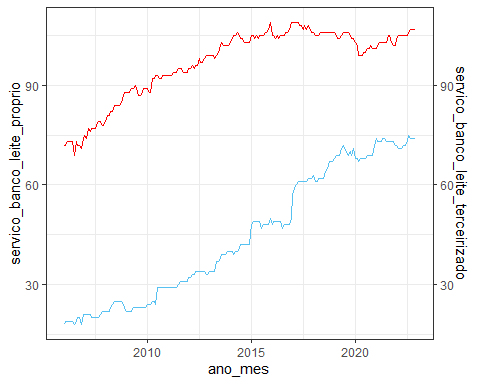


sisvan\_estab %>% mutate\_if(is.integer64, as.integer) %>%   
 mutate(  
 ano\_mes = make\_date(year=ano,month=mes,day=1)) %>%   
 group\_by(ano, mes) %>%   
 summarise(  
 servico\_nutricao\_terceirizado = sum(indicador\_servico\_nutricao\_terceirizado),  
 servico\_lactario\_proprio= sum(indicador\_servico\_lactario\_proprio),   
 servico\_lactario\_terceirizado= sum(indicador\_servico\_lactario\_terceirizado),   
 servico\_banco\_leite\_proprio= sum(indicador\_servico\_banco\_leite\_proprio),   
 servico\_banco\_leite\_terceirizado= sum(indicador\_servico\_banco\_leite\_terceirizado)  
 ) %>%   
 group\_by(ano) %>%   
 summarise(  
 # servico\_nutricao\_terceirizado = mean(servico\_nutricao\_terceirizado),  
 servico\_lactario\_proprio= mean(servico\_lactario\_proprio),  
 servico\_lactario\_terceirizado= mean(servico\_lactario\_terceirizado),  
 # servico\_banco\_leite\_proprio= mean(servico\_banco\_leite\_proprio),   
 # servico\_banco\_leite\_terceirizado= mean(servico\_banco\_leite\_terceirizado)  
 )

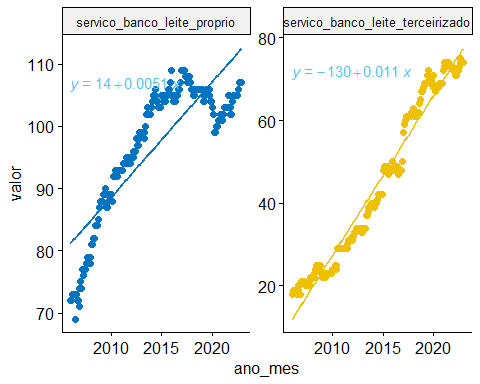
## # A tibble: 17 × 3  
## ano servico\_lactario\_proprio servico\_lactario\_terceirizado  
## <dbl> <dbl> <dbl>  
## 1 2006 421. 63.5  
## 2 2007 419. 66.3  
## 3 2008 422. 62.1  
## 4 2009 430. 63.4  
## 5 2010 431. 68   
## 6 2011 425. 70   
## 7 2012 421 74.4  
## 8 2013 423. 78.8  
## 9 2014 427 86.9  
## 10 2015 421. 91.2  
## 11 2016 421. 91.2  
## 12 2017 425. 92.4  
## 13 2018 429 97.7  
## 14 2019 429. 101   
## 15 2020 427. 101.   
## 16 2021 434. 104.   
## 17 2022 439. 106.

#### Banco de leite

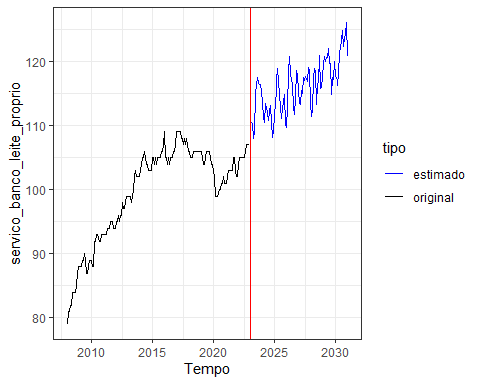
coeff <- 1  
df\_aux %>%   
 ggplot(aes(x = ano\_mes)) +  
 geom\_line(aes(y = servico\_banco\_leite\_proprio), color="red") +  
 geom\_line(aes(y = servico\_banco\_leite\_terceirizado/coeff)) +  
 scale\_y\_continuous(  
 name = "servico\_banco\_leite\_proprio",  
 sec.axis = sec\_axis(~.\*coeff, name="servico\_banco\_leite\_terceirizado")  
 ) +  
 theme\_bw()



