Survival Analysis: Breast Cancer

TABLE OF CONTENTS

Introduction

German Breast Cancer Dataset has 15 variables and 686 observations.

- **Willi Sauerbrei:** A senior statistician and a professor of medical biometry at the University Medical Center Freiburg
- **Patrick Royston:** A professor of Statistics in the Department of Statistical Science at University College London

720 patients with primary node positive breast cancer were recruited from July 1984 to December 1989

Introduction

Variables:

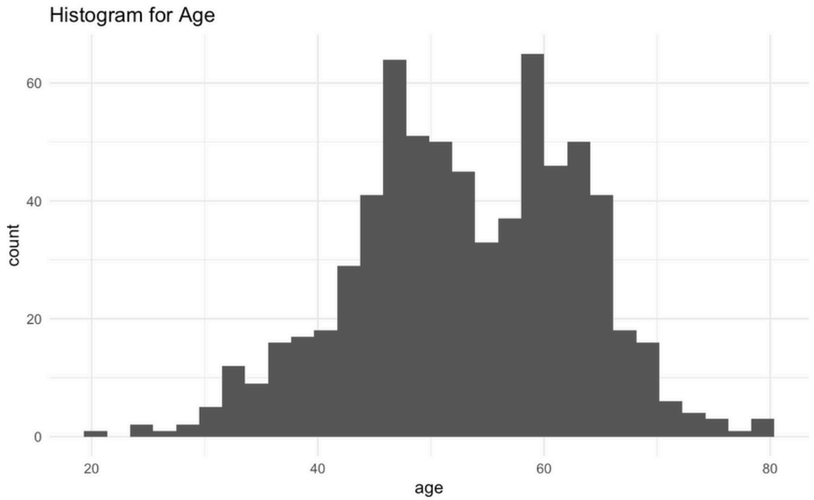
- **diagdateb:** date the patient was diagnosed with cancer
- **recdate:** date of recurrence
- **deathdate:** date of death
- **censdead:** death/censoring indicator (death=1, alive=0)
- **age:** age of the patient in years
- **rectime:** time to recurrence/censoring, whichever occurs first
- **size:** tumor size (in mm)
- **grade:** tumor grade, a ordered factor at levels I < II < III
- **nodes:** number of positive nodes
- **censrec:** recurrence/censoring indicator (recurrence=1, alive=0)
- **survtime:** recurrence free survival time (in days)
- **menopause:** menstrual period stopped or not
- **hormone:** received hormone therapy (yes = 1, no = 0)
- **prog_recp:** progesterone receptor (in fmol)
- **estrg_recp:** estrogen receptor (in fmol)

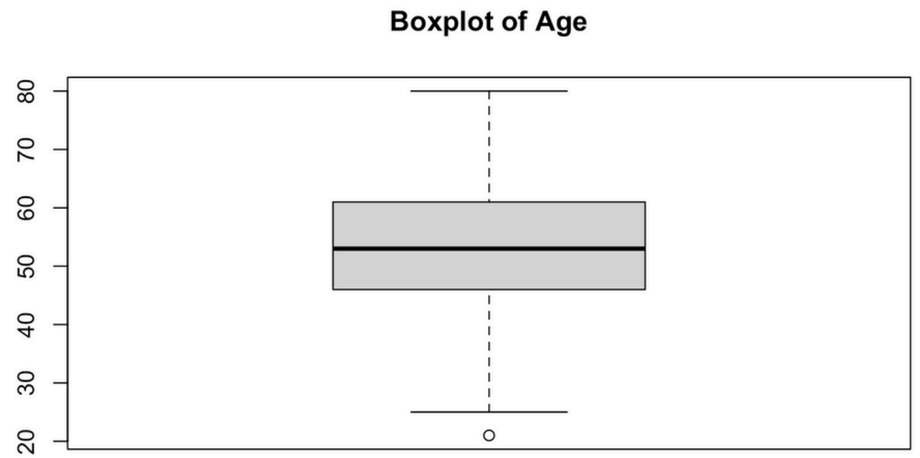
Introduction

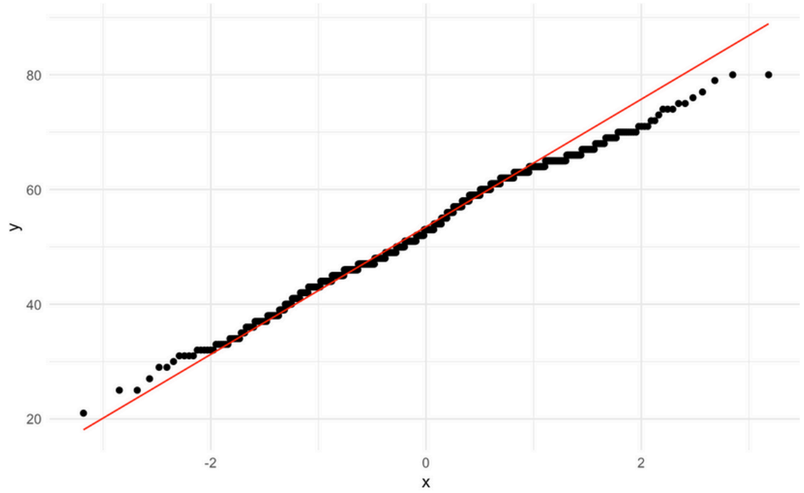
```
tibble [686 × 20] (S3: tbl_df/tbl/data.frame)
 $ diagdateb               : POSIXct[1:686], format: "1984-08-17" "1985-04-25" "1984-10-11" "1984-06-29" ...
 $ recdate                 : POSIXct[1:686], format: "1988-04-15" "1989-03-15" "1988-04-12" "1984-11-24" ...
 $ deathdate               : POSIXct[1:686], format: "1990-11-16" "1990-10-22" "1988-10-06" "1984-11-24" ...
 $ age                     : num [1:686] 38 52 47 40 64 49 53 61 43 74 ...
 $ age_group               : Factor w/ 3 levels "1","2","3": 2 3 2 2 3 2 3 3 2 3 ...
 $ received_hormone_therapy: Factor w/ 2 levels "1","2": 1 1 1 1 2 2 1 2 1 2 ...
 $ menopause_cat           : Factor w/ 2 levels "1","2": 1 1 1 1 2 2 2 2 1 2 ...
 $ size                    : num [1:686] 18 20 30 24 19 56 52 22 30 20 ...
 $ tumor_grade             : Ord.factor w/ 3 levels "1"<"2"<"3": 3 1 2 1 2 1 2 2 2 2 ...
 $ nodes                   : num [1:686] 5 1 1 3 1 3 9 2 1 1 ...
 $ prog_recp               : num [1:686] 141 78 422 25 19 356 6 6 22 462 ...
 $ estrg_recp              : num [1:686] 105 14 89 11 9 64 29 173 0 240 ...
 $ rectime                 : num [1:686] 1337 1420 1279 148 1863 ...
 $ censrec                 : num [1:686] 1 1 1 0 0 0 1 1 0 0 ...
 $ survival_time           : num [1:686] 2282 2006 1456 148 1863 ...
 $ event_status            : num [1:686] 0 0 1 0 0 0 1 0 0 0 ...
 $ sqrt_survival           : num [1:686] 47.8 44.8 38.2 12.2 43.2 ...
 $ log_size                : num [1:686] 2.89 3 3.4 3.18 2.94 ...
 $ log_prog_recp           : num [1:686] 4.96 4.37 6.05 3.26 3 ...
 $ log_estrg_recp          : num [1:686] 4.66 2.71 4.5 2.48 2.3 ...
```

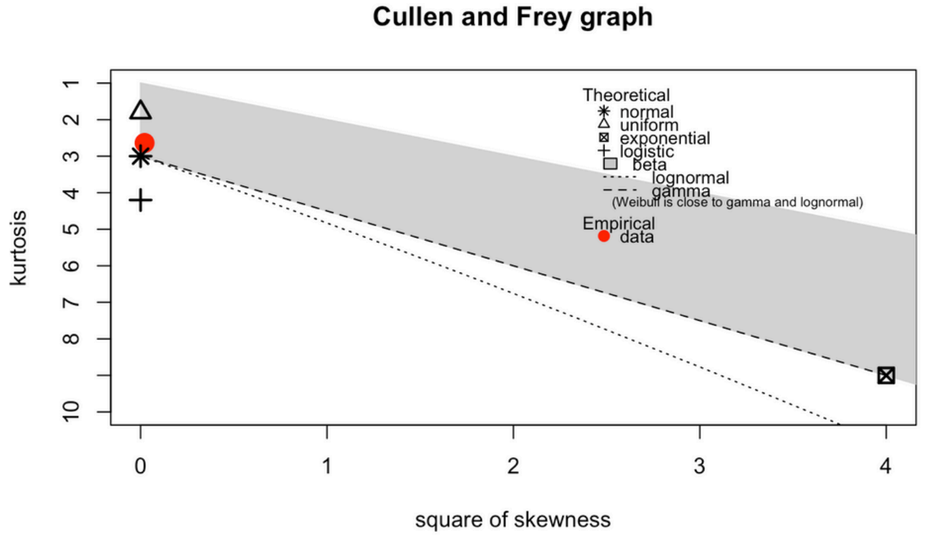
```
  diagdateb                      recdate                       deathdate
Min.   :1984-04-25 00:00:00.000   Min.   :1984-11-24 00:00:00.000   Min.   :1984-11-24 00:00:00.000
1st Qu.:1985-09-10 06:00:00.000   1st Qu.:1988-05-06 06:00:00.000   1st Qu.:1989-06-01 00:00:00.000
Median :1986-08-03 12:00:00.000   Median :1990-04-23 00:00:00.000   Median :1990-11-25 12:00:00.000
Mean   :1986-10-07 21:37:15.569   Mean   :1989-11-05 09:22:33.935   Mean   :1990-05-20 12:18:53.527
3rd Qu.:1987-10-27 18:00:00.000   3rd Qu.:1991-07-10 06:00:00.000   3rd Qu.:1991-09-17 18:00:00.000
Max.   :1989-12-05 00:00:00.000   Max.   :1992-03-25 00:00:00.000   Max.   :1992-06-15 00:00:00.000
     age        age_group received_hormone_therapy menopause_cat      size        tumor_grade     nodes
Min.   :21.00   1:  7     1:440                    1:290          Min.   :  3.00   1: 81       Min.   : 1.00
1st Qu.:46.00   2:282     2:246                    2:396          1st Qu.: 20.00   2:444       1st Qu.: 1.00
Median :53.00   3:397                                             Median : 25.00   3:161       Median : 3.00
Mean   :53.05                                                     Mean   : 29.33               Mean   : 5.01
3rd Qu.:61.00                                                     3rd Qu.: 35.00               3rd Qu.: 7.00
Max.   :80.00                                                     Max.   :120.00               Max.   :51.00
  prog_recp        estrg_recp         rectime         censrec       survival_time     event_status
Min.   :   0.0   Min.   :   0.00   Min.   :   8.0   Min.   :0.0000   Min.   :   8.0   Min.   :0.0000
1st Qu.:   7.0   1st Qu.:   8.00   1st Qu.: 567.8   1st Qu.:0.0000   1st Qu.: 798.8   1st Qu.:0.0000
Median :  32.5   Median :  36.00   Median :1084.0   Median :0.0000   Median :1338.0   Median :0.0000
Mean   : 110.0   Mean   :  96.25   Mean   :1124.5   Mean   :0.4359   Mean   :1320.6   Mean   :0.2493
3rd Qu.: 131.8   3rd Qu.: 114.00   3rd Qu.:1684.8   3rd Qu.:1.0000   3rd Qu.:1824.8   3rd Qu.:0.0000
Max.   :2380.0   Max.   :1144.00   Max.   :2659.0   Max.   :1.0000   Max.   :2668.0   Max.   :1.0000
```

