
Economía Aplicada: Problem Set N°9

Milton Bronstein Felipe García Vassallo Santiago López Franco Riottini

En este Problem Set realizamos simulaciones de *power calculations*.

En el primer gráfico, lo que vemos son las simulaciones para distinto número de observaciones y para distintos efectos del $\hat{\beta}$ hipotético. Como se puede ver, la figura cumple con la aproximación teórica al tema hecha en clases. En este sentido, a medida que el $\hat{\beta}$ aumenta, también lo hace el *power*, mientras que también aumenta el mismo cuando lo que crece es el tamaño de la muestra. Curiosamente, cuando el $\hat{\beta}$ es de un 10 %, tendremos un *power* de 1, al menos con muestras mayores a $n = 1000$. Esto indica que la probabilidad de rechazar la hipótesis nula cuando la misma no sea verdadera se dará en el 100 % de los casos.

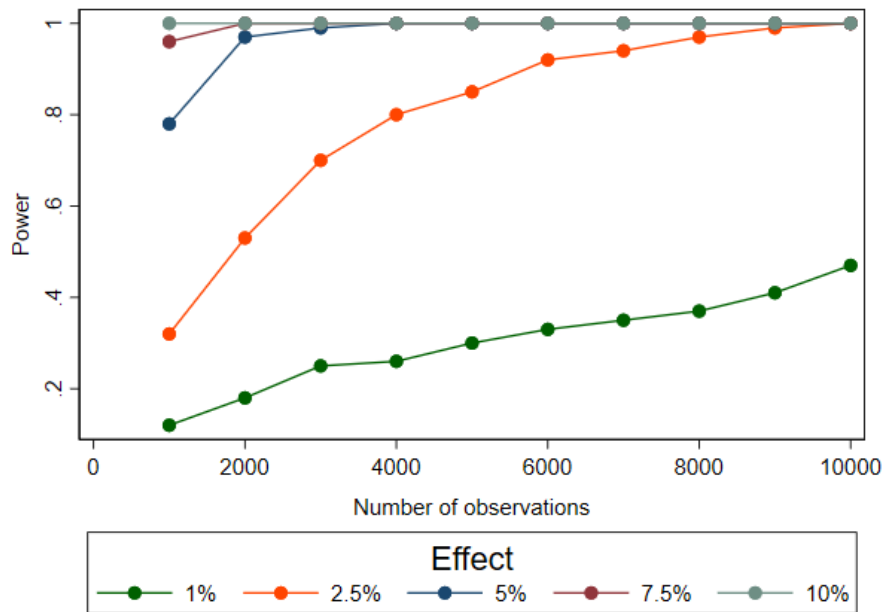


Figura 1: First power calculation

Se puede observar que al aumentar la varianza del término de error aumenta la cantidad de n necesario para mejorar el *power* a valores relativamente buenos. En el caso de intentar captar un efecto del 1 % o 2,5 % parece no ser suficiente ni con 10000 observaciones.

Podemos anticipar esto sabiendo que:

$$V(\hat{\beta}_j) = \frac{\sigma^2}{n(1 - R_j^2)V(X_j)} \quad (1)$$

Y aumentando el σ^2 aumentamos la varianza, lo cual afecta negativamente al *power*

