

Análisis de Sobrevida en Stata

June 3, 2020

Conectamos Stata con SQL mediante ODBC para extraer la base de datos desde el servidor local

```
. odbc load, exec("SELECT * FROM TestFalabella.dbo.SobrevidaUchile") dsn("R_Conexion")
```

Descripción de la información.

Base de datos de 500 registros con 6 columnas (4 covariables)

```
. describe
```

Contains data

```
obs:      500
vars:      6
size:     24,000
```

| variable name | storage type | display format | value label | variable label |
|---------------|--------------|----------------|-------------|----------------|
| tiempo | double | %10.0g | | |
| censura | double | %10.0g | | |
| estadio | double | %10.0g | | |
| tabaco | double | %10.0g | | |
| alcohol | double | %10.0g | | |
| antec | double | %10.0g | | |

Sorted by:

Note: Dataset has changed since last saved.

En la data observamos 79 datos censurados.

```
. tab censura
```

| censura | Freq. | Percent | Cum. |
|---------|-------|---------|--------|
| 0 | 79 | 15.80 | 15.80 |
| 1 | 421 | 84.20 | 100.00 |
| Total | 500 | 100.00 | |

Se observan 3 estadios de la enfermedad siendo el más frecuente el tercer estadio y el menos frecuente el primero, lo que hace suponer que en su mayoría son enfermos terminales.

```
. tab estadio
```

| estadio | Freq. | Percent | Cum. |
|---------|-------|---------|--------|
| 0 | 27 | 5.40 | 5.40 |
| 1 | 200 | 40.00 | 45.40 |
| 2 | 273 | 54.60 | 100.00 |
| Total | 500 | 100.00 | |

Con respecto al hábito del tabaco se observa mucho desbalance, siendo en su gran mayoría fumadores.

```
. tab tabaco
```

| tabaco | Freq. | Percent | Cum. |
|--------|-------|---------|--------|
| 0 | 64 | 12.80 | 12.80 |
| 1 | 436 | 87.20 | 100.00 |
| Total | 500 | 100.00 | |

Con respecto al consumo de alcohol también se observa mucho desbalance observándose mayoritariamente una muestra de la población que no es bebedora

```
. tab alcohol
```

| alcohol | Freq. | Percent | Cum. |
|---------|-------|---------|--------|
| 0 | 473 | 94.60 | 94.60 |
| 1 | 27 | 5.40 | 100.00 |
| Total | 500 | 100.00 | |

El 70% de los pacientes poseen antecedentes familiares de la enfermedad.

```
. tab antec
```

| antec | Freq. | Percent | Cum. |
|-------|-------|---------|--------|
| 0 | 151 | 30.20 | 30.20 |
| 1 | 349 | 69.80 | 100.00 |
| Total | 500 | 100.00 | |

En Stata es necesario declarar la datos como tiempo de supervivencia.

```
. stset tiempo,failure(censura)

      failure event:  censura != 0 & censura < .
obs. time interval:  (0, tiempo]
exit on or before:  failure
```

```
      500  total observations
       0  exclusions
```

```
      500  observations remaining, representing
      421  failures in single-record/single-failure data
18194.791  total analysis time at risk and under observation
              at risk from t =          0
              earliest observed entry t =    0
              last observed exit t =  53.94884
```

En el resumen, del análisis de sobrevida se observa que a los 33 meses estaba vivo el 75% de los pacientes.

```
. stsum
```

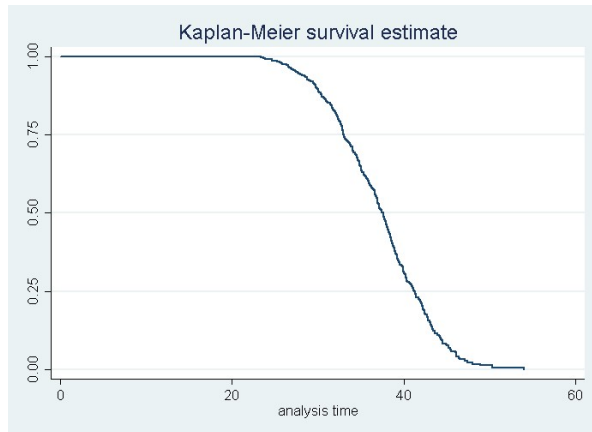
| | failure _d: censura | | | | | |
|-------|--------------------------|----------------|-----------------|---------------|----------|----------|
| | analysis time _t: tiempo | | | | | |
| | time at risk | incidence rate | no. of subjects | Survival time | | |
| | | | | 25% | 50% | 75% |
| total | 18194.79115 | .0231385 | 500 | 33.00543 | 37.57973 | 41.21234 |

```
. sts graph

      failure _d: censura
analysis time _t: tiempo
```

```
. graph export stsgraph.png, replace
(file stsgraph.png written in PNG format)
```

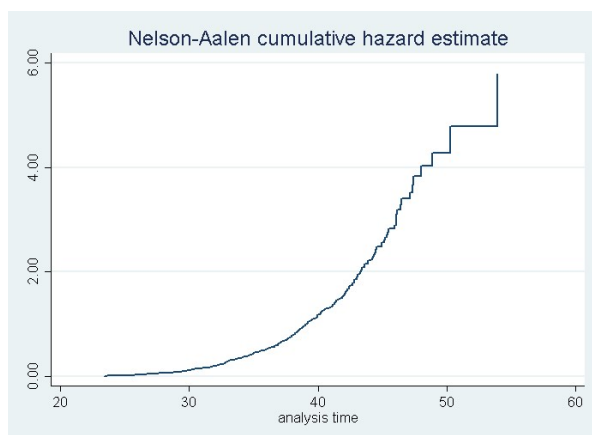
Gráfico de sobrevida de Kaplan-Meier



```
. sts graph, cumhaz  
    failure _d: censura  
    analysis time _t: tiempo
```

```
. graph export stsgraphcum.png, replace  
(file stsgraphcum.png written in PNG format)
```