Exploratory Data Analysis

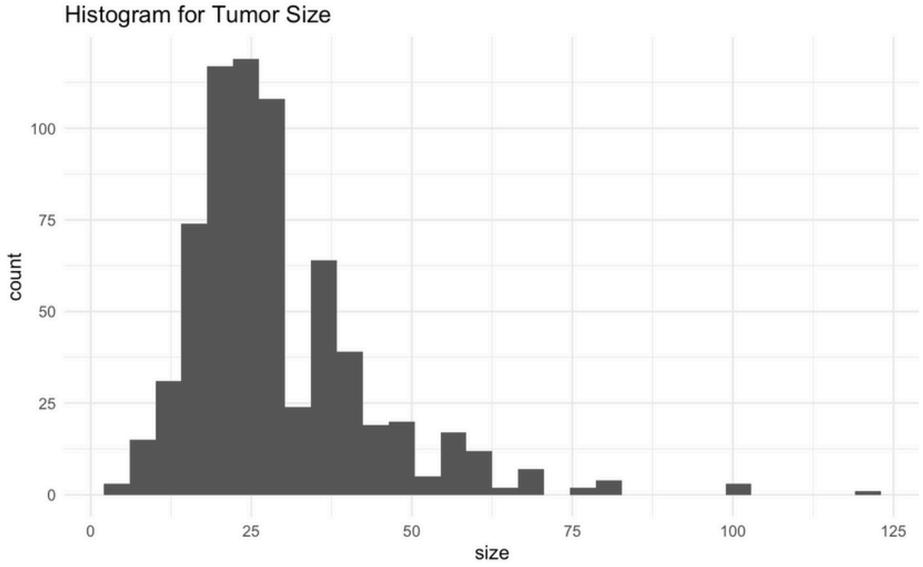


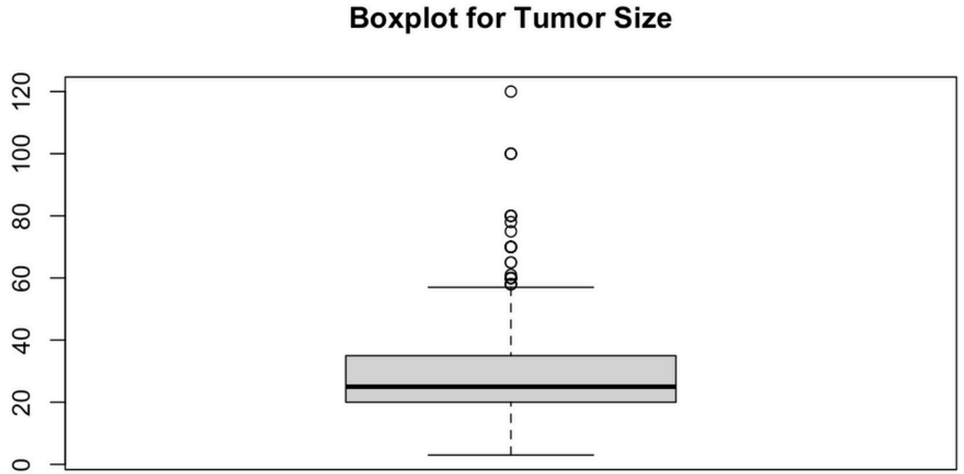


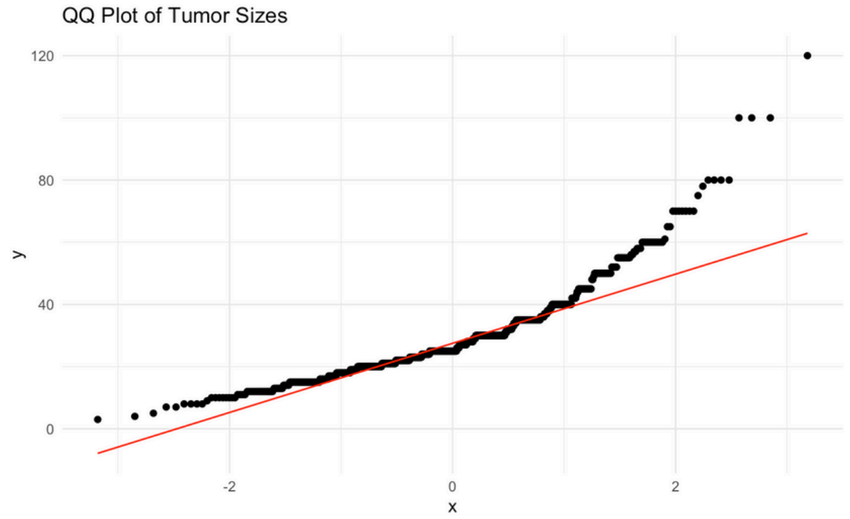


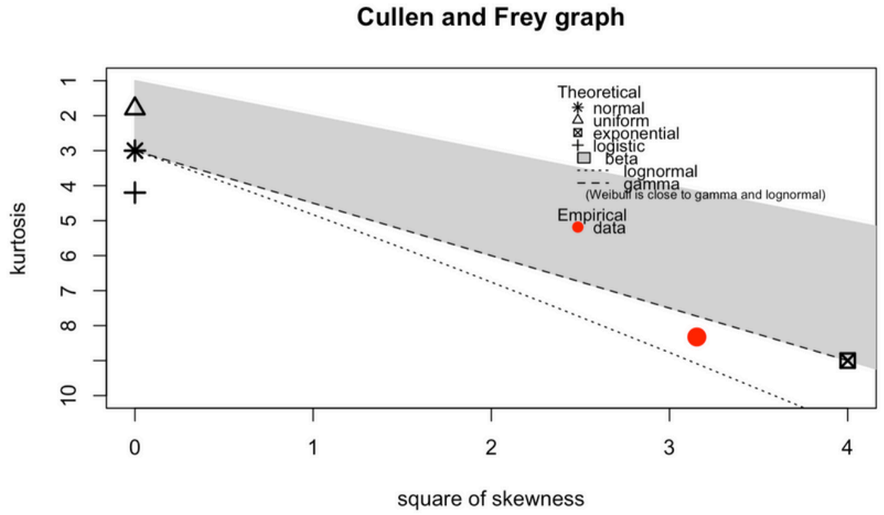


Exploratory Data Analysis

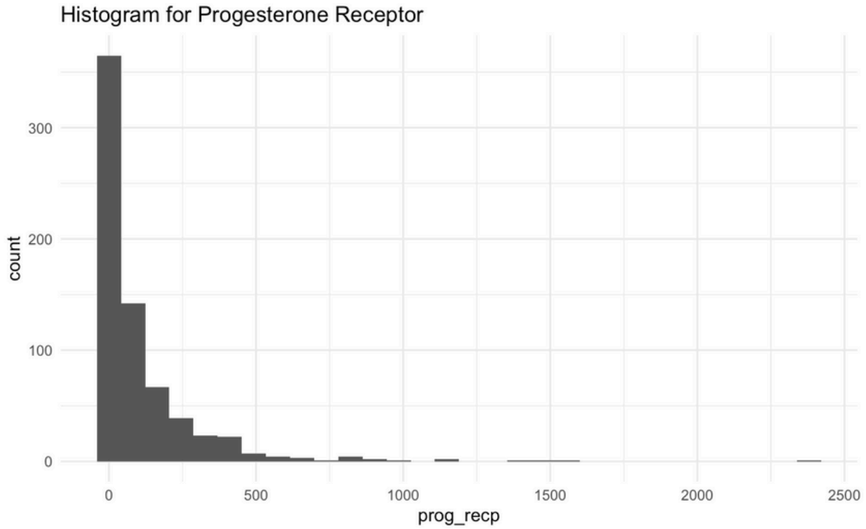


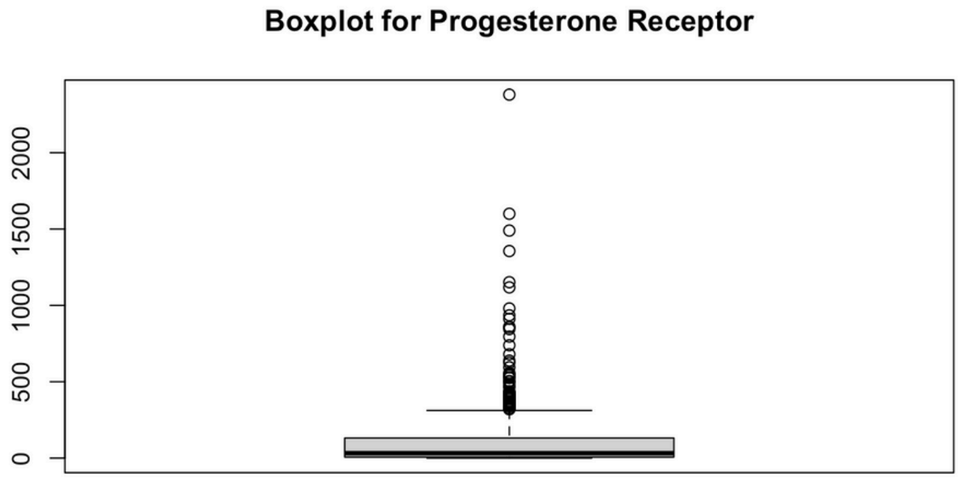


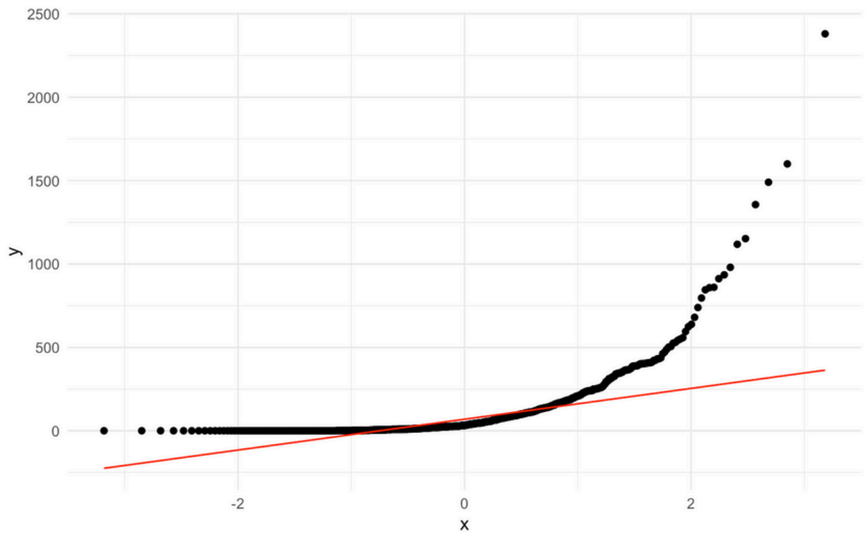


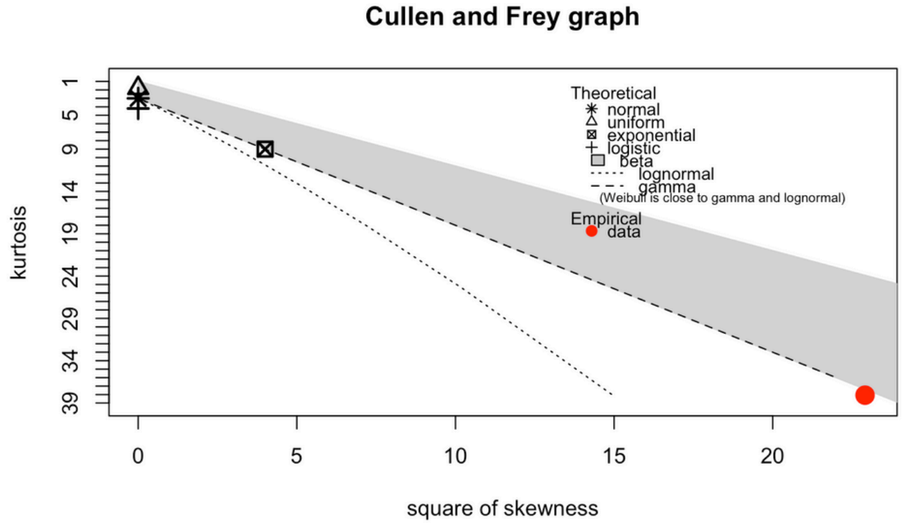


Exploratory Data Analysis

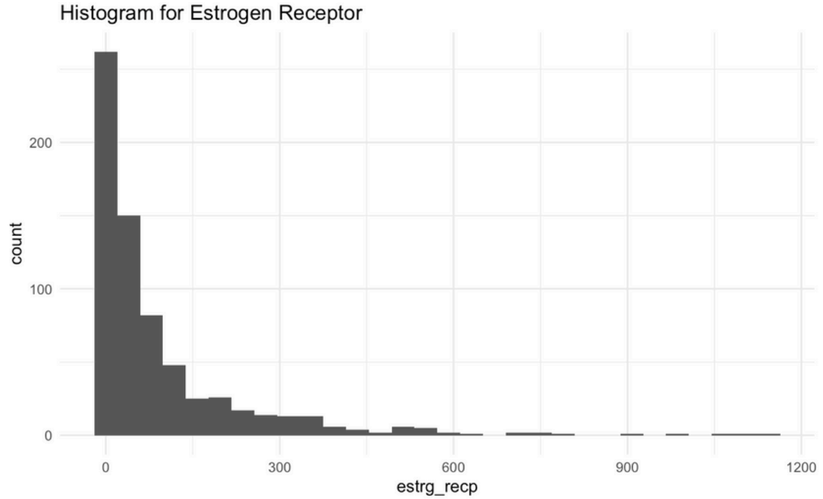


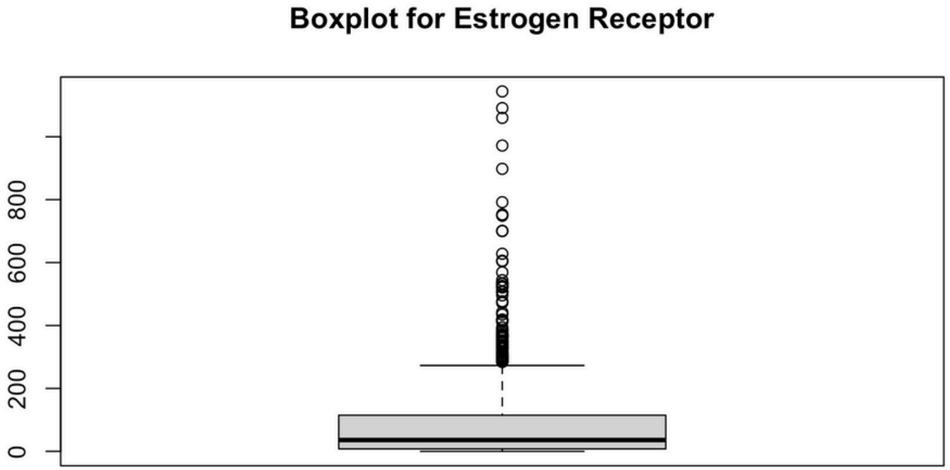


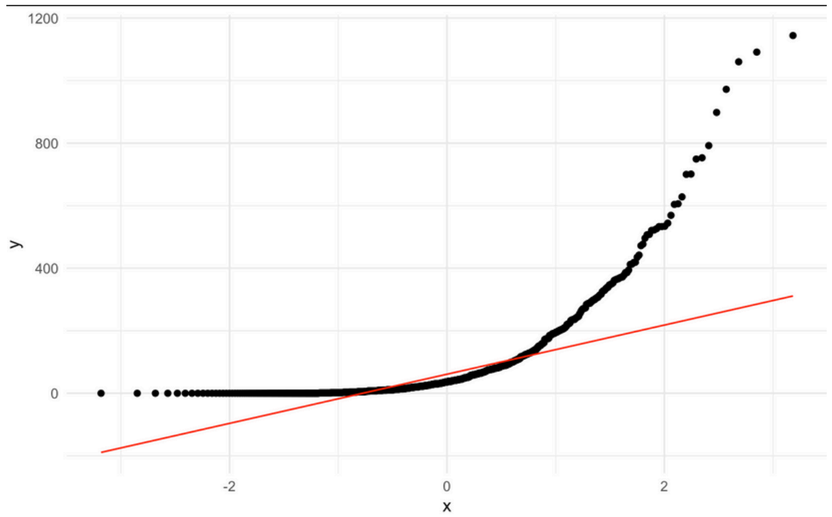


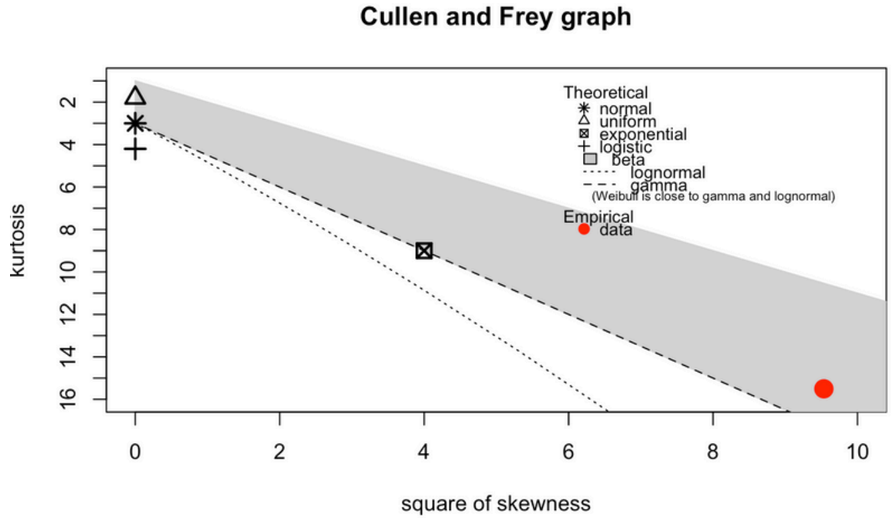


Exploratory Data Analysis

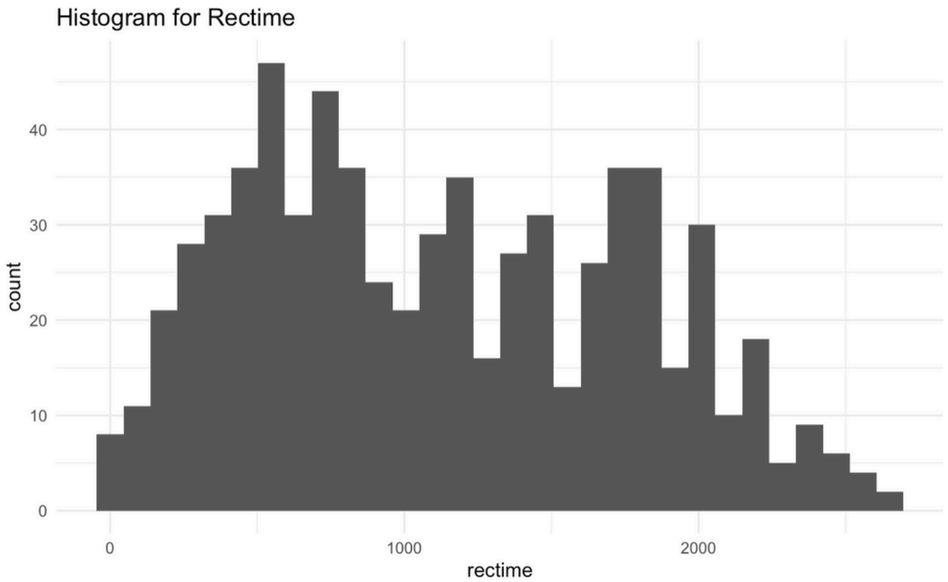


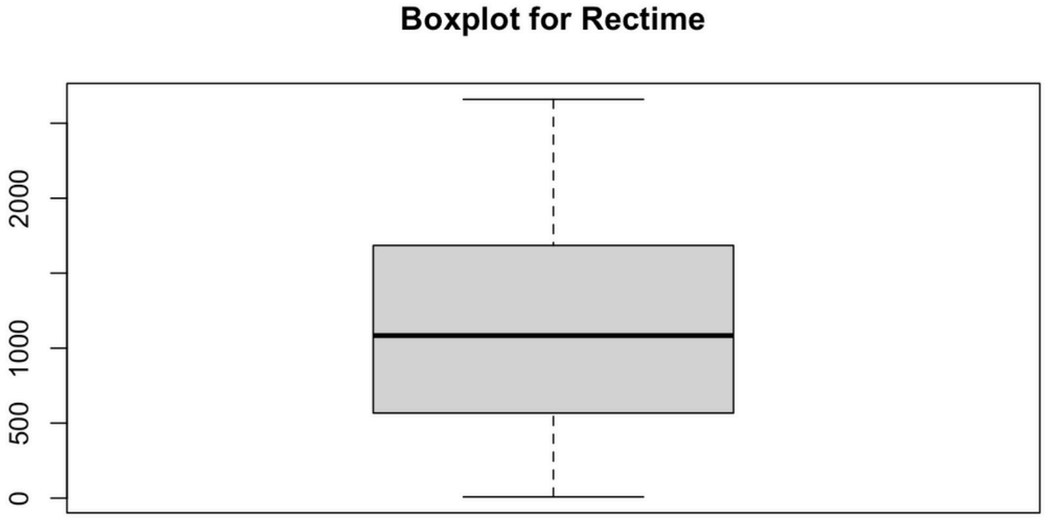


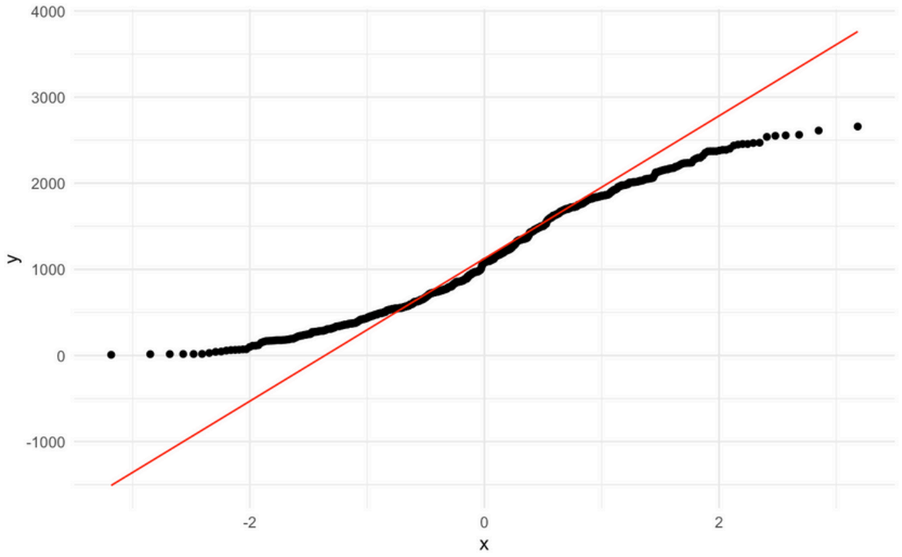


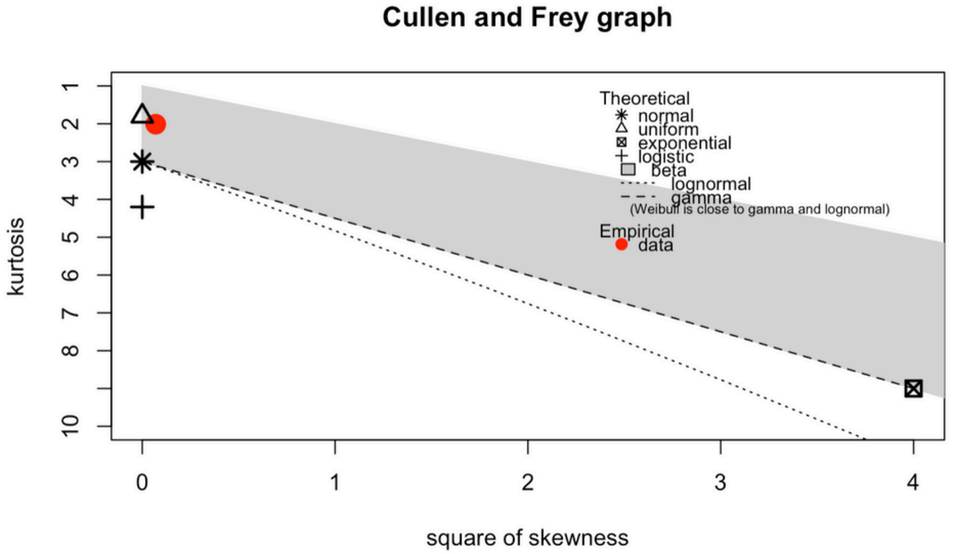


Exploratory Data Analysis

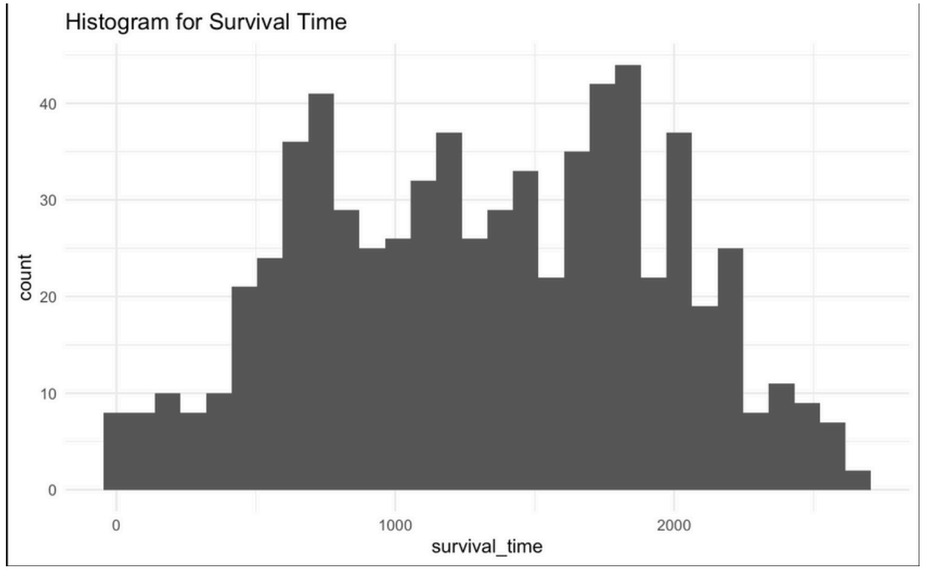


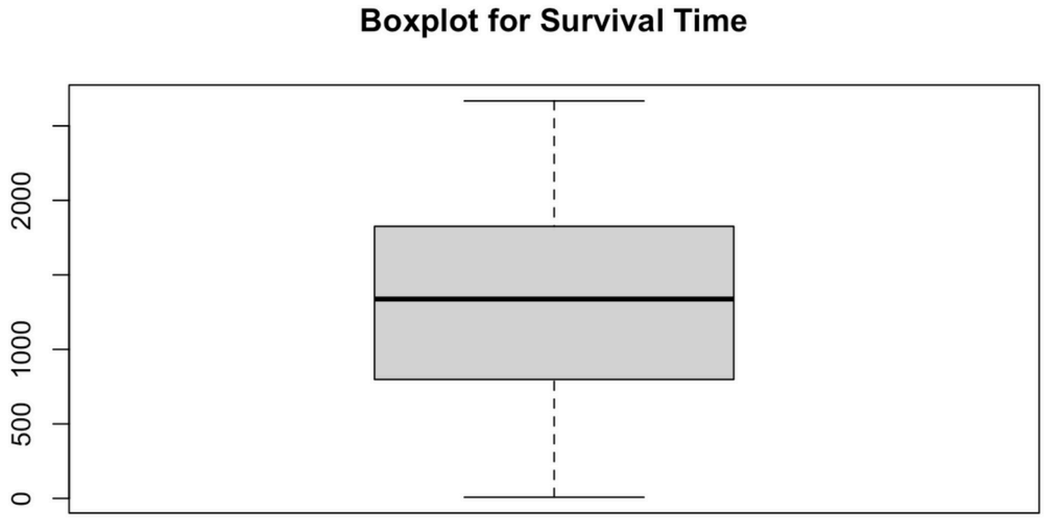