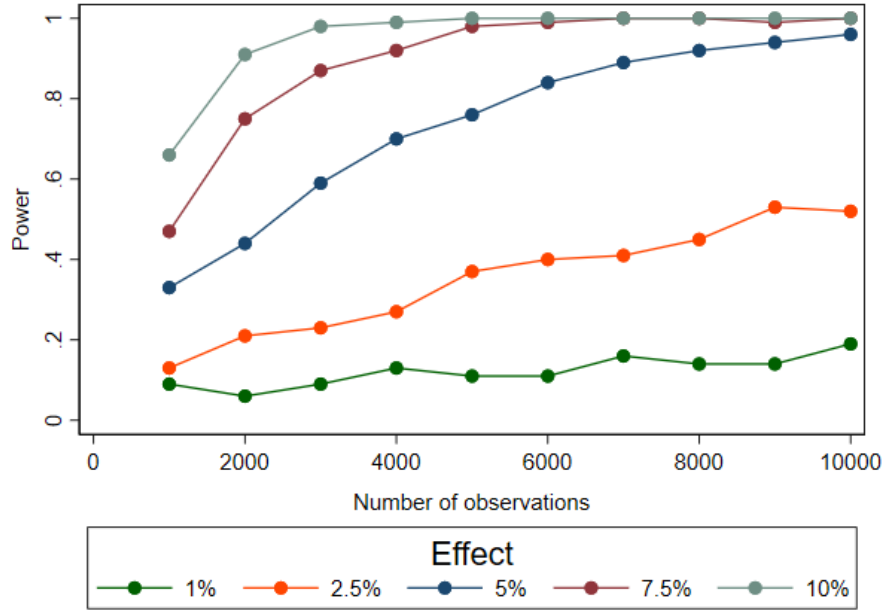


Figura 2: Second power calculation - aumento de la varianza del término de error a 5000

En los dos siguientes gráficos podemos ver que al tener una asignación más alejada de la óptima (50% en tratamiento y 50% en control) el *power* en comparación con la figura 2 de los dos son peores y casi en misma medida (cuanto más grande el n más igual es el *power* en promedio). Esto se debe a que tenemos una peor $V(X_j)$, que viéndolo en la ecuación 1 un valor más chico de esta varianza termina en menor *power*, por una mayor $V(\hat{\beta}_j)$.

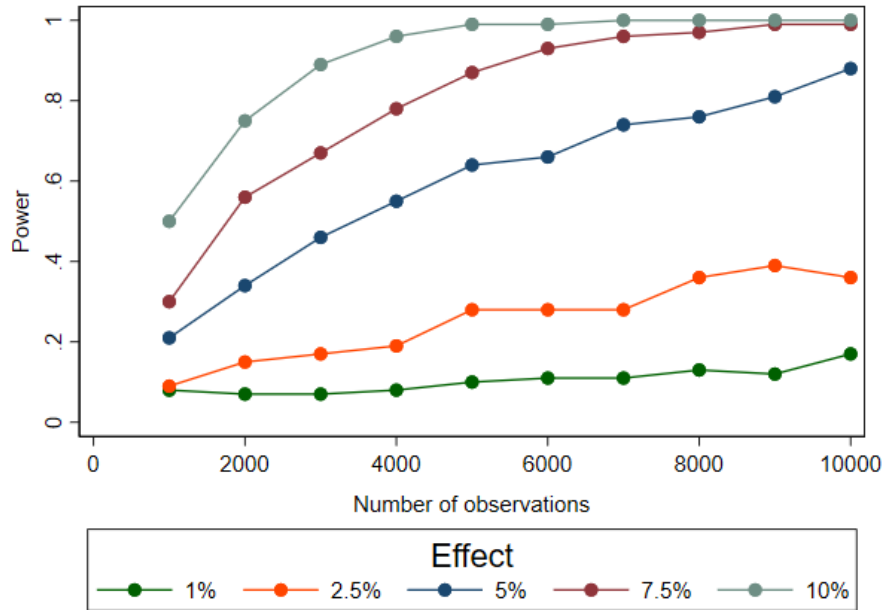


Figura 3: Third power calculation - varianza del término de error en 5000 y 20% de observaciones en tratamiento