Gráfico de riesgo acumulado de Nelson Aalen



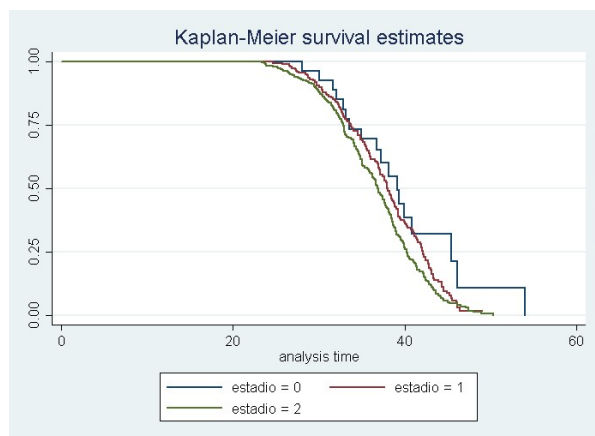
Análisis Sobrevida univariado

Analizaremos las curvas de sobrevida en cada grupo de cada covariable y aplicaremos la d cima de Log Rank para evaluar la existencia de diferencias significativas entre grupos. Por otro lado evaluaremos el supuesto de los riesgos proporcionales para validar la utilizaci n del modelo semiparam trico de Cox.

Estad o:

El estadio 2, que corresponde al estadio m s avanzado de la enfermedad, se proyecta como el m s determinante a la hora de observar el evento de inter s.

```
. sts graph, by(estadio)  
    failure _d: censura  
    analysis time _t: tiempo  
. graph export stsgraphesta.png, replace  
(file stsgraphesta.png written in PNG format)
```



La d cima de Log Rank confirma la existencia de diferencias significativas entre los 3 grupos

```
. sts test estadio
      failure _d: censura
      analysis time _t: tiempo

Log-rank test for equality of survivor functions
```

| estadio | Events observed | Events expected |
|---------|-----------------|-----------------|
| 0 | 18 | 29.25 |
| 1 | 160 | 175.92 |
| 2 | 243 | 215.84 |
| Total | 421 | 421.00 |

```

      chi2(2) = 9.42
      Pr>chi2 = 0.0090

```

Proporcionalidad de los riesgos:

David Kleinbaum en su libro "Survival Analysis" indica que este supuesto puede ser evaluado mediante la interacci n de la covariable con el tiempo, es decir evaluando la significancia de esta correlaci n.

Para la interacci n del estadio con el tiempo, la correlaci n es menor a 5% (coef= -.3624337; $p < 0.05$)

Entonces los riesgos no ser n proporcionales, sin embargo tanto en R como en Python si se demuestra la existencia de proporcionalidad.

```
. gen est_t=estadio*tiempo
. stcox estadio est_t, nohr
      failure _d: censura
      analysis time _t: tiempo

Iteration 0:   log likelihood = -2201.3044
Iteration 1:   log likelihood = -1804.9885
Iteration 2:   log likelihood = -1771.6775
Iteration 3:   log likelihood = -1770.133
Iteration 4:   log likelihood = -1770.1273
Refining estimates:
Iteration 0:   log likelihood = -1770.1273

Cox regression -- no ties

No. of subjects = 500           Number of obs   = 500
No. of failures = 421
Time at risk   = 18194.79115
Log likelihood = -1770.1273     LR chi2(2)      = 862.35
                                Prob > chi2          = 0.0000
```

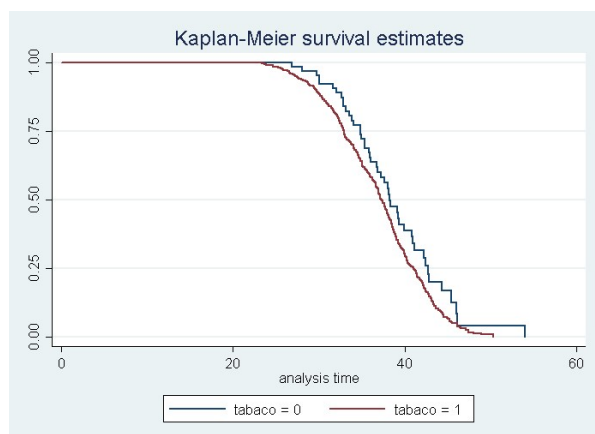
| _t | Coef. | Std. Err. | z | P> z | [95% Conf. Interval] |
|----|-------|-----------|---|------|----------------------|
| | | | | | |

| | | | | | | |
|---------|-----------|----------|--------|-------|-----------|-----------|
| estadio | 14.71005 | .6823831 | 21.56 | 0.000 | 13.37261 | 16.0475 |
| est_t | -.3624337 | .0168738 | -21.48 | 0.000 | -.3955059 | -.3293616 |

Tabaco:

Los pacientes con hábito de tabaquismo parecen ser más riesgosos.

```
. sts graph, by(tabaco)
    failure _d: censura
    analysis time _t: tiempo
. graph export stsgraphtab.png, replace
(file stsgraphtab.png written in PNG format)
```



La d cima de Log Rank confirma un efecto significativo de la covariable sobre la variable de inter s.

```
. sts test tabaco
    failure _d: censura
    analysis time _t: tiempo

Log-rank test for equality of survivor functions
```

| tabaco | Events observed | Events expected |
|--------|--------------------|--------------------|
| 0 | 46 | 60.52 |
| 1 | 375 | 360.48 |
| Total | 421 | 421.00 |

```

chi2(1) = 4.14
Pr>chi2 = 0.0418
```

