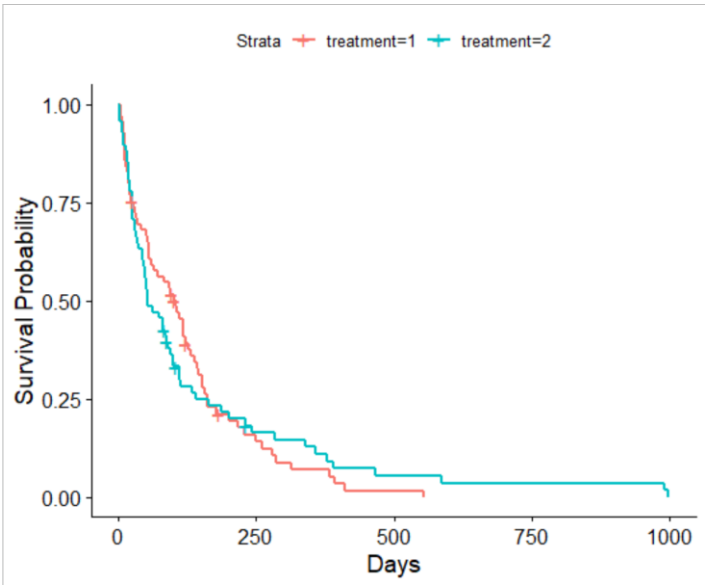
SURVIVAL ANALYSIS PROJECT_AR

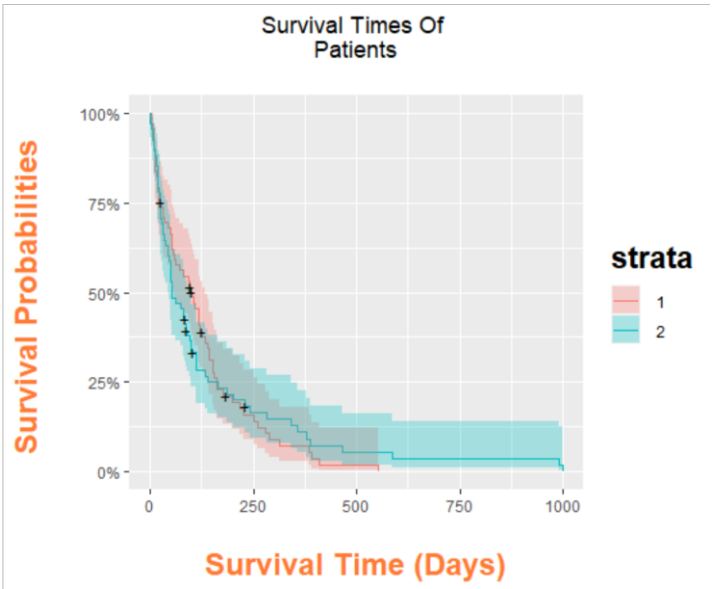
This project is to understand the effect of chemotherapy (standard treatment) vs chemotherapy combined with a new drug (test treatment) on the Veteran's Administration Lung Cancer Trial data which consists of data of 137 patients with advanced, inoperable lung cancer who underwent this treatment.

Source: Kalbfleisch and Prentice (pp. 223-224)

 # DATA DICTIONARY- Variables:
Treatment 1=standard, 2=test
Cell type 1=squamous, 2=small cell, 3=adeno, 4=large
Survival in days
Status 1=dead, 0=censored
Karnofsky score (measure of general performance, 100=best)
Months from Diagnosis
Age in years
Prior chemotherapy 0=no, 10=yes

Kaplan-Meier Survival Graphs





From the Kaplan-Meier survival graphs, we can draw the following inferences-

- Test Treatment results in longer survival days overall in the long run for time greater than 250 days compared to standard treatment.
- **Days 0 to 25 -** No difference in survival days for test treatment i.e., treatment 2+ chemo for first 25 days.
- **Days 45 to 150 -** The standard treatment is better than the test treatment with higher survival probability from around 45 days to 180 days.
- **Days 150 to 250 -** Between 180 to 250 days, both treatments have almost similar survival probability.
- **Days 250+ -** Post 250 days, test treatment has higher survival probability.
- This could mean that it takes about **6 months to start seeing the effects of the new drug.**
- Also, the confidence levels with respect to the spread (in blue and red on the left graph) are a little higher in area, which implies that test treatment has more confidence on the survival probability than standard treatment.