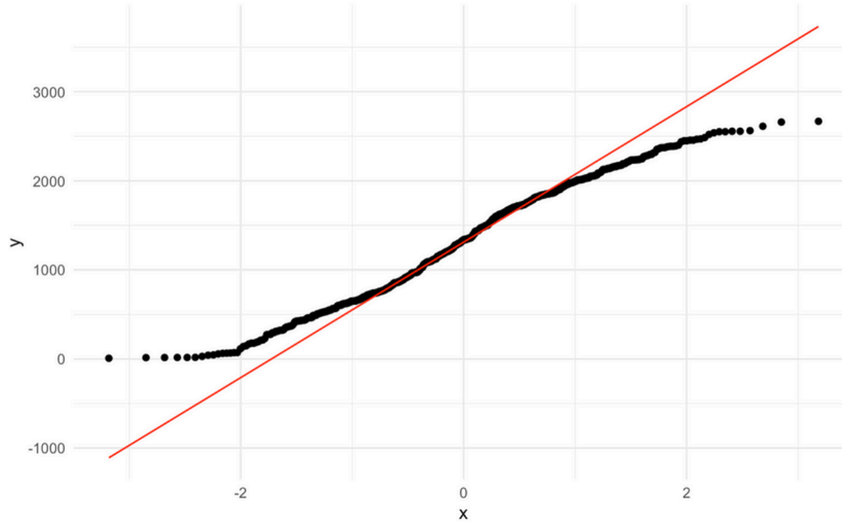


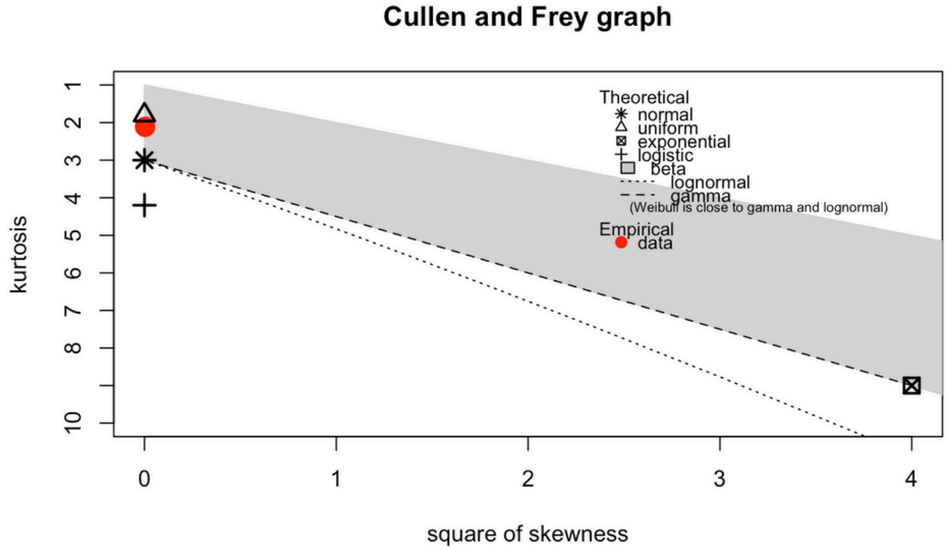


Exploratory Data Analysis

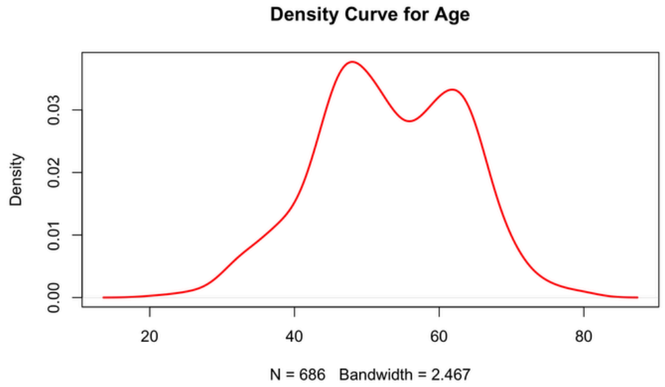


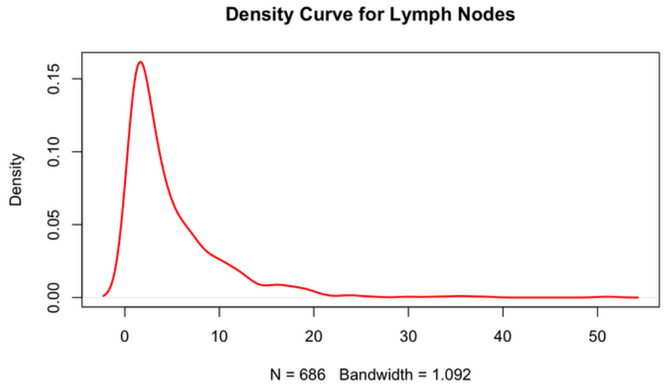


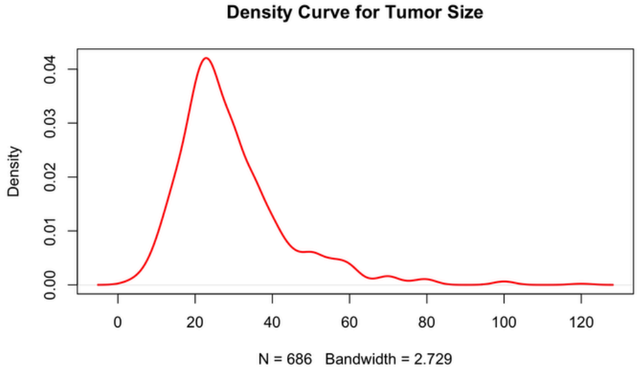


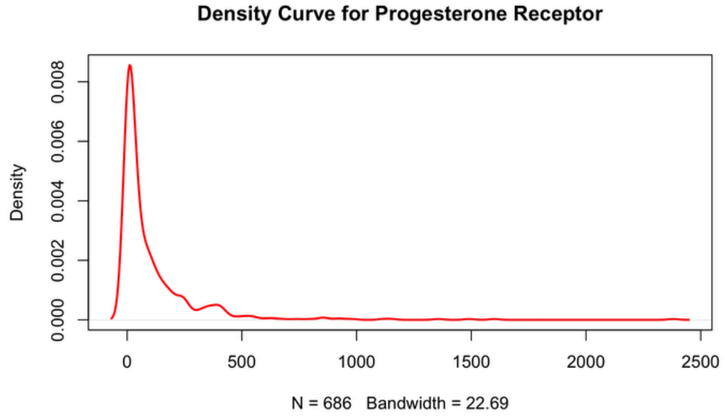


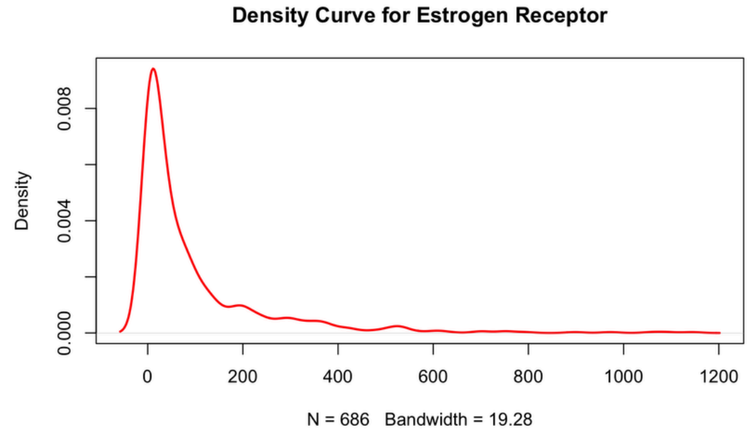
Exploratory Data Analysis

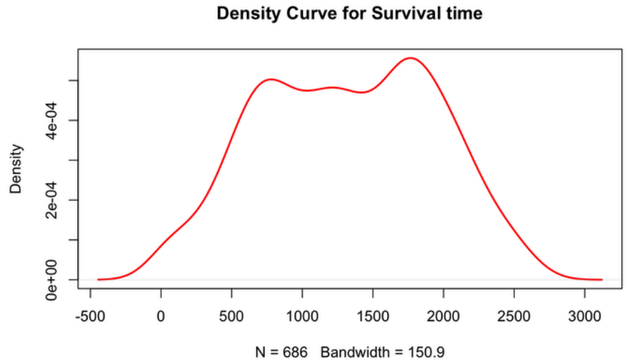




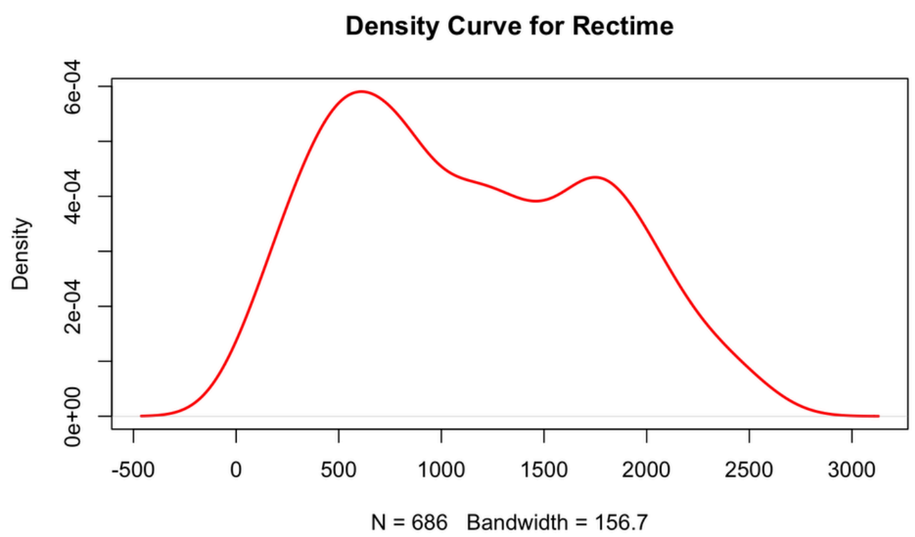


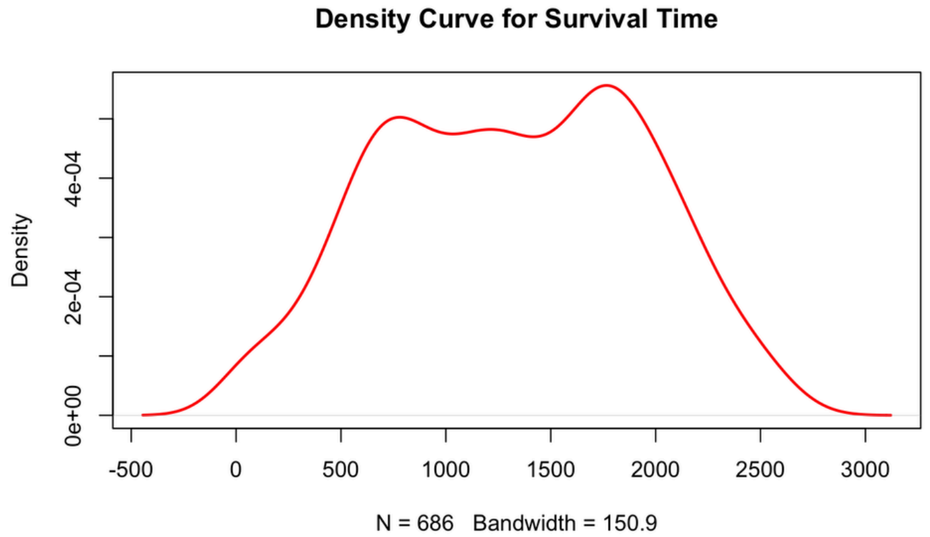




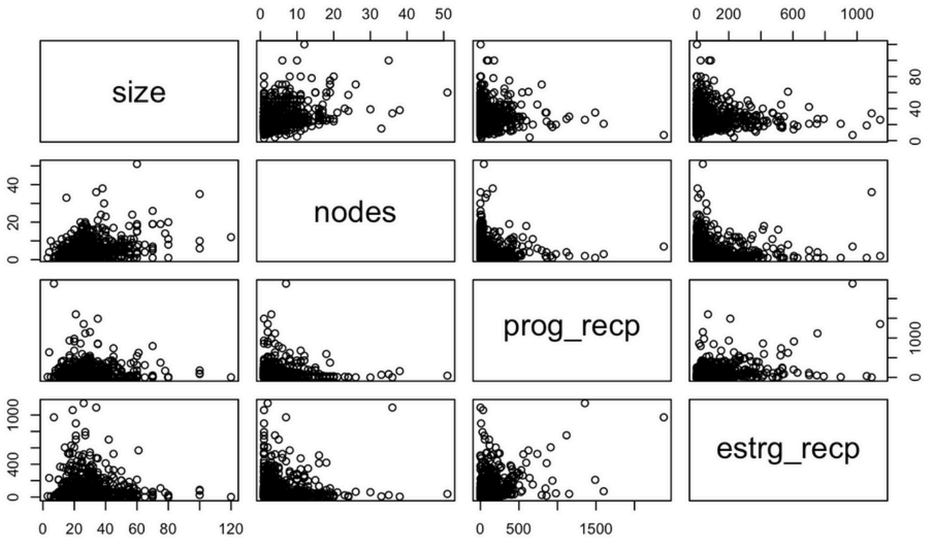


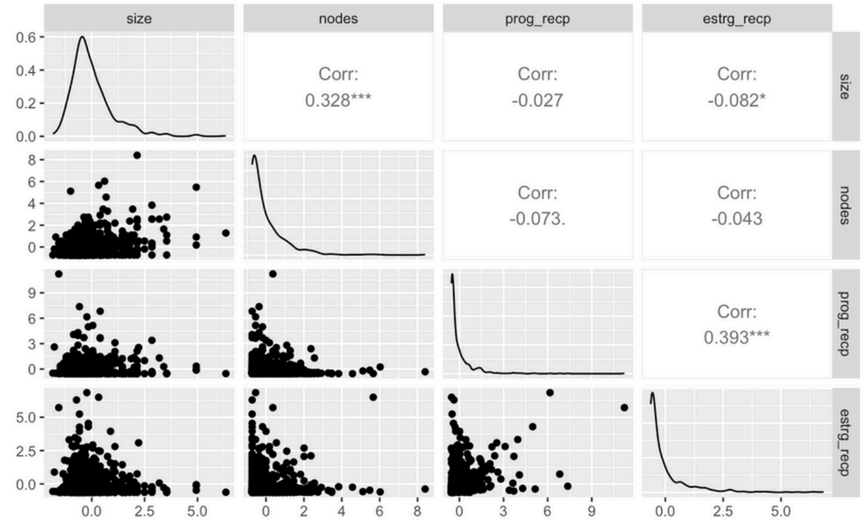
Exploratory Data Analysis

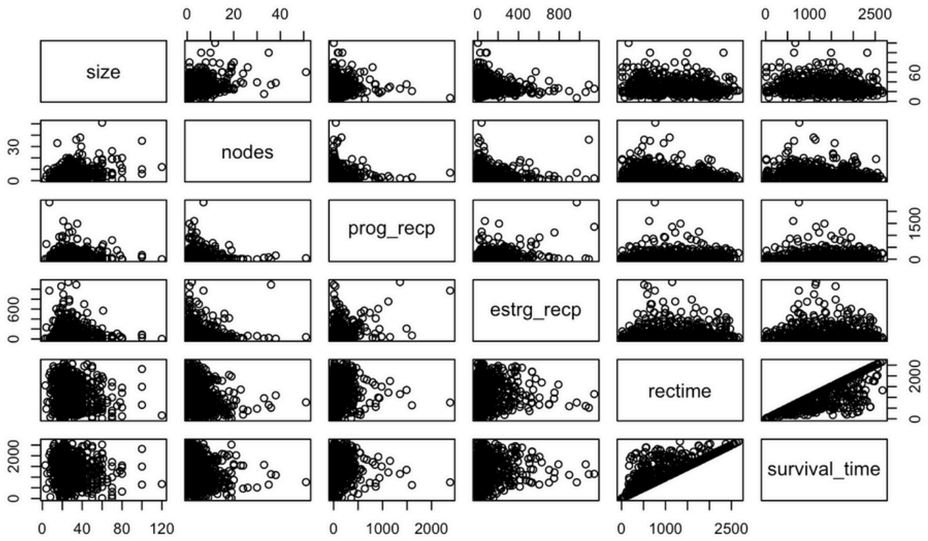


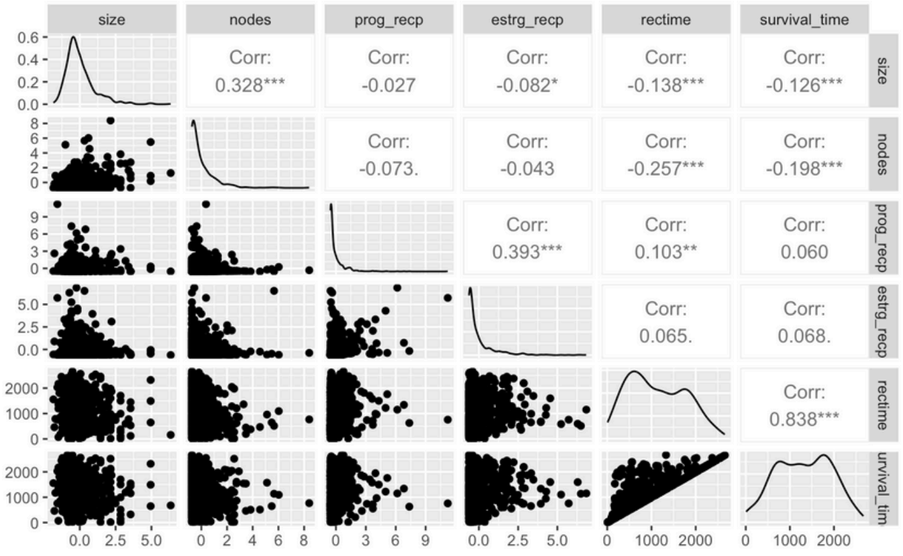


Exploratory Data Analysis

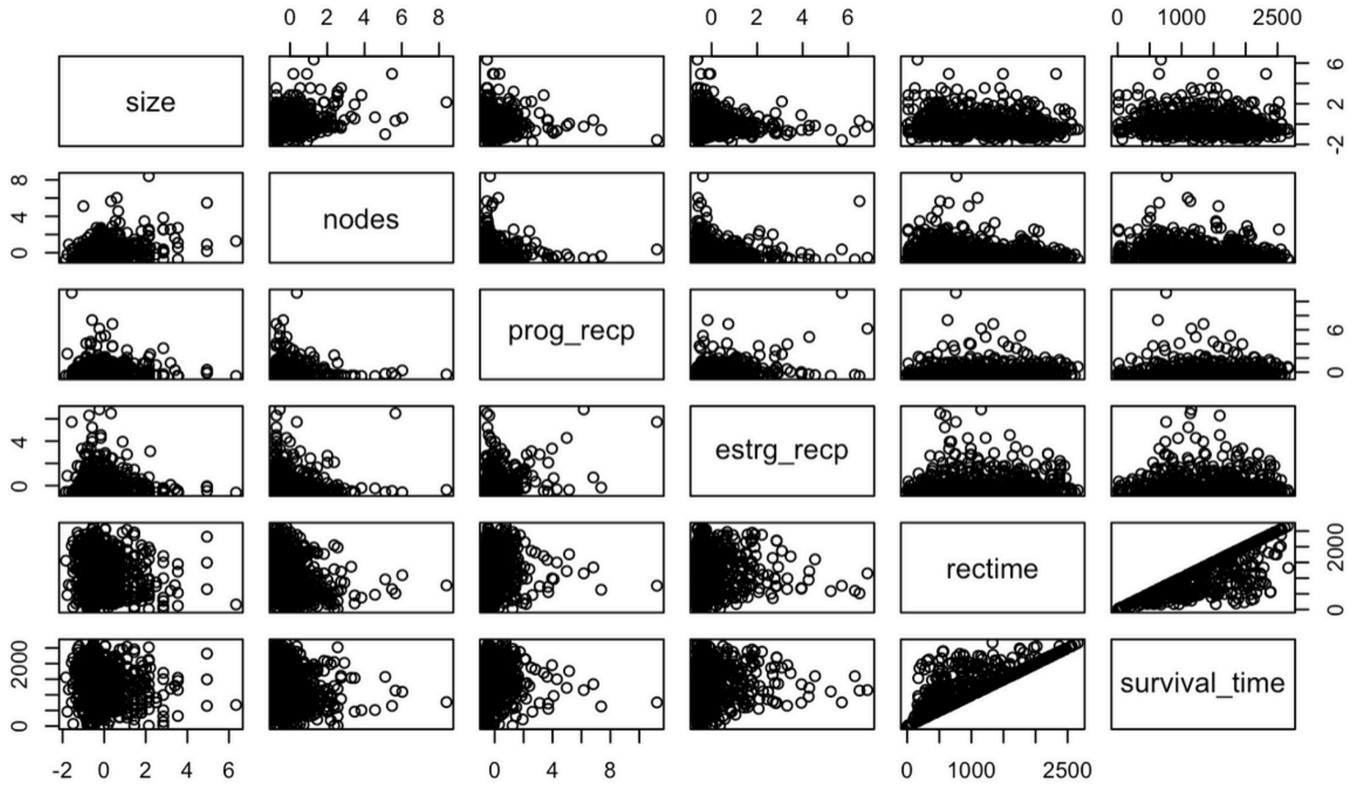




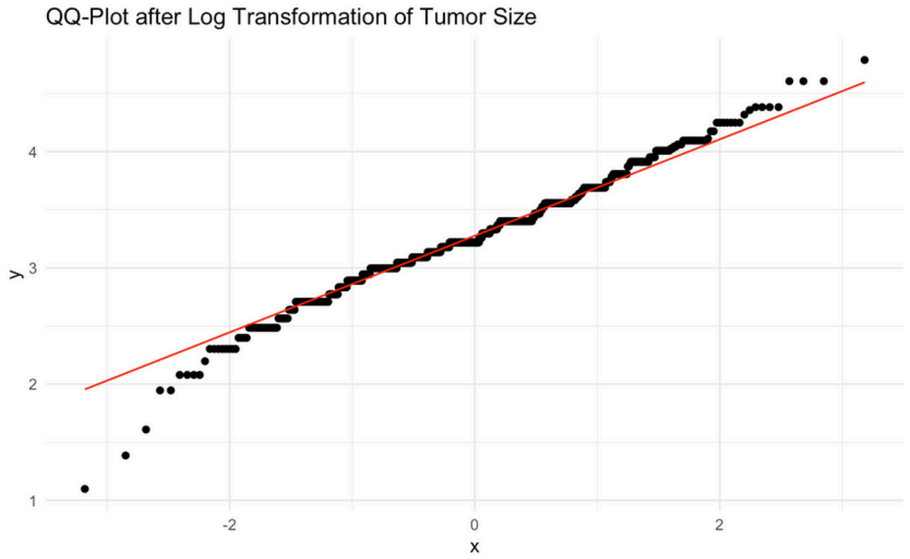


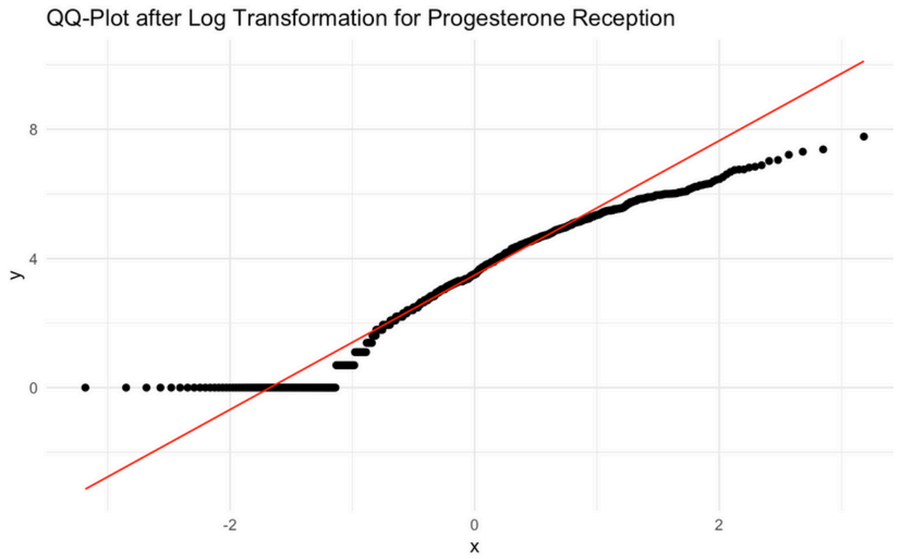


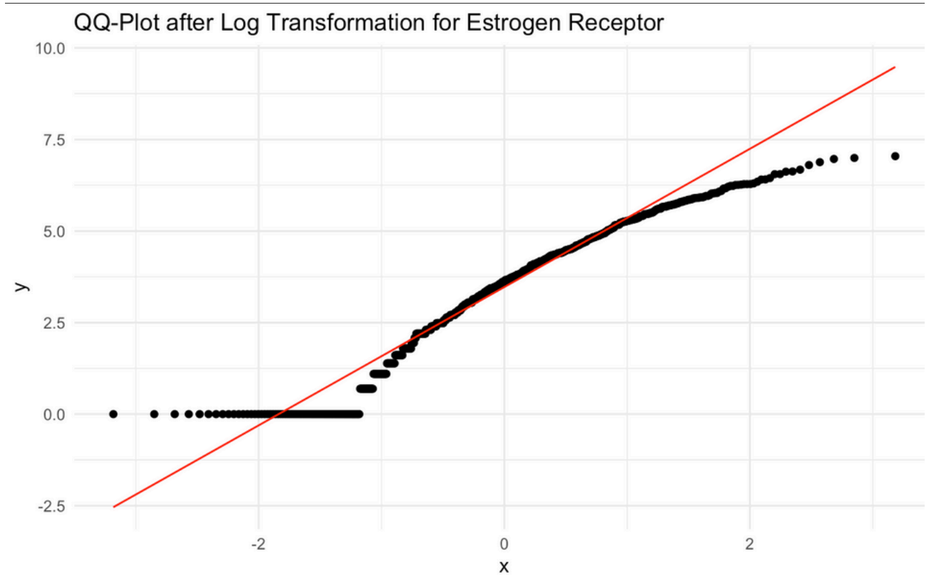