Proporcionalidad de los riesgos:

Bajo el mismo criterio utilizado en la covariable anterior, los riesgos de muerte en cada categor a no son proporcionales en cada instante de tiempo (coef=-1.062536; $p < 0.05$), aunque en R y Python los test utilizados indican lo contrario.

```
. gen tab_t=tabaco*tiempo
. stcox tabaco tab_t, nohr
    failure _d: censura
    analysis time _t: tiempo

Iteration 0: log likelihood = -2201.3044
Iteration 1: log likelihood = -1755.4298
Iteration 2: log likelihood = -1661.1197
Iteration 3: log likelihood = -1637.8488
Iteration 4: log likelihood = -1635.9645
Iteration 5: log likelihood = -1635.9495
Refining estimates:
Iteration 0: log likelihood = -1635.9495

Cox regression -- no ties
```

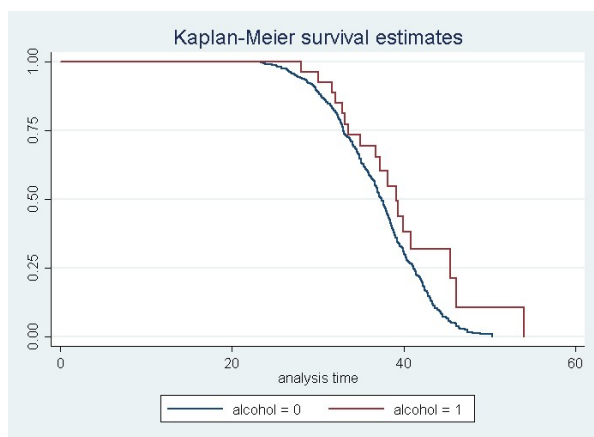
| | | | |
|-------------------|-------------|-----------------|---------|
| No. of subjects = | 500 | Number of obs = | 500 |
| No. of failures = | 421 | | |
| Time at risk = | 18194.79115 | | |
| Log likelihood = | -1635.9495 | LR chi2(2) = | 1130.71 |
| | | Prob > chi2 = | 0.0000 |

| _t | Coef. | Std. Err. | z | P> z | [95% Conf. Interval] |
|--------|-----------|-----------|--------|-------|----------------------|
| tabaco | 45.41826 | 2.299485 | 19.75 | 0.000 | 40.91136 49.92517 |
| tab_t | -1.062536 | .0529256 | -20.08 | 0.000 | -1.166269 -.9588043 |

Alcohol:

Los pacientes con hábito de consumo de alcohol presentan una sobrevida superior.

```
. sts graph, by(alcohol)
      failure _d: censura
      analysis time _t: tiempo
. graph export stsgraphalc.png, replace
(file stsgraphalc.png written in PNG format)
```



La d cima de Log Rank muestra un efecto significativo de la covariable sobre la variable de inter s.

```
. sts test alcohol
      failure _d: censura
      analysis time _t: tiempo

Log-rank test for equality of survivor functions
```

| alcohol | Events observed | Events expected |
|---------|--------------------|--------------------|
| 0 | 403 | 391.75 |
| 1 | 18 | 29.25 |
| Total | 421 | 421.00 |

```

      chi2(1) = 4.88
      Pr>chi2 = 0.0272
```

Proporcionalidad de los riesgos:

Nuevamente los resultados indican falta de proporcionalidad seg n el criterio de Kleinbaum (coef= -.253404 ; $p < 0.05$), sin embargo esto tambi n difiere con los resultados arrojados por los test ejecutados en R y Python.

```
. gen alc_t=alcohol*tiempo
. stcox alcohol alc_t, nohr
      failure _d: censura
```

```

analysis time _t: tiempo
Iteration 0: log likelihood = -2201.3044
Iteration 1: log likelihood = -2189.5174
Iteration 2: log likelihood = -2185.8833
Iteration 3: log likelihood = -2185.3153
Iteration 4: log likelihood = -2185.2969
Iteration 5: log likelihood = -2185.2969
Refining estimates:
Iteration 0: log likelihood = -2185.2969

Cox regression -- no ties

No. of subjects =          500          Number of obs   =          500
No. of failures =          421
Time at risk   = 18194.79115
Log likelihood = -2185.2969          LR chi2(2)        =          32.01
                                      Prob > chi2       =          0.0000

```

| _t | Coef. | Std. Err. | z | P> z | [95% Conf. Interval] |
|---------|----------|-----------|-------|-------|----------------------|
| alcohol | 9.462451 | 2.051501 | 4.61 | 0.000 | 5.441584 13.48332 |
| alc_t | -.253404 | .0557719 | -4.54 | 0.000 | -.3627149 -.1440931 |

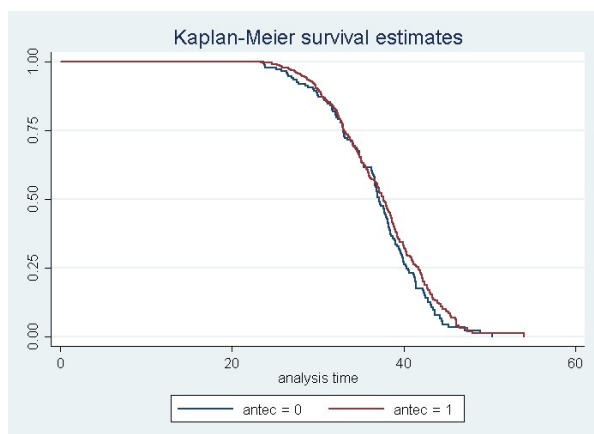
Antecedentes Familiares:

Parece no existir un efecto significativo de esta covariable sobre el evento de interés.

```

. sts graph, by(antec)
    failure _d: censura
    analysis time _t: tiempo
. graph export stsgraphantec.png, replace
(file stsgraphantec.png written in PNG format)

```



La d cima de Log Rank no muestra un efecto significativo de la covariable sobre la variable de inter s.

```

. sts test antec
    failure _d: censura
    analysis time _t: tiempo

Log-rank test for equality of survivor functions

```

| antec | Events observed | Events expected |
|-------|-----------------|-----------------|
| 0 | 131 | 120.13 |
| 1 | 290 | 300.87 |
| Total | 421 | 421.00 |

```

    chi2(1) = 1.38
    Pr>chi2 = 0.2394

```

Proporcionalidad de los riesgos:

Para esta covariable el criterio también difiere de lo observado en R y Python (coef= -0.470457; $p < 0.05$)

```
. gen antec_t=antec*tiempo
. stcox antec antec_t, nohr
      failure _d:  censura
    analysis time _t:  tiempo
Iteration 0:    log likelihood = -2201.3044
Iteration 1:    log likelihood = -1936.9499
Iteration 2:    log likelihood = -1920.4473
Iteration 3:    log likelihood = -1920.0691
Iteration 4:    log likelihood = -1920.0687
Refining estimates:
Iteration 0:    log likelihood = -1920.0687

Cox regression -- no ties

No. of subjects =          500          Number of obs   =          500
No. of failures =          421
Time at risk   = 18194.79115
Log likelihood = -1920.0687          LR chi2(2)         =          562.47
                                      Prob > chi2        =          0.0000
```

| | _t | Coef. | Std. Err. | z | P> z | [95% Conf. Interval] |
|--|---------|----------|-----------|--------|-------|----------------------|
| | antec | 18.79549 | 1.000702 | 18.78 | 0.000 | 16.83415 20.75683 |
| | antec_t | -.470457 | .0247648 | -19.00 | 0.000 | -.518995 -.4219189 |

En resumen, el criterio de David Kleinbaum nos indicaría que para ninguna de las variables categóricas se cumple el supuesto de los riesgos proporcionales, sin embargo los tests entregados por R y Python si indican proporcionalidad de los riesgos.

Lo anterior podría estar relacionado a una imprecisión del criterio ante variables altamente desbalanceadas, o a un error en la interpretación del criterio lo cual aún no investigo en profundidad ya que no es el propósito de este estudio. Sin embargo la decisión es que los riesgos son proporcionales dado los resultados gráficos y estadísticos que entrega R y Python.

Buscaremos un ajuste paramétrico dentro de la familia de las exponenciales.

Los datos distribuyen exponencial?

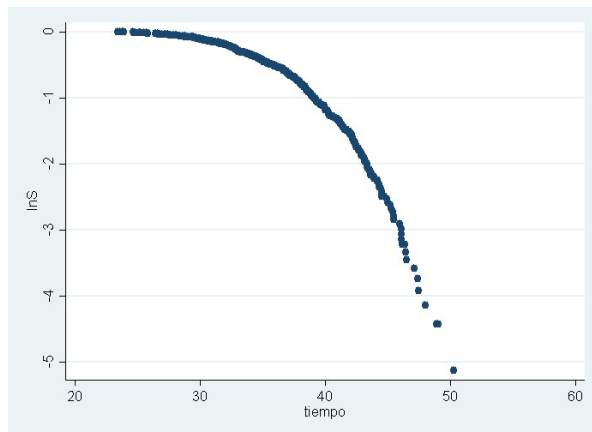
```
. sts gen S=s
```

```
. gen lnS=ln(S)
(1 missing value generated)
```

```
. scatter lnS tiempo
```

```
. graph export scatterlnS_t.png, replace
(file scatterlnS_t.png written in PNG format)
```


Si una colección de datos proviene de una distribución exponencial, existirá una relación lineal entre el logaritmo natural de la sobrevida y el tiempo.



Al evaluar la regresión lineal:

```
. reg lnS tiempo, noconstant
```

| Source | SS | df | MS | Number of obs | = | 499 |
|----------|------------|-----|------------|---------------|---|--------|
| Model | 433.356021 | 1 | 433.356021 | F(1, 498) | = | 873.86 |
| Residual | 246.963056 | 498 | .495909751 | Prob > F | = | 0.0000 |
| Total | 680.319077 | 499 | 1.36336488 | R-squared | = | 0.6370 |
| | | | | Adj R-squared | = | 0.6363 |
| | | | | Root MSE | = | .70421 |

| lnS | Coef. | Std. Err. | t | P> t | [95% Conf. Interval] | |
|--------|-----------|-----------|--------|-------|----------------------|-----------|
| tiempo | -.0253748 | .0008584 | -29.56 | 0.000 | -.0270613 | -.0236883 |

Los datos distribuyen Weibull?