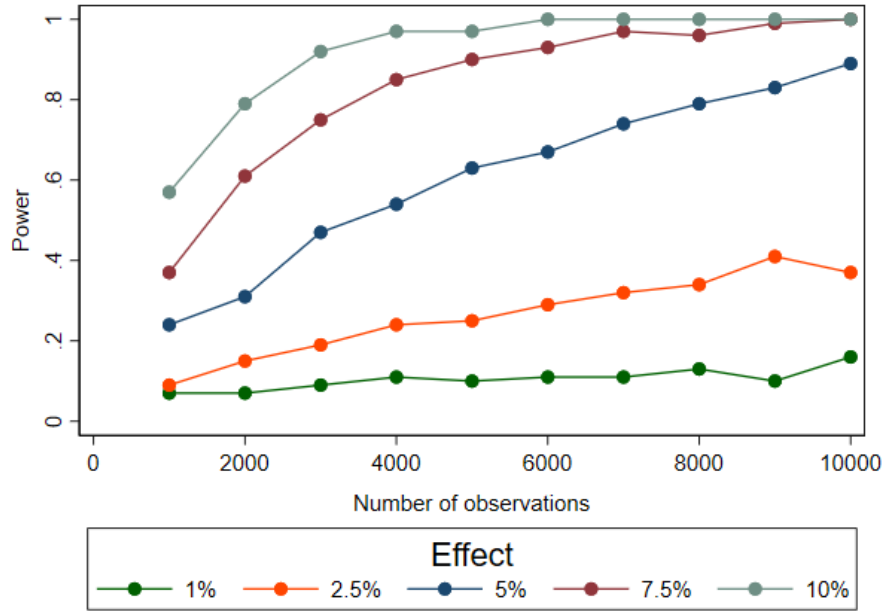


Figura 4: Fourth power calculation - varianza del término de error en 5000 y 80 % de observaciones en tratamiento

El agregar una variable como control purga el término de error y por lo tanto baja el σ^2 que se encuentra en la ecuación 1, disminuyendo la $V(\hat{\beta})$ y mejorando el *power* relativo en comparación con el de la figura 1.

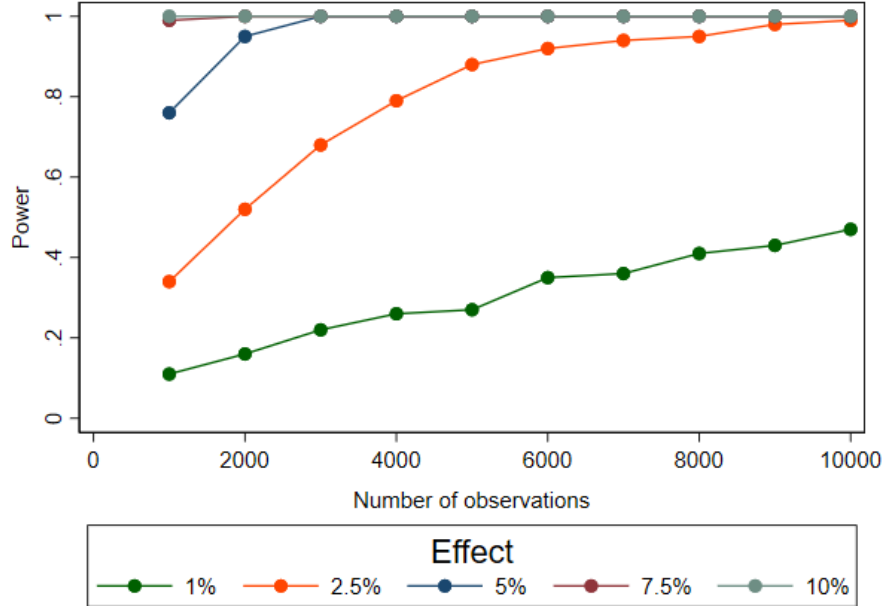


Figura 5: Fifth power calculation - se agrega como control las ganancias estimadas para 2019



```

/*****
Semana 10: Power calculations

Universidad de San Andrés
Economía Aplicada

2022

>
*****/
*      Bronstein      García Vassallo      López      Riottini
/*****/
Este archivo sigue la siguiente estructura:

0) Set up environment

1) Replicación de gráficos

*****/
* 0) Set up environment
*=====

global main "C:\Users\Franco\Documents\GitHub\Problem-Sets\PS 9"
global output "$main/output"
global input "$main/input"

cd "$main"

* Generamos una muestra simulada con datos de pagos de impuestos de empresas
*=====
clear all
set seed 123 // seteamos semilla para poder replicar los resultados
set obs 15000
gen ganancias_estimadas = rnormal(10000,2000)
drop if ganancias_estimadas<0

gen impuestos_pagados = 0.2*ganancias_estimadas + rnormal(0,500) // este es el termino
> de error que hay que modificar en el 2.
drop if impuestos_pagados<0
*=====

* Simulamos un efecto del 2.5% con 2000 obs
*=====
preserve

sample 2000, count // aleatoriamente me quedo con 2000 obs

* Aleatoriamente asigno el tratamiento al 50% de las obs:
gen temp = runiform()
gen T=0
replace T = 1 if temp<0.5

* A los tratados les aumento el outcome un 2.5%
replace impuestos_pagados = impuestos_pagados * (1+0.025) if T==1

* Regreso el outcome en el tratamiento
reg impuestos_pagados T, robust

restore
*=====

* Repetimos 500 veces y nos fijamos el % de veces que rechazamos H0
*=====
mat R = J(500,2,.)
forvalues x=1(1)500 {
preserve

sample 2000, count

gen temp = runiform()
gen T=0
replace T = 1 if temp<0.5