HALF-YEARLY AND YEARLY SURVIVAL DAYS ESTIMATE

```
> summary(km, times=365)
Call: survfit(formula = Surv(d$survivaldays, d$status) ~ d$treatment)

                d$treatment=1
      time       n.risk      n.event     survival      std.err lower 95% CI upper 95% CI
   365.0000       4.0000     59.0000       0.0719       0.0341       0.0284       0.1821

                d$treatment=2
      time       n.risk      n.event     survival      std.err lower 95% CI upper 95% CI
   365.0000       6.0000     58.0000       0.1098       0.0407       0.0530       0.2272

> summary(km, times=183)
Call: survfit(formula = Surv(d$survivaldays, d$status) ~ d$treatment)

                d$treatment=1
      time       n.risk      n.event     survival      std.err lower 95% CI upper 95% CI
   183.0000      12.0000     51.0000       0.2156       0.0521       0.1343       0.3462

                d$treatment=2
      time       n.risk      n.event     survival      std.err lower 95% CI upper 95% CI
   183.0000      14.0000     51.0000       0.2329       0.0529       0.1492       0.3634
```

Survival Probability	6 MONTHS (183 DAYS)	1 YEAR (365 DAYS)
STANDARD TREATMENT	21.56%	7.19%
TEST TREATMENT	23.29%	10.98%

PREDICTOR TABLE

(Choosing Variables for modeling)

Y Variables: (survivaldays,status)		
PREDICTOR VARIABLES:	**Effect Direction**	**Rationale**
treatment	?	We need to test Y variable against effectiveness of tests vs standard treatment. Required for analysis.
cell_type	"+/-"	Included as a factor variable to understand if a certain cell type has less survival probability.
months	+	Longer Diagnosis usually means better chance of longer survival.
age	-	Important variable to understand how increase in age affects survival days.
priorchem	+	Required to understand effectiveness of chemotherapy treatment on survival.
EXCLUDED VARIABLES		
kscore	?	Inclusion of Karnofsky score for analysis is subject to debate as per research and we do not know at what stage i.e., in which month/ post how many days of observation this score was calculated for a patient and is hence, excluded from analysis.Also,it measures the ability of cancer patients to perform ordinary tasks and it is unclear how it may be related to treatment. Karnofsky Score Reference - https://ascopubs.org/doi/10.1200/jco.1984.2.3.187 & https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7018389/

MEDIAN PATIENT SURVIVAL TIME BASED ON TREATMENT

Median survival time for patients in standard treatment is 103 days whereas on test treatment it is 52 days i.e., standard treatment survival time is double that of test treatment.

```
> print(km)
Call: survfit(formula = y ~ treatment, data = d)

             n events median 0.95LCL 0.95UCL
treatment=1 69     64  103.0      59     132
treatment=2 68     64   52.5      44      95
```

Survival Models used for this analysis-

```
#Analysis with selected variables
y <- Surv(d$survivaldays,d$status)
cox2 <- coxph(y ~ treatment + cell_type +  months+ age + priorchem)
exp2 <- survreg(y ~ treatment + cell_type +  months+ age + priorchem, dist="exponential")
weibull2 <- survreg(y ~ treatment + cell_type +  months+ age + priorchem, dist="weibull")
loglogistic2 <- survreg(y ~ treatment + cell_type +  months+ age + priorchem, dist="loglogistic")

library(stargazer)
stargazer(cox2, exp2, weibull2, loglogistic2, type="text",single.row=TRUE)
```

Model outputs and Interpretations below.

Stargazer output of Model's beta coefficients

```
==========================================================================================
                                              Dependent variable:
                     ---------------------------------------------------------------------
                                                        y
                            Cox              exponential          Weibull      survreg: loglogistic
                        prop. hazards
                             (1)                 (2)                (3)                (4)
------------------------------------------------------------------------------------------
treatment2             0.167 (0.202)      -0.136 (0.199)     -0.135 (0.202)      -0.210 (0.225)
cell_type2            1.040*** (0.278)  -1.081*** (0.264) -1.081*** (0.268)  -1.076*** (0.298)
cell_type3            1.188*** (0.299)  -1.233*** (0.274) -1.230*** (0.279)  -0.939*** (0.339)
cell_type4             0.290 (0.285)      -0.308 (0.276)     -0.303 (0.282)       0.091 (0.327)
months                 0.010 (0.009)      -0.010 (0.009)     -0.010 (0.009)      -0.010 (0.013)
age                    0.005 (0.010)      -0.007 (0.010)     -0.007 (0.010)       0.001 (0.011)
priorchem10           -0.069 (0.232)       0.095 (0.228)      0.092 (0.232)      -0.195 (0.269)
Constant                                  5.963*** (0.579)  5.952*** (0.591)    4.934*** (0.706)
------------------------------------------------------------------------------------------
Observations                137                 137                137                137
R2                         0.180
Max. Possible R2           0.999
Log Likelihood            -491.890            -733.085           -733.046           -738.397
chi2 (df = 7)                                 36.272***          30.091***          23.737***
Wald Test              26.510*** (df = 7)
LR Test                27.117*** (df = 7)
Score (Logrank) Test 28.073*** (df = 7)
==========================================================================================
Note:                                                          *p<0.1; **p<0.05; ***p<0.01
```

Cox proportional hazard model estimates hazard ratio whereas Exponential and Weibull models give time to event. For ease of interpretation, we can choose Cox PH Model. (Light Red implies increase of death risk and green implies decrease of death risk).