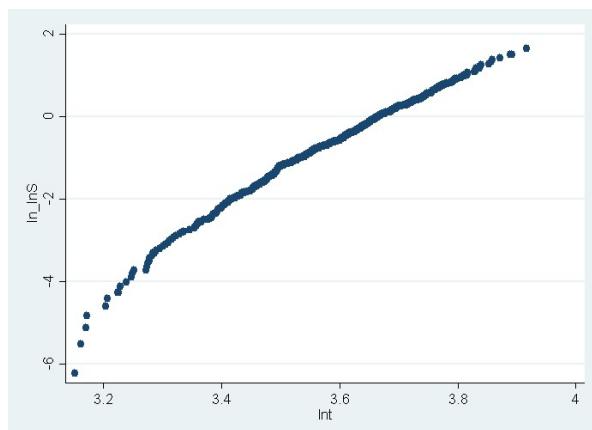
```
. gen ln_lnS=ln(-lnS)
(1 missing value generated)
```

```
. gen lnt=ln(tiempo)
```

```
. scatter ln_lnS lnt
```

```
. graph export scatterln_lnS_lnt.png, replace
(file scatterln_lnS_lnt.png written in PNG format)
```

Si una colección de datos proviene de una distribución Weibull, existirá una relación lineal entre el logaritmo natural del logaritmo natural negativo de la sobrevida y el logaritmo natural del tiempo.



Al evaluar la regresión lineal:

```
. reg ln_lns lnt
```

| Source | SS | df | MS | Number of obs | = | 499 |
|----------|------------|-----|------------|---------------|---|----------|
| Model | 752.098301 | 1 | 752.098301 | F(1, 497) | = | 26038.47 |
| Residual | 14.3554065 | 497 | .028884118 | Prob > F | = | 0.0000 |
| | | | | R-squared | = | 0.9813 |
| Total | 766.453707 | 498 | 1.53906367 | Adj R-squared | = | 0.9812 |
| | | | | Root MSE | = | .16995 |

| ln_lns | Coef. | Std. Err. | t | P> t | [95% Conf. Interval] |
|--------|-----------|-----------|---------|-------|----------------------|
| lnt | 8.323197 | .0515801 | 161.36 | 0.000 | 8.221855 8.424539 |
| _cons | -30.55307 | .1849513 | -165.20 | 0.000 | -30.91646 -30.18969 |

Los datos distribuyen Gompertz?

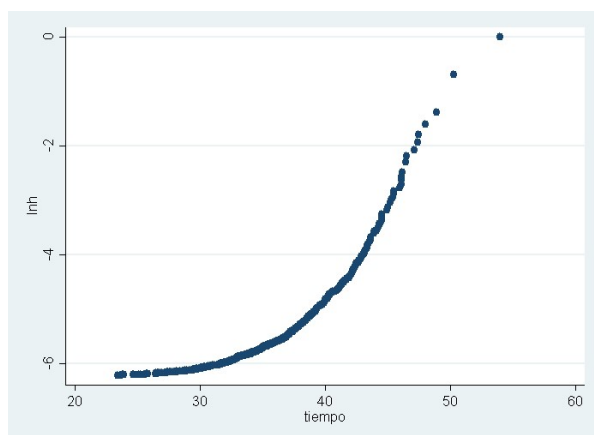
```
. sts gen risk=h
```

```
. gen lnh =ln(risk)
(79 missing values generated)
```

```
. scatter lnh tiempo
```

```
. graph export scatterlnh.png, replace
(file scatterlnh.png written in PNG format)
```

Si una colección de datos proviene de una distribución de Gompertz, existirá una relación lineal entre el logaritmo natural del riesgo (ht) y el tiempo.



Al evaluar la regresión lineal:

```
. reg lnh tiempo, noconstant
```

| Source | SS | df | MS | Number of obs | = | 421 |
|----------|------------|-----|------------|---------------|---|---------|
| Model | 10724.6845 | 1 | 10724.6845 | F(1, 420) | = | 3797.84 |
| Residual | 1186.03468 | 420 | 2.82389209 | Prob > F | = | 0.0000 |
| | | | | R-squared | = | 0.9004 |
| Total | 11910.7192 | 421 | 28.2914945 | Adj R-squared | = | 0.9002 |
| | | | | Root MSE | = | 1.6804 |

| lnh | Coef. | Std. Err. | t | P> t | [95% Conf. Interval] |
|--------|-----------|-----------|--------|-------|----------------------|
| tiempo | -.1374445 | .0022303 | -61.63 | 0.000 | -.1418284 -.1330606 |

Ajustaremos los datos a un modelo paramétrico

creamos las variables dummy de la categoría estadio para este efecto.

```
. tabulate estadio , generate(est)
```

| estadio | Freq. | Percent | Cum. |
|---------|-------|---------|--------|
| 0 | 27 | 5.40 | 5.40 |
| 1 | 200 | 40.00 | 45.40 |
| 2 | 273 | 54.60 | 100.00 |
| Total | 500 | 100.00 | |

Ensayamos el modelo Exponencial

```
. streg tabaco alcohol antec est3, dist(exponential) nolog
      failure _d:  censura
      analysis time _t:  tiempo

Exponential regression -- log relative-hazard form

No. of subjects =          500          Number of obs   =          500
No. of failures =          421
Time at risk   = 18194.79115
Log likelihood = -496.88316          LR chi2(4)         =          3.48
                                      Prob > chi2        =          0.4816
```

| _t | Haz. Ratio | Std. Err. | z | P> z | [95% Conf. Interval] |
|---------|------------|-----------|--------|-------|----------------------|
| tabaco | 1.081071 | .2249305 | 0.37 | 0.708 | .7190367 1.62539 |
| alcohol | .8661425 | .2616694 | -0.48 | 0.634 | .4791083 1.565831 |
| antec | 1.066147 | .1372068 | 0.50 | 0.619 | .8284619 1.372024 |
| est3 | 1.160185 | .149048 | 1.16 | 0.247 | .9019328 1.492384 |
| _cons | .0191529 | .0043791 | -17.30 | 0.000 | .0122353 .0299815 |