Exploratory Data Analysis



Exploratory Data Analysis



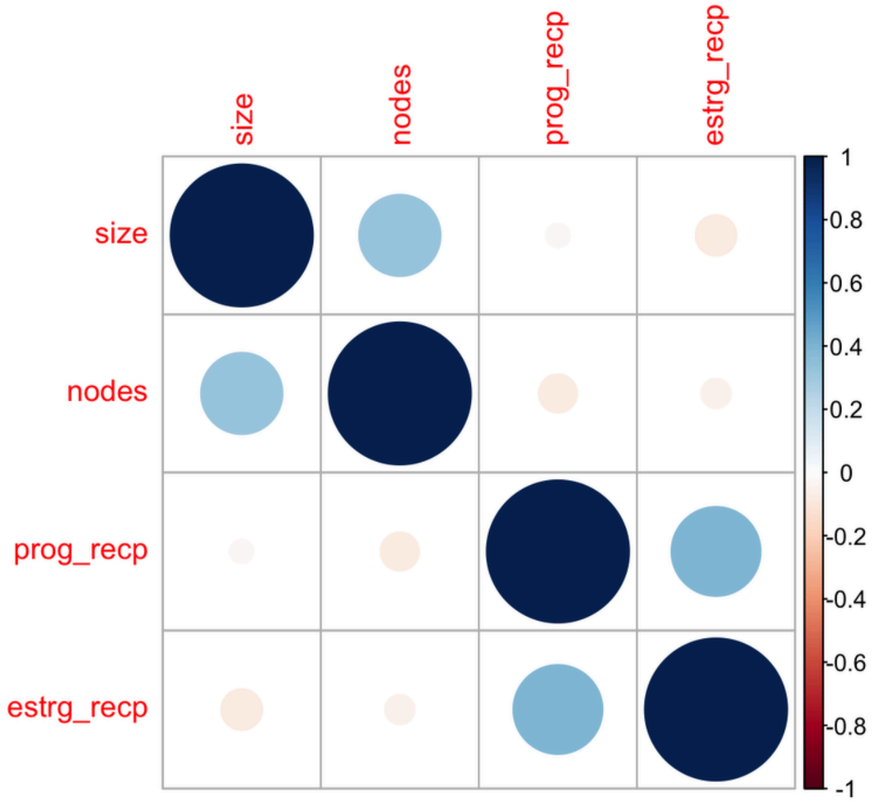


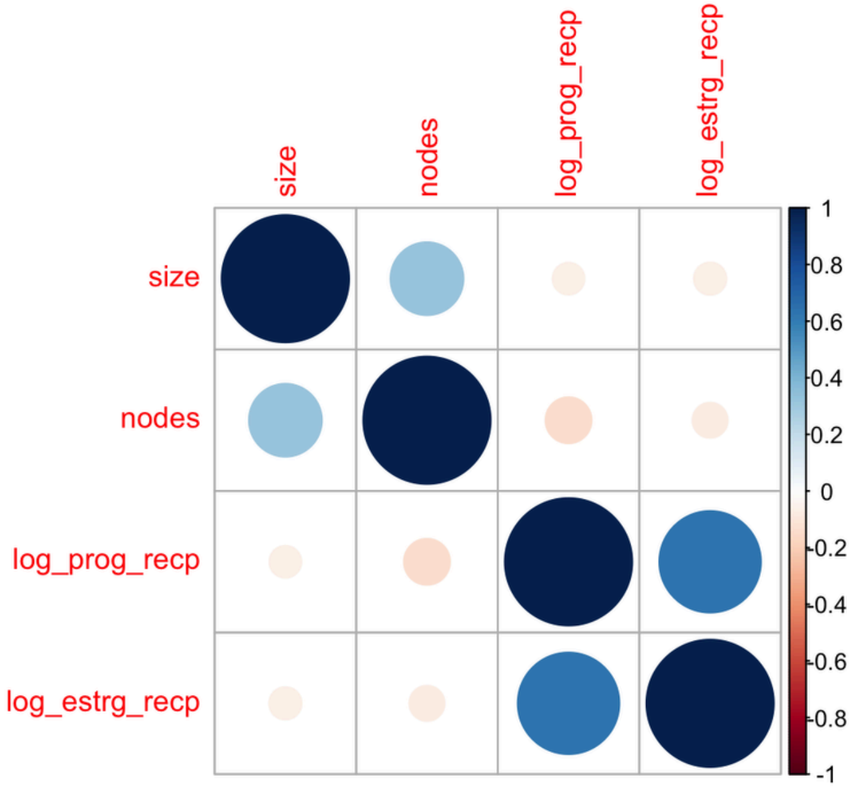


Exploratory Data Analysis

	size	nodes	prog_recp	estrg_recp
size	1.00000000	0.32766498	-0.02741477	-0.08176636
nodes	0.32766498	1.00000000	-0.07253389	-0.04318344
prog_recp	-0.02741477	-0.07253389	1.00000000	0.39260134
estrg_recp	-0.08176636	-0.04318344	0.39260134	1.00000000

	size	nodes	log_prog_recp	log_estrg_recp
size	1.00000000	0.32766498	-0.06202947	-0.06424583
nodes	0.32766498	1.00000000	-0.13087355	-0.07585719
log_prog_recp	-0.06202947	-0.13087355	1.00000000	0.63566439
log_estrg_recp	-0.06424583	-0.07585719	0.63566439	1.00000000





Kaplan-Meier Curve

The Kaplan-Meier Curve is a statistical tool used to estimate the survival function.

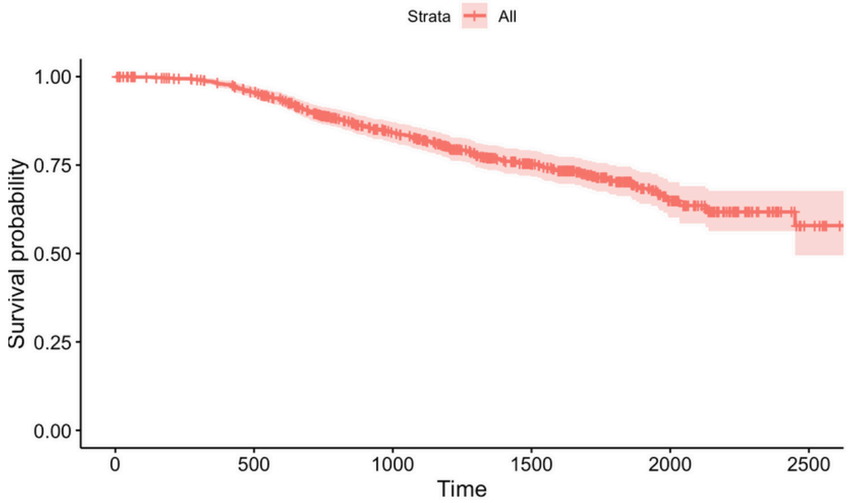
$$S(t) = \prod_{i:\, t_i <= t} (1 - d_i/n_i)$$