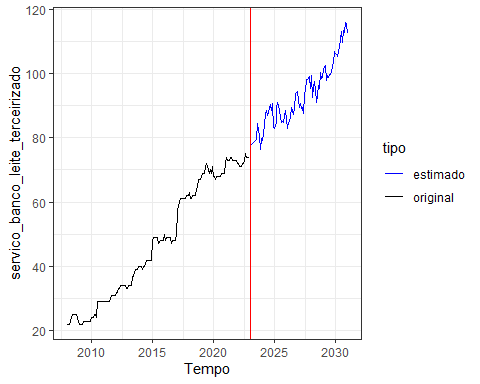
df\_aux %>%   
 select(ano\_mes,servico\_banco\_leite\_proprio,servico\_banco\_leite\_terceirizado) %>%   
 pivot\_longer(cols = c("servico\_banco\_leite\_proprio","servico\_banco\_leite\_terceirizado"),  
 names\_to = "tipo",  
 values\_to = "valor") %>%   
 ggscatter(  
 x = "ano\_mes", y = "valor",  
 color = "tipo", palette = "jco",  
 add = "reg.line"  
 ) +  
 facet\_wrap(~as.factor(tipo), scale="free") +  
 # stat\_cor(label.y = 70) +   
 stat\_regline\_equation()+  
 theme( legend.position = "none")



previsao\_sisvan\_estab(servico\_banco\_leite\_proprio,ano\_corte=2007,ano\_final=2030, janela = 3)



previsao\_sisvan\_estab(servico\_banco\_leite\_terceirizado,ano\_corte=2015,ano\_final=2030, janela = 8)



sisvan\_estab %>% mutate\_if(is.integer64, as.integer) %>%   
 mutate(  
 ano\_mes = make\_date(year=ano,month=mes,day=1)) %>%   
 group\_by(ano, mes) %>%   
 summarise(  
 servico\_nutricao\_terceirizado = sum(indicador\_servico\_nutricao\_terceirizado),  
 servico\_lactario\_proprio= sum(indicador\_servico\_lactario\_proprio),   
 servico\_lactario\_terceirizado= sum(indicador\_servico\_lactario\_terceirizado),   
 servico\_banco\_leite\_proprio= sum(indicador\_servico\_banco\_leite\_proprio),   
 servico\_banco\_leite\_terceirizado= sum(indicador\_servico\_banco\_leite\_terceirizado)  
 ) %>%   
 group\_by(ano) %>%   
 summarise(  
 # servico\_nutricao\_terceirizado = mean(servico\_nutricao\_terceirizado),  
 # servico\_lactario\_proprio= mean(servico\_lactario\_proprio),   
 # servico\_lactario\_terceirizado= mean(servico\_lactario\_terceirizado),   
 servico\_banco\_leite\_proprio= mean(servico\_banco\_leite\_proprio),   
 servico\_banco\_leite\_terceirizado= mean(servico\_banco\_leite\_terceirizado)  
 )

## # A tibble: 17 × 3  
## ano servico\_banco\_leite\_proprio servico\_banco\_leite\_terceirizado  
## <dbl> <dbl> <dbl>  
## 1 2006 72.2 19.1  
## 2 2007 77.1 20.8  
## 3 2008 83.4 23.8  
## 4 2009 88.4 22.7  
## 5 2010 91.5 26.7  
## 6 2011 94 29.9  
## 7 2012 96.8 33.3  
## 8 2013 101. 37.1  
## 9 2014 104 41   
## 10 2015 106. 48.5  
## 11 2016 106. 48.5  
## 12 2017 108. 60.7  
## 13 2018 106. 63.6  
## 14 2019 105. 69.8  
## 15 2020 101. 68.6  
## 16 2021 103. 73.2  
## 17 2022 106. 72.8