Ensayamos el modelo weibull

```
. streg tabaco alcohol antec est3, dist(weibull) nolog
      failure _d:  censura
      analysis time _t:  tiempo

Weibull regression -- log relative-hazard form

No. of subjects =          500          Number of obs   =          500
No. of failures =          421
Time at risk   = 18194.79115
Log likelihood = 133.71324          LR chi2(4)         =          14.21
                                      Prob > chi2        =          0.0067
```

| _t | Haz. Ratio | Std. Err. | z | P> z | [95% Conf. Interval] |
|---------|------------|-----------|--------|-------|----------------------|
| tabaco | 1.014875 | .2112408 | 0.07 | 0.943 | .6749001 1.526109 |
| alcohol | .5870482 | .1782087 | -1.75 | 0.079 | .3237998 1.064317 |
| antec | 1.084536 | .1395738 | 0.63 | 0.528 | .8427501 1.395689 |
| est3 | 1.292027 | .166015 | 1.99 | 0.046 | 1.004382 1.66205 |
| _cons | 5.17e-13 | 5.51e-13 | -26.53 | 0.000 | 6.39e-14 4.18e-12 |
| /ln_p | 2.033966 | .0368925 | 55.13 | 0.000 | 1.961658 2.106274 |
| p | 7.644345 | .2820191 | | | 7.111109 8.217567 |
| 1/p | .1308157 | .0048261 | | | .1216905 .140625 |

Ensayamos el modelo Gompertz

```
. streg tabaco alcohol antec est3, dist(gompertz) nolog
      failure _d:  censura
      analysis time _t:  tiempo

Gompertz regression -- log relative-hazard form

No. of subjects =          500          Number of obs   =          500
No. of failures =          421
Time at risk   = 18194.79115
Log likelihood = 115.57133          LR chi2(4)         =          18.78
                                      Prob > chi2        =          0.0009
```

| _t | Haz. Ratio | Std. Err. | z | P> z | [95% Conf. Interval] |
|---------|------------|-----------|--------|-------|----------------------|
| tabaco | .9969326 | .2075299 | -0.01 | 0.988 | .6629374 1.499198 |
| alcohol | .4961582 | .1514109 | -2.30 | 0.022 | .272811 .9023571 |
| antec | 1.083969 | .1395005 | 0.63 | 0.531 | .8423102 1.394959 |
| est3 | 1.2835 | .1649099 | 1.94 | 0.052 | .9977681 1.651057 |
| _cons | .0000644 | .0000219 | -28.39 | 0.000 | .0000331 .0001254 |

Ensayamos el modelo Loglogistic

```
failure _d: censura
analysis time _t: tiempo
```

| | | | |
|-------------------|-------------|-----------------|--------|
| No. of subjects = | 500 | Number of obs = | 500 |
| No. of failures = | 421 | | |
| Time at risk = | 18194.79115 | | |
| | | LR chi2(4) = | 8.67 |
| Log likelihood = | 129.8512 | Prob > chi2 = | 0.0698 |

| _t | Coef. | Std. Err. | z | P> z | [95% Conf. Interval] | |
|---------|-----------|-----------|--------|-------|----------------------|-----------|
| tabaco | -.017286 | .0291109 | -0.59 | 0.553 | -.0743423 | .0397704 |
| alcohol | .0198935 | .0423577 | 0.47 | 0.639 | -.0631261 | .1029131 |
| antec | -.0187611 | .0193792 | -0.97 | 0.333 | -.0567437 | .0192215 |
| est3 | -.0403017 | .0193104 | -2.09 | 0.037 | -.0781493 | -.002454 |
| _cons | 3.661917 | .0325237 | 112.59 | 0.000 | 3.598172 | 3.725663 |
| /ln_gam | -2.411134 | .0400407 | -60.22 | 0.000 | -2.489612 | -2.332656 |
| gamma | .0897135 | .0035922 | | | .0829421 | .0970377 |

Ensayamos el modelo Lognormal

```
failure _d: censura
analysis time t: tiempo
```

| | | | |
|-------------------|-------------|-----------------|--------|
| No. of subjects = | 500 | Number of obs = | 500 |
| No. of failures = | 421 | | |
| Time at risk = | 18194.79115 | | |
| | | LR chi2(4) = | 9.89 |
| Log likelihood = | 133.39885 | Prob > chi2 = | 0.0423 |

| _t | Coef. | Std. Err. | z | P> z | [95% Conf. Interval] | |
|---------|-----------|-----------|--------|-------|----------------------|-----------|
| tabaco | -.0222277 | .0300123 | -0.74 | 0.459 | -.0810507 | .0365953 |
| alcohol | .0245913 | .0422851 | 0.58 | 0.561 | -.058286 | .1074685 |
| antec | -.013729 | .0194645 | -0.71 | 0.481 | -.0518787 | .0244208 |
| est3 | -.0380358 | .0192504 | -1.98 | 0.048 | -.0757659 | -.0003057 |
| _cons | 3.655301 | .0334457 | 109.29 | 0.000 | 3.589749 | 3.720854 |
| /ln_sig | -1.857104 | .0345766 | -53.71 | 0.000 | -1.924873 | -1.789335 |
| sigma | .1561241 | .0053982 | | | .1458943 | .1670712 |