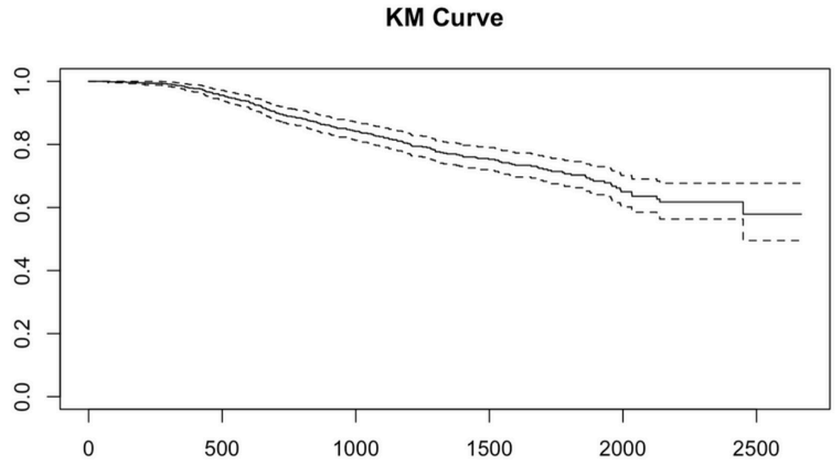
where di is the number of events at time ti and ni is the number of subjects at risk just before time ti

Characteristics:

- **Non-Parametric:** does not assume/require a specific distribution of the data
- **Censoring:** can handle censored data
- **Non-Continuous Nature:** does not require the data to be continuous
- **Comparison of Groups:** helps in assessing differences in survival probabilities in different groups

Kaplan-Meier Curve

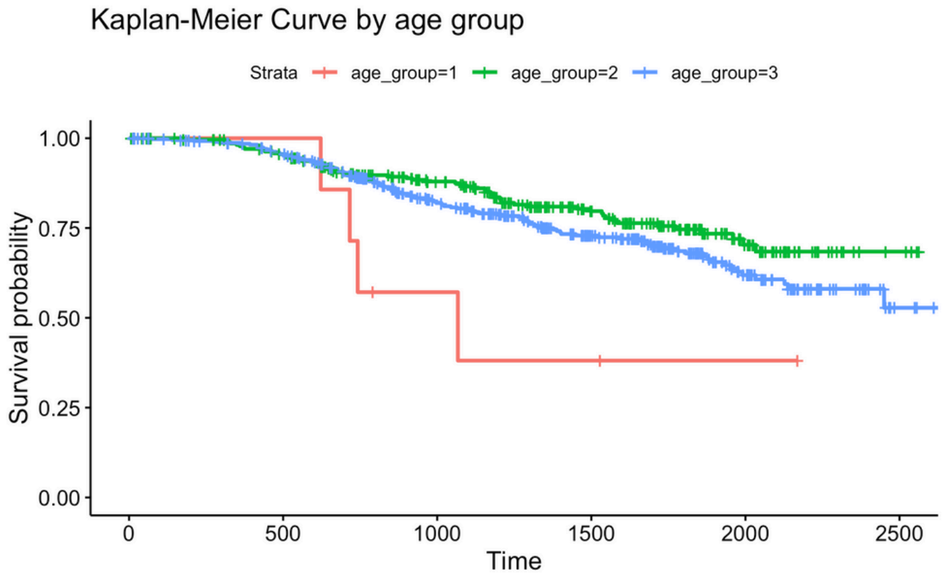




- Survival probabilities are being estimated for the entire population
- The variable is **censdead**: death/censoring indicator (death=1, alive=0)
- Curve is flat, then steep drop, and then flattens again
- Confidence interval is narrower early on
- Median survival rate is undefined

```
      n events median 0.95LCL 0.95UCL
[1,] 686    171     NA    2450      NA
```