```

```

replace impuestos_pagados = impuestos_pagados * (1+0.025) if T==1
reg impuestos_pagados T, robust
mat R[`x',1]=_b[T]/_se[T]

restore
}

preserve
clear
svmat R
gen reject = 0
replace reject = 1 if (R1>1.65)
drop if reject==.
sum reject
restore
*=====

* Repetimos la simulación pero para distintos tamaños de muestra y para distintos efec
> tos
*=====
local i=1
mat resultados = J(50,4,.)

foreach efecto in 0.01 0.025 0.05 0.075 0.1{
forvalues size = 1000(1000)10000 {
mat R = J(500,2,.)

forvalues x=1(1)500 {

preserve

sample `size' , count

gen temp = runiform()
gen T=0
replace T = 1 if temp<0.5

replace impuestos_pagados = impuestos_pagados * (1+`efecto') if T==1

reg impuestos_pagados T, robust
mat R[`x',1]=_b[T]/_se[T]

restore
}

preserve
clear
svmat R
gen reject = 0
replace reject = 1 if (R1>1.65)
drop if reject==.
sum reject
scalar media = r(mean)

mat resultados[`i',3] = `efecto'
mat resultados[`i',2] = media
mat resultados[`i',1] = `size'
restore

local i=`i'+1
}
}

clear
svmat resultados

```

```

rename resultados1 sample_size
rename resultados2 st_power
rename resultados3 efecto

replace st_power=round(st_power,.01)
separate st_power, by(efecto)
*=====

* Gráfico
*=====
set scheme slcolor
twoway (connected st_power1 sample_size) (connected st_power2 sample_size) ///
      (connected st_power3 sample_size) (connected st_power4 sample_size) ///
      (connected st_power5 sample_size), ytitle("Power") ///
      xtitle("Number of observations") ///
      legend(label(1 "1%") label(2 "2.5%") label(3 "5%") label(4 "7.5%") label(5 "10%")) ///
      legend(rows(1) title("Effect")) xscale(titlegap(3)) yscale(titlegap(3))
graph export "$main/output/Graph 1.png", replace
*=====

* Generamos una muestra simulada con datos de pagos de impuestos de empresas
* Ahora seteamos la var. del error en 5000
*=====
clear all
set seed 1234 // seteamos semilla para poder replicar los resultados
set obs 15000
gen ganancias_estimadas = rnormal(10000,2000)
drop if ganancias_estimadas<0

gen impuestos_pagados = 0.2*ganancias_estimadas + rnormal(0,5000) // este es el termin
> o de error que hay que modificar en el 2.
drop if impuestos_pagados<0
*=====

* Repetimos la simulación pero para distintos tamaños de muestra y para distintos efec
> tos
*=====
local i=1
mat resultados = J(50,4,.)

foreach efecto in 0.01 0.025 0.05 0.075 0.1{
  forvalues size = 1000(1000)10000 {
    mat R = J(500,2,.)

    forvalues x=1(1)500 {

      preserve

      sample `size' , count

      gen temp = runiform()
      gen T=0
      replace T = 1 if temp<0.5

      replace impuestos_pagados = impuestos_pagados * (1+`efecto') if T==1

      reg impuestos_pagados T, robust

      mat R[`x',1]=_b[T]/_se[T]

      restore
    }
  }
}