Ensayamos un modelo gengamma que en primera instancia entrega el mejor ajuste según R (ver archivo Análisis de Sobrevida en R)

```
failure _d: censura
analysis time t: tiempo
```

| | | | |
|-------------------|-------------|-----------------|--------|
| No. of subjects = | 500 | Number of obs = | 500 |
| No. of failures = | 421 | | |
| Time at risk = | 18194.79115 | | |
| | | LR chi2(4) = | 10.91 |
| Log likelihood = | 140.00783 | Prob > chi2 = | 0.0276 |

| _t | Coef. | Std. Err. | z | P> z | [95% Conf. Interval] | |
|---------|-----------|-----------|--------|-------|----------------------|-----------|
| tabaco | -0.111362 | .028834 | -0.39 | 0.694 | -.0678756 | .0451515 |
| alcohol | .0434369 | .0415093 | 1.05 | 0.295 | -.1247936 | .1247936 |
| antec | -.0133162 | .0182258 | -0.73 | 0.465 | -.049038 | .0224057 |
| est3 | -.0372781 | .0181015 | -2.06 | 0.039 | -.0727565 | -.0017998 |
| _cons | 3.680231 | .0323514 | 113.76 | 0.000 | 3.616823 | 3.743638 |
| /ln_sig | -1.938281 | .0433821 | -44.68 | 0.000 | -2.023309 | -1.853254 |
| /kappa | .5031948 | .1367526 | 3.68 | 0.000 | .2351646 | .7712249 |

| | | | | |
|-------|----------|----------|----------|----------|
| sigma | .1439511 | .0062449 | .1322173 | .1567264 |
|-------|----------|----------|----------|----------|

Proceso stepwise

Aplicamos un proceso stepwise para obtener sólo aquellas variables estadísticamente significativas para la respuesta

exponencial

```

. sw, pr(0.1) pe(0.05) forward: streg tabaco alcohol antec est3, dist(exponential) nolog
p >= 0.0500
Exponential regression -- log relative-hazard form

No. of subjects =          500          Number of obs      =          500
No. of failures =          421
Time at risk   = 18194.79115

LR chi2(0)      =          -0.00
Log likelihood  =   -498.621      Prob > chi2       =          .

```

| _t | Haz. Ratio | Std. Err. | z | P> z | [95% Conf. Interval] | |
|-------|------------|-----------|--------|-------|----------------------|----------|
| _cons | .0231385 | .0011277 | -77.28 | 0.000 | .0210305 | .0254578 |

weibull

```

. sw, pr(0.1) pe(0.05) forward: streg tabaco alcohol antec est3, dist(weibull) nolog
                                begin with empty model
p = 0.0041 < 0.0500 adding est3
p = 0.0292 < 0.0500 adding alcohol

weibull regression -- log relative-hazard form

No. of subjects =          500                Number of obs      =          500
No. of failures =          421
Time at risk    = 18194.79115

                                LR chi2(2)      =          13.80
Log likelihood   = 133.51238                  Prob > chi2      =          0.0010

```

| _t | Haz. Ratio | Std. Err. | z | P> z | [95% Conf. Interval] | |
|---------|------------|-----------|--------|-------|----------------------|----------|
| est3 | 1.238936 | .1261509 | 2.10 | 0.035 | 1.014794 | 1.512586 |
| alcohol | .5798959 | .1448883 | -2.18 | 0.029 | .3553656 | .9462908 |
| _cons | 5.65e-13 | 5.93e-13 | -26.87 | 0.000 | 7.22e-14 | 4.42e-12 |
| /ln_p | 2.034103 | .0368802 | 55.15 | 0.000 | 1.961819 | 2.106387 |
| p | 7.645393 | .2819635 | | | 7.112256 | 8.218495 |
| 1/p | .1307977 | .0048238 | | | .1216768 | .1406024 |

gomPERTZ

```

. sw, pr(0.1) pe(0.05) forward: streg tabaco alcohol antec est3, dist(gompertz) nolog
               begin with empty model
p = 0.0008 < 0.0500 adding alcohol
p = 0.0443 < 0.0500 adding est3

Gompertz regression -- log relative-hazard form

No. of subjects =          500          Number of obs   =          500
No. of failures =          421
Time at risk   = 18194.79115

Log likelihood =    115.37543          LR chi2(2)       =         18.39
                                      Prob > chi2      =         0.0001

```

| _t | Haz. Ratio | Std. Err. | z | P> z | [95% Conf. Interval] | |
|---------|------------|-----------|--------|-------|----------------------|----------|
| alcohol | .4974441 | .1250736 | -2.78 | 0.005 | .3038957 | .8142616 |
| est3 | 1.227255 | .1249497 | 2.01 | 0.044 | 1.005246 | 1.498296 |
| _cons | .0000697 | .0000187 | -35.65 | 0.000 | .0000412 | .0001179 |
| /gamma | .1974663 | .0070934 | 27.84 | 0.000 | .1835636 | .2113691 |

llogistic

```
. sw, pr(0.1) pe(0.05) forward: streg tabaco alcohol antec est3, dist(llogistic) nolog
begin with empty model
p = 0.0110 < 0.0500 adding est3
```

Loglogistic regression -- accelerated failure-time form

```
No. of subjects =      500      Number of obs   =      500
No. of failures =      421
Time at risk    = 18194.79115
Log likelihood   = 128.74213      LR chi2(1)     =      6.45
                                      Prob > chi2    =      0.0111
```

| _t | Coef. | Std. Err. | z | P> z | [95% Conf. Interval] | |
|---------|-----------|-----------|--------|-------|----------------------|-----------|
| est3 | -.0370046 | .0145557 | -2.54 | 0.011 | -.0655332 | -.0084759 |
| _cons | 3.632941 | .0108949 | 333.45 | 0.000 | 3.611588 | 3.654295 |
| /ln_gam | -2.409832 | .0400429 | -60.18 | 0.000 | -2.488315 | -2.331349 |
| gamma | .0898304 | .0035971 | | | .0830498 | .0971646 |

lnormal

```
. sw, pr(0.1) pe(0.05) forward: streg tabaco alcohol antec est3, dist(lnormal) nolog
begin with empty model
p = 0.0069 < 0.0500 adding est3
```