Predictor	β	exp(β)	Interpretation
Treatment	0.167	1.180	The new drug treatment ***increases* risk of patient death by 18.1 %** compared to patients who do not receive this drug, when all other variables are invariant.
Cancer Type 4 (adeno)	0.290	1.336	Large cell cancer has an **33.6% higher risk of death** compared to squamous cell cancer, when all other variables are invariant.
Cancer Type 3 (adeno)	1.188	3.280	Adeno cell cancer has an **228% higher risk of death** compared to squamous cell cancer, when all other variables are invariant.
Cancer Type 2 (small cell)	1.040	2.829	Small cell cancer has an **183% higher risk of death** compared to squamous cell cancer, when all other variables are invariant.
Age	0.005	1.003	With each year of increase in age, risk of death increases by 0.5%, when all other variables are invariant.
Months from diagnosis	0.010	1.010	With each additional month on diagnosis, risk of death increases by 1% when all other variables are invariant.
Previous chemo	-0.069	0.933	**Previous chemotherapy reduces risks of death by 6.67%** compared to patients with no prior chemo with all other variables invariant.

In summary, for our samples, we can draw the below conclusions based on the coefficients-

- Test Treatment had 18.1% higher hazard rate than standard treatment.
- Compared to Squamous cell type, Small Cell, Adeno, and Large cell types had 183%,228% and 33.6% higher risk of death respectively.
- A 1-month addition of diagnosis time increases hazard rate by 1%.
- 1 year addition in age increases death risk by 0.5%.
- Prior Chemotherapy reduces death risk by 6.7%.

Model with interaction effects of treatment against cell type-

```
cox <- coxph(y ~ treatment*cell_type + treatment*age + treatment*months
             + treatment*priorchem, data=d, method="breslow")

exp <- survreg(y ~ treatment*cell_type + treatment*age + treatment*months
               + treatment*priorchem, data=d, dist="exponential")

weibull <- survreg(y ~ treatment*cell_type + treatment*age + treatment*months
                   + treatment*priorchem, data=d, dist="weibull")

loglogistic <- survreg(y ~ treatment*cell_type + treatment*age + treatment*months
                       + treatment*priorchem, data=d, dist="loglogistic")

stargazer(cox, exp, weibull, type="text", single.row=TRUE)
```

	Dependent variable:		
	y		
	Cox prop. hazards	exponential	Weibull
	(1)	(2)	(3)
treatment2	0.338 (1.257)	-0.475 (1.252)	-0.479 (1.209)
cell_type2	0.448 (0.346)	-0.467 (0.343)	-0.458 (0.331)
cell_type3	0.845* (0.466)	-0.822* (0.460)	-0.815* (0.443)
cell_type4	-0.325 (0.398)	0.314 (0.394)	0.308 (0.379)
age	0.003 (0.013)	-0.005 (0.012)	-0.005 (0.012)
months	0.006 (0.019)	-0.003 (0.019)	-0.002 (0.018)
priorchem10	0.266 (0.308)	-0.311 (0.306)	-0.302 (0.296)
treatment2:cell_type2	1.275** (0.529)	-1.120** (0.500)	-1.135** (0.482)
treatment2:cell_type3	0.676 (0.590)	-0.590 (0.576)	-0.603 (0.555)
treatment2:cell_type4	1.279** (0.574)	-1.157** (0.549)	-1.154** (0.528)
treatment2:age	-0.014 (0.020)	0.015 (0.020)	0.015 (0.019)
treatment2:months	0.008 (0.022)	-0.012 (0.022)	-0.012 (0.021)
treatment2:priorchem10	-0.916* (0.496)	0.949** (0.479)	0.942** (0.462)
Constant		5.375*** (0.790)	5.385*** (0.759)
Observations	137	137	137
R2	0.243		
Max. Possible R2	0.999		
Log Likelihood	-486.809	-727.298	-727.148
chi2 (df = 13)		47.847***	41.887***
Wald Test	35.950*** (df = 13)		
LR Test	38.150*** (df = 13)		
Score (Logrank) Test	39.063*** (df = 13)		
Note:			*p<0.1; **p<0.05; ***p<0.01

EQUATION = dy/dx = d(hazard risk)/d(treatment2) = 0.338+1.275(celltype2)+0.676(celltype3)+1.279(celltype4)-0.014(age)+0.008(months)-0.916(priorchem).

FINAL INTERPRETATION & RECOMMENDATION

- The new drug treatment **increases death risk of patient by 33.8%** for patients with celltype1(squamous), **161.3%** for celltype2(small cell),**101.4%** for celltype3(adeno) **and 161.7%** for patients with celltype4(large cell).
- This **death risk drops by 91.6% for patients with prior chemotherapy**.
- Therefore, this treatment can be suggested only to patients with prior chemotherapy who have celltype1(squamous). For this set of patients, **death risk will go down by 57.8%.**