```

```

preserve
clear
svmat R
gen reject = 0
replace reject = 1 if (R1>1.65)
drop if reject==.
sum reject
scalar media = r(mean)

mat resultados['i',3] = `efecto'
mat resultados['i',2] = media
mat resultados['i',1] = `size'
restore

local i=`i'+1
}
}

clear
svmat resultados

rename resultados1 sample_size
rename resultados2 st_power
rename resultados3 efecto

replace st_power=round(st_power,.01)
separate st_power, by(efecto)
*=====

* Gráfico
*=====
set scheme slcolor
twoway (connected st_power1 sample_size) (connected st_power2 sample_size) ///
(connectd st_power3 sample_size) (connected st_power4 sample_size) ///
(connectd st_power5 sample_size), ytitle("Power") ///
xtitle("Number of observations") ///
legend(label(1 "1%") label(2 "2.5%") label(3 "5%") label(4 "7.5%") label(5 "10%")) ///
legend(rows(1) title("Effect")) xscale(titlegap(3)) yscale(titlegap(3))
graph export "$main/output/Graph 2.png", replace
*=====

* Generamos una muestra simulada con datos de pagos de impuestos de empresas
* Ahora seteamos la var. del error en 5000
*=====
clear all
set seed 1234 // seteamos semilla para poder replicar los resultados
set obs 15000
gen ganancias_estimadas = rnormal(10000,2000)
drop if ganancias_estimadas<0

gen impuestos_pagados = 0.2*ganancias_estimadas + rnormal(0,5000) // este es el termin
> o de error que hay que modificar en el 2.
drop if impuestos_pagados<0
*=====

* Repetimos la simulación pero para distintos tamaños de muestra y para distintos efec
> tos
* Asignando tratamiento al 20% de las obs.
*=====
local i=1
mat resultados = J(50,4,.)

foreach efecto in 0.01 0.025 0.05 0.075 0.1{
forvalues size = 1000(1000)10000 {
mat R = J(500,2,.)

forvalues x=1(1)500 {

preserve

```

```

sample `size' , count

gen temp = runiform()
gen T=0
replace T = 1 if temp<0.2

replace impuestos_pagados = impuestos_pagados * (1+`efecto') if T==1

reg impuestos_pagados T, robust

mat R[`x',1]=_b[T]/_se[T]

restore
}

preserve
clear
svmat R
gen reject = 0
replace reject = 1 if (R1>1.65)
drop if reject==.
sum reject
scalar media = r(mean)

mat resultados[`i',3] = `efecto'
mat resultados[`i',2] = media
mat resultados[`i',1] = `size'
restore

local i=`i'+1
}
}

clear
svmat resultados

rename resultados1 sample_size
rename resultados2 st_power
rename resultados3 efecto

replace st_power=round(st_power,.01)
separate st_power, by(efecto)
*=====

* Gráfico
*=====
set scheme slcolor
twoway (connected st_power1 sample_size) (connected st_power2 sample_size) ///
(connectd st_power3 sample_size) (connected st_power4 sample_size) ///
(connectd st_power5 sample_size), ytitle("Power") ///
xtitle("Number of observations") ///
legend(label(1 "1%") label(2 "2.5%") label(3 "5%") label(4 "7.5%") label(5 "10%")) ///
legend(rows(1) title("Effect")) xscale(titlegap(3)) yscale(titlegap(3))
graph export "$main/output/Graph 3.png", replace
*=====

* Generamos una muestra simulada con datos de pagos de impuestos de empresas
* Ahora seteamos la var. del error en 5000
*=====
clear all
set seed 1234 // seteamos semilla para poder replicar los resultados
set obs 15000
gen ganancias_estimadas = rnormal(10000,2000)
drop if ganancias_estimadas<0

gen impuestos_pagados = 0.2*ganancias_estimadas + rnormal(0,5000) // este es el termin
> o de error que hay que modificar en el 2.
drop if impuestos_pagados<0
*=====

```



```

* Repetimos la simulación pero para distintos tamaños de muestra y para distintos efec
> tos
* Asignando tratamiento al 80% de las obs.
*=====*
local i=1
mat resultados = J(50,4,.)

foreach efecto in 0.01 0.025 0.05 0.075 0.1{
forvalues size = 1000(1000)10000 {
mat R = J(500,2,.)

forvalues x=1(1)500 {

preserve

sample `size' , count

gen temp = runiform()
gen T=0
replace T = 1 if temp<0.8

replace impuestos_pagados = impuestos_pagados * (1+`efecto') if T==1

reg impuestos_pagados T, robust

mat R[`x',1]=_b[T]/_se[T]

restore
}

preserve
clear
svmat R
gen reject = 0
replace reject = 1 if (R1>1.65)
drop if reject==.
sum reject
scalar media = r(mean)

mat resultados[`i',3] = `efecto'
mat resultados[`i',2] = media
mat resultados[`i',1] = `size'
restore

local i=`i'+1
}
}

clear
svmat resultados

rename resultados1 sample_size
rename resultados2 st_power
rename resultados3 efecto

replace st_power=round(st_power,.01)
separate st_power, by(efecto)
*=====*