Kaplan-Meier Curve

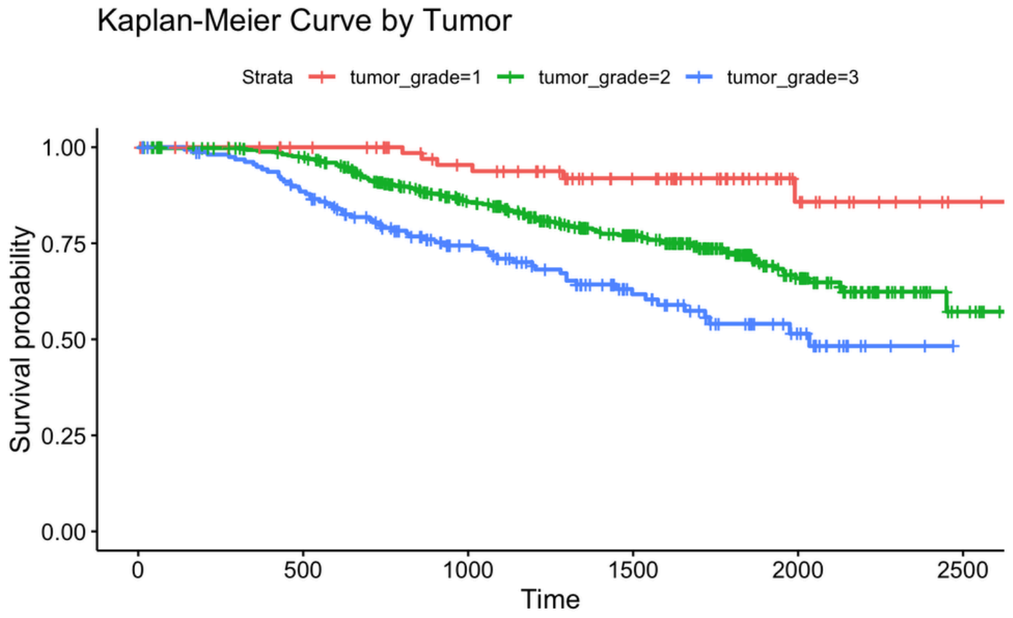


N: number of patients in each group
Observed: number of observed events
Expected: expected number of events if there was no difference
(O-E)^2/E: term to see how much the observed events deviate from expected
(O-E)^2/V: it adjusts using variance
v: variance of the difference between observed and expected events

```
Call:
survdiff(formula = Surv(survival_time, event_status) ~ age_group,
    data = X_copy)

               N Observed Expected (O-E)^2/E (O-E)^2/V
age_group=1    7        4     1.39      4.94      4.98
age_group=2  282       58    70.74      2.29      3.91
age_group=3  397      109    98.88      1.04      2.46

 Chisq= 8.3  on 2 degrees of freedom, p= 0.02
```

```
        Pairwise comparisons using Log-Rank test

data:  X_copy and age_group

  1     2
2 0.031 -
3 0.142 0.202

P value adjustment method: bonferroni
```