Lognormal regression -- accelerated failure-time form

```
No. of subjects =      500      Number of obs   =      500
No. of failures =      421
Time at risk    = 18194.79115
Log likelihood   = 132.09753      LR chi2(1)     =      7.29
                                      Prob > chi2    =      0.0069
```

| _t | Coef. | Std. Err. | z | P> z | [95% Conf. Interval] | |
|---------|-----------|-----------|--------|-------|----------------------|-----------|
| est3 | -.0394154 | .0145864 | -2.70 | 0.007 | -.0680043 | -.0108266 |
| _cons | 3.628279 | .0109215 | 332.22 | 0.000 | 3.606873 | 3.649685 |
| /ln_sig | -1.85547 | .0345906 | -53.64 | 0.000 | -1.923267 | -1.787674 |
| sigma | .1563794 | .0054093 | | | .1461289 | .167349 |

ggamma

```
. sw, pr(0.1) pe(0.05) forward: streg tabaco alcohol antec est3, dist(ggamma) nolog
begin with empty model
p = 0.0058 < 0.0500 adding est3
```

Generalized gamma regression -- accelerated failure-time form

```
No. of subjects =      500      Number of obs   =      500
No. of failures =      421
Time at risk    = 18194.79115
Log likelihood   = 138.37931      LR chi2(1)     =      7.65
                                      Prob > chi2    =      0.0057
```

| _t | Coef. | Std. Err. | z | P> z | [95% Conf. Interval] | |
|---------|-----------|-----------|--------|-------|----------------------|-----------|
| est3 | -.0382762 | .0138644 | -2.76 | 0.006 | -.0654499 | -.0111026 |
| _cons | 3.662128 | .0139156 | 263.17 | 0.000 | 3.634854 | 3.689402 |
| /ln_sig | -1.931398 | .0423689 | -45.59 | 0.000 | -2.014439 | -1.848356 |
| /kappa | .4795642 | .1324952 | 3.62 | 0.000 | .2198784 | .7392501 |
| sigma | .1449455 | .0061412 | | | .1333952 | .1574959 |

Ensayaremos el modelo semiparamétrico de riesgos proporcionales de Cox

```
. stcox tabaco alcohol antec est3, nolog
```

```
failure _d: censura
analysis time _t: tiempo
```

Cox regression -- no ties

```
No. of subjects =      500      Number of obs   =      500
No. of failures =      421
```

| Time at risk | = | 18194.79115 | LR chi2(4) | = | 10.39 |
|----------------|------------|-------------|-------------|-------|----------------------|
| Log likelihood | = | -2196.1091 | Prob > chi2 | = | 0.0343 |
| _t | Haz. Ratio | Std. Err. | z | P> z | [95% Conf. Interval] |
| tabaco | 1.017542 | .2120198 | 0.08 | 0.933 | .6763819 1.530779 |
| alcohol | .66734 | .2064772 | -1.31 | 0.191 | .3639005 1.223803 |
| antec | 1.073881 | .138634 | 0.55 | 0.581 | .8338146 1.383067 |
| est3 | 1.283994 | .1653833 | 1.94 | 0.052 | .9975275 1.652726 |

Con stepwise

| . sw, pr(0.1) pe(0.05) forward: stcox tabaco alcohol antec est3, nolog | | | | | | |
|--|------------|-----------------|------|--------|----------------------|----------|
| begin with empty model | | | | | | |
| p = 0.0080 < 0.0500 | | adding est3 | | | | |
| Cox regression -- no ties | | | | | | |
| No. of subjects = 500 | | Number of obs = | | 500 | | |
| No. of failures = 421 | | | | | | |
| Time at risk = 18194.79115 | | | | | | |
| Log likelihood = -2197.7503 | | LR chi2(1) = | | 7.11 | | |
| | | Prob > chi2 = | | 0.0077 | | |
| _t | Haz. Ratio | Std. Err. | z | P> z | [95% Conf. Interval] | |
| est3 | 1.300464 | .128842 | 2.65 | 0.008 | 1.070943 | 1.579174 |

Los resultados muestran similitudes entre el ajuste paramétrico de los modelos Exponencial Weibull y Gompertz los cuales consideran el estadio 3 como una variable de riesgo que se mueve entre un 24% y 30%, y sugiere al consumo de alcohol como variable de protección (coeficiente negativo).

Por otro lado los modelos loglogistic, lognormal y gengama muestran un grado de incongruencia al comparar con los resultados de sobrevida univariados.

Al utilizar el proceso stepwise los modelos weibull y gompertz son los que matienen ambas covariables, sin embargo al realizar un stepwise sobre el modelo de riesgos proporcionales de Cox, este sugiere sólo a la covariable estadio como importante dentro de un eventual modelo.

En términos generales los pacientes fumadores presentan mayor riesgo de ocurrencia del evento (menor tiempo de sobrevida), aunque este resultado no es significativo.

Finalmente considerar la variable alcohol debería ser una descisión fuertemente apoyada por un experto en enfermedades. Una conclusión que podría explicar en parte la desición de incorporar esta covariable podría ser el hecho de que los pacientes que consumen alcohol presentan un 42% de menor riesgo debido a un eventual fortalecimiento del sistema inmunológico pensando que tal vez el alcohol ayuda a los pacientes a sobrellevar la enfermedad terminal de una manera más relajada, aunque todo esto no es más que un supuesto y podría simplemente estar relacionado al elevado desvalance existente en las covariables. Como sea se hace necesaria la opinión de un experto en la enfermedad ya que estos son sólo los resultados estadísticos.