```

```

* Gráfico
*=====
set scheme slcolor
twoway (connected st_power1 sample_size) (connected st_power2 sample_size) ///
      (connected st_power3 sample_size) (connected st_power4 sample_size) ///
      (connected st_power5 sample_size), ytitle("Power") ///
xtitle("Number of observations") ///
legend(label(1 "1%") label(2 "2.5%") label(3 "5%") label(4 "7.5%") label(5 "10%")) ///
legend(rows(1) title("Effect")) xscale(titlegap(3)) yscale(titlegap(3))
graph export "$main/output/Graph 4.png", replace
*=====

*Generamos una muestra simulada con datos de pagos de impuestos de empresas
*Agregamos datos de pagos de impuestos de empresas en 2019 a nuestra muestra simulada
*=====
clear all
set seed 1235 // seteamos semilla para poder replicar los resultados
set obs 15000
gen ganancias_estimadas = rnormal(10000,2000)
drop if ganancias_estimadas<0
gen ganancias_estimadas_2019 = rnormal(10000,2000)
drop if ganancias_estimadas<0

gen impuestos_pagados = 0.2*ganancias_estimadas + rnormal(0,500)
drop if impuestos_pagados<0
*=====

* Repetimos simulación de un efecto del 2.5% con 2000 obs
*=====
preserve

sample 2000, count // aleatoriamente me quedo con 2000 obs

* Aleatoriamente asigno el tratamiento al 50% de las obs:
gen temp = runiform()
gen T=0
replace T = 1 if temp<0.5

* A los tratados les aumento el outcome un 2.5%
replace impuestos_pagados = impuestos_pagados * (1+0.025) if T==1

* Regreso el outcome en el tratamiento
reg impuestos_pagados T ganancias_estimadas_2019, robust

restore
*=====

* Repetimos 500 veces y nos fijamos el % de veces que rechazamos H0
*=====
mat R = J(500,2,.)
forvalues x=1(1)500 {
preserve

sample 2000, count

gen temp = runiform()
gen T=0
replace T = 1 if temp<0.5

replace impuestos_pagados = impuestos_pagados * (1+0.025) if T==1

reg impuestos_pagados T ganancias_estimadas_2019, robust

mat R[`x',1]=_b[T]/_se[T]

restore
}

```

```

preserve
clear
svmat R
gen reject = 0
replace reject = 1 if (R1>1.65)
drop if reject==.
sum reject
restore
*=====

* Repetimos la simulación pero para distintos tamaños de muestra y para distintos efec
> tos
*=====
local i=1
mat resultados = J(50,4,.)

foreach efecto in 0.01 0.025 0.05 0.075 0.1{
forvalues size = 1000(1000)10000 {
mat R = J(500,2,.)

forvalues x=1(1)500 {

preserve

sample `size' , count

gen temp = runiform()
gen T=0
replace T = 1 if temp<0.5

replace impuestos_pagados = impuestos_pagados * (1+`efecto') if T==1

reg impuestos_pagados T ganancias_estimadas_2019, robust

mat R[`x',1]=_b[T]/_se[T]

restore
}

preserve
clear
svmat R
gen reject = 0
replace reject = 1 if (R1>1.65)
drop if reject==.
sum reject
scalar media = r(mean)

mat resultados[`i',3] = `efecto'
mat resultados[`i',2] = media
mat resultados[`i',1] = `size'
restore

local i=`i'+1
}
}

clear
svmat resultados

rename resultados1 sample_size
rename resultados2 st_power
rename resultados3 efecto

replace st_power=round(st_power,.01)
separate st_power, by(efecto)
*=====

```

```

* Gráfico
*=====
set scheme slcolor
twoway (connected st_power1 sample_size) (connected st_power2 sample_size) ///
      (connected st_power3 sample_size) (connected st_power4 sample_size) ///
      (connected st_power5 sample_size), ytitle("Power") ///
xtitle("Number of observations") ///
legend(label(1 "1%") label(2 "2.5%") label(3 "5%") label(4 "7.5%") label(5 "10%")) ///
legend(rows(1) title("Effect")) xscale(titlegap(3)) yscale(titlegap(3))
graph export "$main/output/Graph 5.png", replace
*=====

*Exportamos a PDF
*=====
translate "$main/programs/PS 9.do" "$output/PS 9 do-file.pdf", translator(txt2pdf) rep
> lace

```