Kaplan-Meier Curve

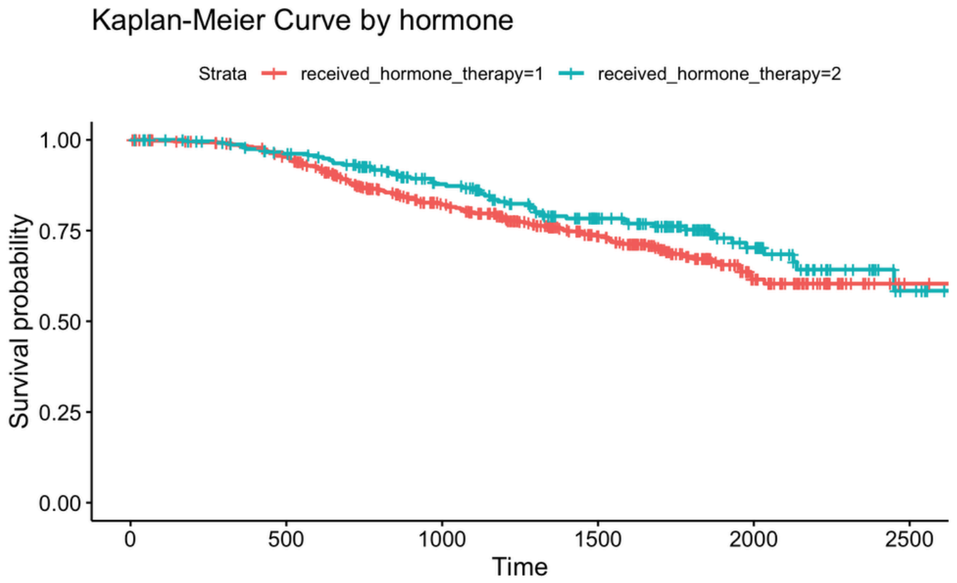


- Significant differences in tumor grade
- grade 1 vs grade 2: statistically significant
- grade 1 vs grade 3: highly statistically significant
- grade 2 vs grade 3: statistically significant

```
Call:
survdiff(formula = Surv(survival_time, event_status) ~ tumor_grade,
    data = X_copy)

                N Observed Expected (O-E)^2/E (O-E)^2/V
tumor_grade=1  81        6     22.3    11.922      13.7
tumor_grade=2 444      107    115.0     0.557       1.7
tumor_grade=3 161       58     33.7    17.543      21.9

 Chisq= 30.1  on 2 degrees of freedom, p= 3e-07
```

```
        Pairwise comparisons using Log-Rank test

data:  X_copy and tumor_grade

  1       2
2 0.00455 -
3 4.2e-06 0.00036

P value adjustment method: bonferroni
```

Kaplan-Meier Curve

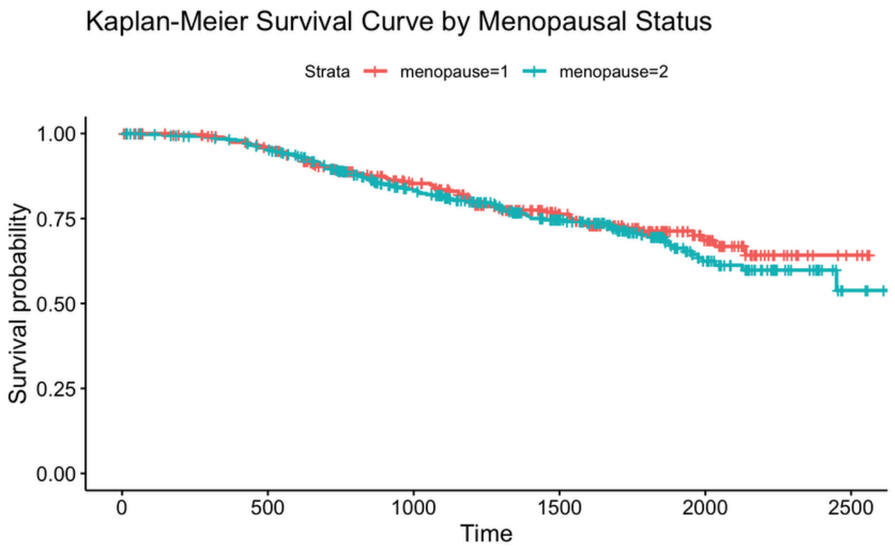


- 1: did receive, 2: did not receive
- Patients who recieved hormone therapy have slightly higher survival rate??

```
Call:
survdiff(formula = Surv(survival_time, event_status) ~ received_hormone_therapy,
    data = X_copy)

                             N Observed Expected (O-E)^2/E (O-E)^2/V
received_hormone_therapy=1 440      115    104.8     0.986      2.56
received_hormone_therapy=2 246       56     66.2     1.562      2.56

 Chisq= 2.6  on 1 degrees of freedom, p= 0.1
```

```
        Pairwise comparisons using Log-Rank test

data:  X_copy and received_hormone_therapy

  1
2 0.11

P value adjustment method: bonferroni
```

Kaplan-Meier Curve

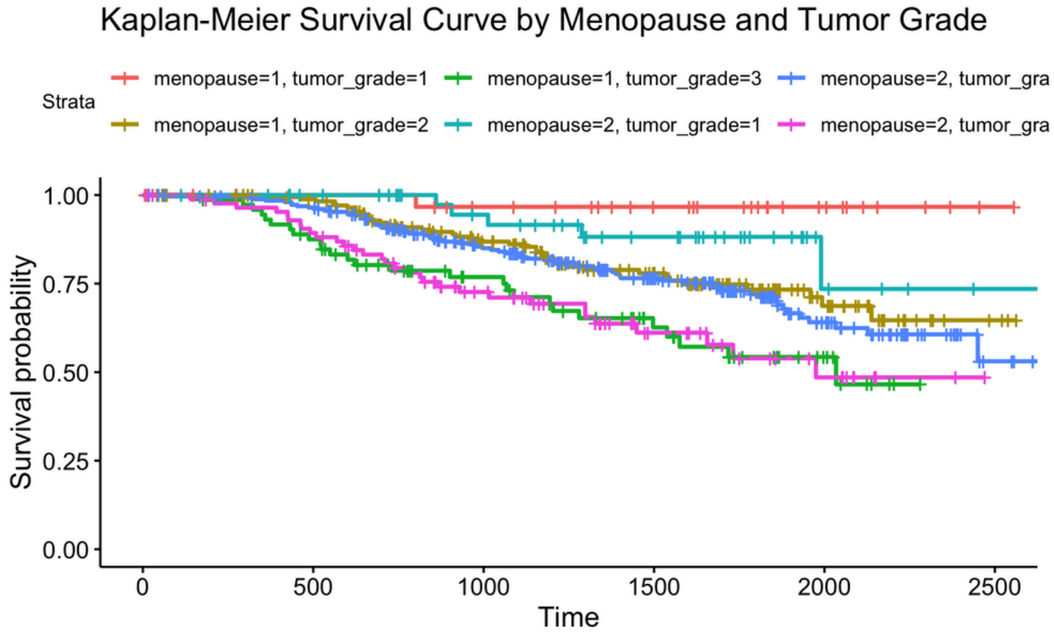


- 1: menopausal, 2: not menopausal

```
Call:
survdiff(formula = Surv(survival_time, event_status) ~ menopause,
    data = X_copy)

                N Observed Expected (O-E)^2/E (O-E)^2/V
menopause=1 290       67     71.5     0.285      0.49
menopause=2 396      104     99.5     0.205      0.49

 Chisq= 0.5  on 1 degrees of freedom, p= 0.5
```

```
        Pairwise comparisons using Log-Rank test

data:  X_copy and menopause

  1
2 0.48

P value adjustment method: bonferroni
```

Kaplan-Meier Curve



- meno 1, grade 1 vs meno 1, grade 3
- meno 2, grade 3 vs meno 1, grade 1
- meno 2, grade 3 vs meno 1, grade 2
- meno 2, grade 1 vs meno 1, grade 3
- meno 2, grade 3 vs meno 2, grade 1

```
Call:
survdiff(formula = Surv(survival_time, event_status) ~ menopause +
    tumor_grade, data = X_copy)

                                N Observed Expected (O-E)^2/E (O-E)^2/V
menopause=1, tumor_grade=1  33        1     10.1    8.2183    8.7563
menopause=1, tumor_grade=2 183       39     45.7    0.9823    1.3411
menopause=1, tumor_grade=3  74       27     15.7    8.1408    8.9767
menopause=2, tumor_grade=1  48        5     12.2    4.2401    4.5682
menopause=2, tumor_grade=2 261       68     69.3    0.0245    0.0412
menopause=2, tumor_grade=3  87       31     18.0    9.4027   10.5434

 Chisq= 31.1  on 5 degrees of freedom, p= 9e-06
```

```
        Pairwise comparisons using Log-Rank test

data:  X_copy and menopause + tumor_grade

                            menopause=1, tumor_grade=1 menopause=1, tumor_grade=2
menopause=1, tumor_grade=2 0.1511                     -
menopause=1, tumor_grade=3 0.0032                     0.0640
menopause=2, tumor_grade=1 1.0000                     1.0000
menopause=2, tumor_grade=2 0.0677                     1.0000
menopause=2, tumor_grade=3 0.0033                     0.0461
                            menopause=1, tumor_grade=3 menopause=2, tumor_grade=1
menopause=1, tumor_grade=2 -                          -
menopause=1, tumor_grade=3 -                          -
menopause=2, tumor_grade=1 0.0268                     -
menopause=2, tumor_grade=2 0.1801                     0.8064
menopause=2, tumor_grade=3 1.0000                     0.0263
                            menopause=2, tumor_grade=2
menopause=1, tumor_grade=2 -
menopause=1, tumor_grade=3 -
menopause=2, tumor_grade=1 -
menopause=2, tumor_grade=2 -
menopause=2, tumor_grade=3 0.1193

P value adjustment method: bonferroni
```

Kaplan-Meier Curve

- All pairwise comparisons have p-values of 1.00 or greater than 0.1, indicating that none of the pairwise differences are statistically significant after Bonferroni correction
- The Kaplan-Meier plot visually suggests some separation between the curves, particularly between pre-menopausal women in age group 2 and post-menopausal women in age group 3. However, the log-rank test (both overall and pairwise comparisons) fails to find statistically significant differences

```
Call:
survdiff(formula = Surv(survival_time, event_status) ~ menopause +
    age_group, data = X_copy)

                                N Observed Expected (O-E)^2/E (O-E)^2/V
menopause=1, age_group=1    7        4     1.39     4.936     4.982
menopause=1, age_group=2  248       53    61.40     1.150     1.795
menopause=1, age_group=3   35       10     8.73     0.186     0.196
menopause=2, age_group=2   34        5     9.33     2.011     2.129
menopause=2, age_group=3  362       99    90.15     0.868     1.838

 Chisq= 9.2  on 4 degrees of freedom, p= 0.06
```

```
        Pairwise comparisons using Log-Rank test

data:  X_copy and menopause + age_group

                            menopause=1, age_group=1 menopause=1, age_group=2
menopause=1, age_group=2 0.14                     -
menopause=1, age_group=3 1.00                     1.00
menopause=2, age_group=2 0.12                     1.00
menopause=2, age_group=3 0.46                     1.00
                            menopause=1, age_group=3 menopause=2, age_group=2
menopause=1, age_group=2 -                        -
menopause=1, age_group=3 -                        -
menopause=2, age_group=2 1.00                     -
menopause=2, age_group=3 1.00                     1.00

P value adjustment method: bonferroni
```

Cox-Proportional Hazards Model

Cox Proportional Hazards Model is a statistical method to analyze the relationship between the survival time and one or more predictors

$\ln(Haz) = \ln(h0(t)) + (B1X1 + B2X2 + \ldots + BkXk) \Rightarrow Haz = h0(t) * \exp(B1X1 + B2X2 + \ldots + BkXk)$

- **Non-informative censoring:** random chance of survival
- **Survival time:** it is independent
- **ln(Haz):** is linear function of X
- Hazards are proportional over time
- Values of X do not change over time (treatment A and treatment B, dosage, etc)

Cox-Proportional Hazards Model

```
Call:
coxph(formula = Surv(survival_time, event_status) ~ size + tumor_grade +
    nodes + log_prog_recp, data = X)

  n= 686, number of events= 171

                  coef exp(coef)  se(coef)      z Pr(>|z|)
size           0.010807  1.010866  0.004786  2.258   0.0239 *
tumor_grade.L  0.676998  1.967960  0.316568  2.139   0.0325 *
tumor_grade.Q -0.184674  0.831375  0.192905 -0.957   0.3384
nodes          0.054667  1.056188  0.009738  5.614 1.98e-08 ***
log_prog_recp -0.270348  0.763114  0.041897 -6.453 1.10e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

               exp(coef) exp(-coef) lower .95 upper .95
size              1.0109     0.9893    1.0014    1.0204
tumor_grade.L     1.9680     0.5081    1.0582    3.6600
tumor_grade.Q     0.8314     1.2028    0.5696    1.2134
nodes             1.0562     0.9468    1.0362    1.0765
log_prog_recp     0.7631     1.3104    0.7030    0.8284

Concordance= 0.751  (se = 0.017 )
Likelihood ratio test= 116.6  on 5 df,   p=<2e-16
Wald test            = 125  on 5 df,   p=<2e-16
Score (logrank) test = 143.3  on 5 df,   p=<2e-16
```

- **coef:** for each unit increase in the variable, log hazards ratio increases by *coef*
- **exp(coef):** a unit increase in the variable, increases the hazards ratio by exp(coef)
- **exp(-coef): a unit decrease in the variable, decreases the hazards ratio by (1-exp(-coef))%**
- **confidence interval:** we are 95% confident that the true hazard ratio is in that interval
- **Likelihood, Wald**, and **Logrank** are statistically significant (p-value < 0.05)

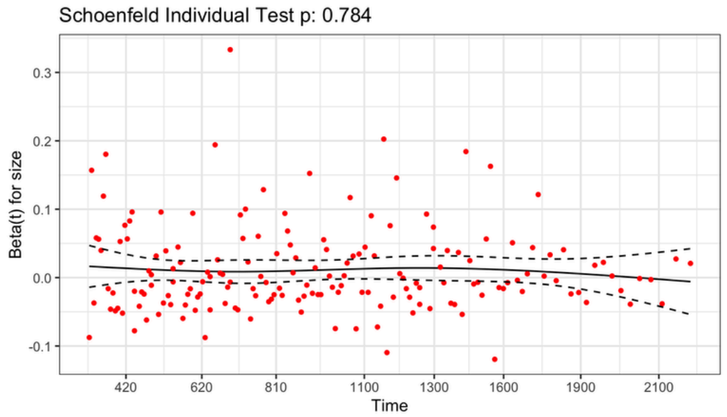
Cox-Proportional Hazards Model

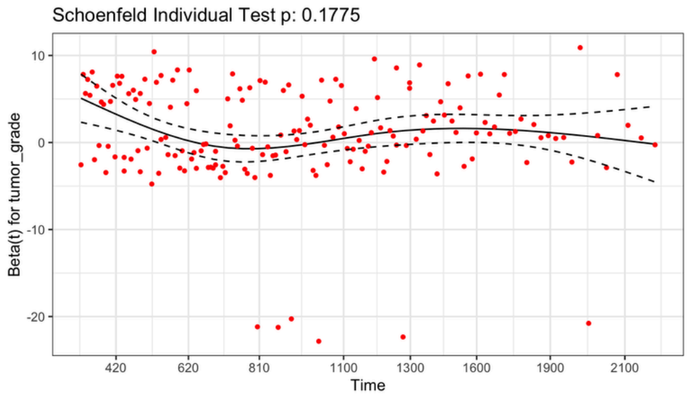
```
test_data <- data.frame(
  size = c(2.5, 3),
  tumor_grade = as.factor(c(1, 2)),
  nodes = c(0, 3),
  log_prog_recp = c(0.5, 1)
)

surv_fit <- survfit(cox_transformed.rm, newdata = test_data)

summary(surv_fit, times = (c(100, 200, 300, 400, 500)))
```

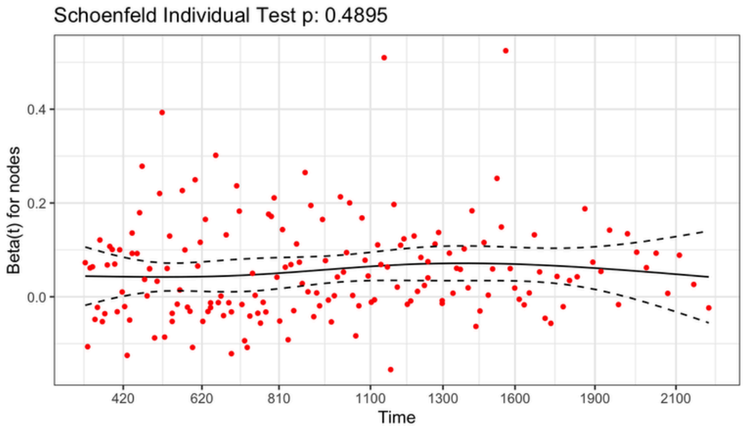
```
 Call: survfit(formula = cox_transformed.rm, newdata = test_data)

  time n.risk n.event survival1 survival2
   100    671       1     0.999     0.998
   200    663       2     0.997     0.993
   300    655       3     0.993     0.986
   400    642       9     0.983     0.965
   500    622      14     0.966     0.931
```

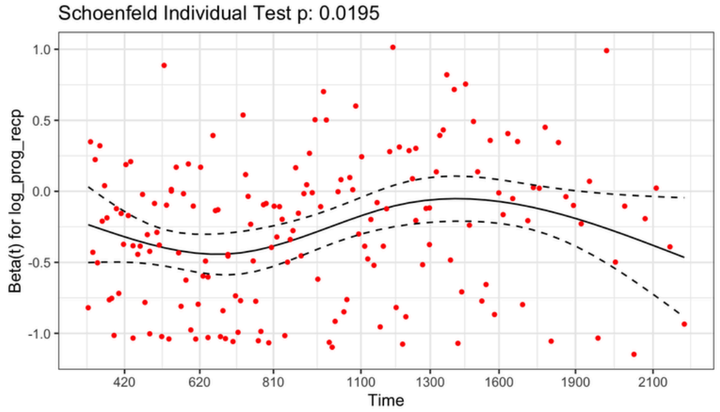
Cox-Proportional Hazards Model

Schoenfeld's Residuals









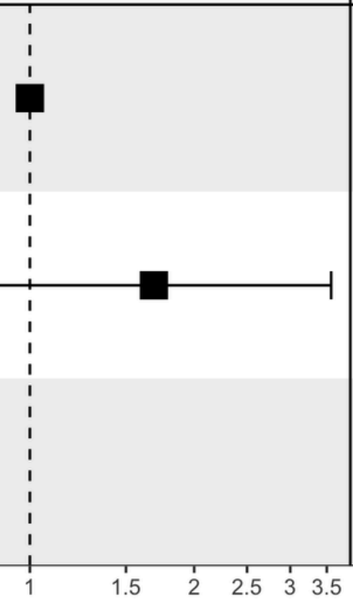
Cox-Proportional Hazards Model

```
Call:
coxph(formula = Surv(survival_time, event_status) ~ menopause_cat +
    tumor_grade + menopause_cat:tumor_grade, data = X_copy)

  n= 686, number of events= 171

                               coef exp(coef) se(coef)      z Pr(>|z|)
menopause_cat2                0.5234    1.6877   0.3815  1.372  0.17013
tumor_grade.L                 2.0257    7.5817   0.7203  2.812  0.00492 **
tumor_grade.Q                -0.5950    0.5516   0.4358 -1.365  0.17222
menopause_cat2:tumor_grade.L -1.0079    0.3650   0.7967 -1.265  0.20586
menopause_cat2:tumor_grade.Q  0.4716    1.6025   0.4885  0.965  0.33434
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

                              exp(coef) exp(-coef) lower .95 upper .95
menopause_cat2                  1.6877     0.5925   0.79900     3.565
tumor_grade.L                   7.5817     0.1319   1.84788    31.107
tumor_grade.Q                   0.5516     1.8130   0.23476     1.296
menopause_cat2:tumor_grade.L    0.3650     2.7399   0.07657     1.740
menopause_cat2:tumor_grade.Q    1.6025     0.6240   0.61520     4.174

Concordance= 0.618  (se = 0.021 )
Likelihood ratio test= 34.67  on 5 df,   p=2e-06
Wald test            = 25.84  on 5 df,   p=1e-04
Score (logrank) test = 31.15  on 5 df,   p=9e-06
```

- **coef of menopause_cat2:** for being menopausal, log hazards ratio increases by 0.5234
- **exp(coef):** A person who is not menopausal is 1.69 times more likely to die, than a person who is not menopausal OR a person who is menopausal is 69% more likely to die than a person who is not
- **exp(-coef):** Hazard ratio for someone who is menopausal, i.e, they are 0.59 times more likely to die than someone who is not menopausal
- **tumor_grade.L:** captures direct relationship between tumor_grade and hazard_rate
- **tumor_grade.Q:** captures non-linear relationship
- **menopause_cat2:tumor_grade.L:** decreases the hazard by a factor of 0.3650, i.e, by 36.5%
- **menopause_cat2:tumor_grade.Q:** increases the hazard by a factor of 1.6025, meaning about a 60.25% higher risk
- **C-index:** goodness-of-fit (percentage of observations which are concordant)
- **confidence interval:** we are 95% confident that the true hazard ratio is in that interval
- **Likelihood, Wald**, and **Logrank** are statistically significant (p-value < 0.05)

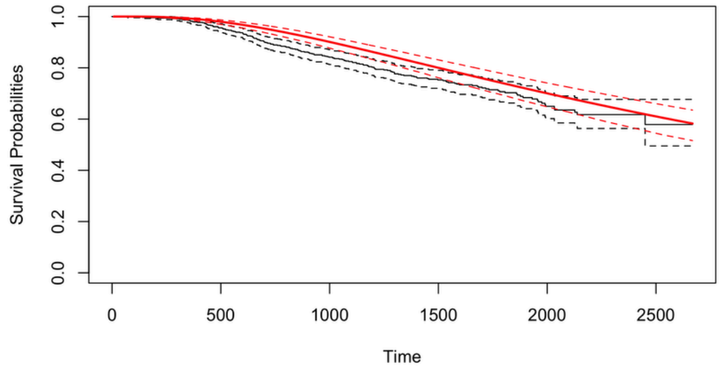
Cox-Proportional Hazards Model

Variable	N	Hazard ratio	p
menopause_cat	1 290		Reference
	2 396		1.69 (0.80, 3.56) 0.2
tumor_grade	686		



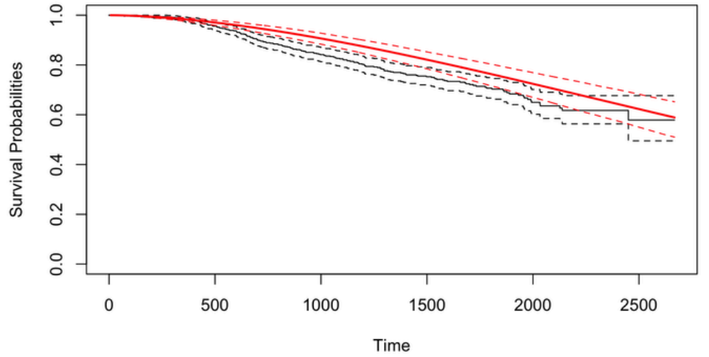
Additional Models

Updated Lognormal with Categorical Variables



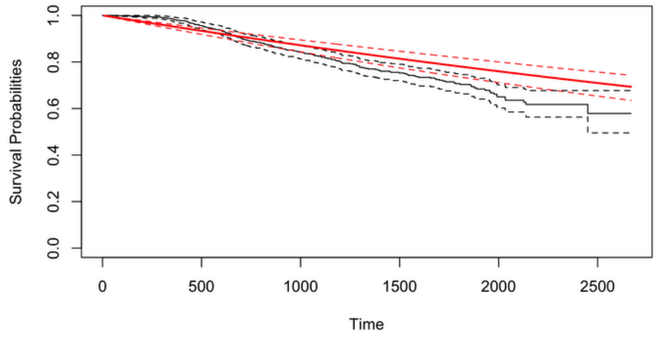
lognormal_cat_rm_fs <- flexsurvreg(Surv(survival_time, event_status) ~ size + nodes + prog_recp + age_group + tumor_grade + received_hormone_therapy, data = X_copy, dist = "lognormal")

Updated Weibull with Categorical Variables



weibull_cat_rm_fs <- flexsurvreg(Surv(survival_time, event_status) ~ size + nodes + prog_recp + age_group + tumor_grade + received_hormone_therapy, data = X_copy, dist = "weibull")

Updated Exponential Model with Categorical Variables



exp_model_cat_rm_fs <- flexsurvreg(Surv(survival_time, event_status) ~ size + nodes + tumor_grade + prog_recp + age_group, data = X_copy, dist = "exp")

Additional Models

$$AIC = 2k - 2\ln(\hat{L})$$

$$BIC = \ln(n)k - 2\ln(\hat{L})$$

```
AIC for Exponential with Categories: 3183.263
 AIC for Weibull with Categories: 3130.189
 AIC for Log-normal with Categories: 3118.848
```

```
BIC for Exponential with Categories: 3219.51
 BIC for Weibull with Categories: 3175.497
 BIC for Log-normal with Categories: 3164.157
```

```
Call:
concordance.formula(object = Surv(X_copy$survival_time, X_copy$event_status) ~
    lognormal_pred)

n= 686
Concordance= 0.7552 se= 0.01754
concordant discordant     tied.x     tied.y    tied.xy
     58701      19033          0          7          0
```

```
Call:
concordance.formula(object = Surv(X_copy$survival_time, X_copy$event_status) ~
    weibull_pred)

n= 686
Concordance= 0.7553 se= 0.01762
concordant discordant     tied.x     tied.y    tied.xy
     58715      19019          0          7          0
```

```
Call:
concordance.formula(object = Surv(X_copy$survival_time, X_copy$event_status) ~
    exp_pred)

n= 686
Concordance= 0.7515 se= 0.01783
concordant discordant     tied.x     tied.y    tied.xy
     58418      19315          1          7          0
```

Conclusion

- Higher tumor grades, grade 3 being the worst, are directly associated with worst survival rate (Kaplan-Meier Curve)
- Larger tumor size is associated with a slightly increased risk of worse survival (Cox PH Model)
- A higher number of lymph nodes involved is associated with a significantly increased risk of worse survival (Cox Ph Model)
- Higher (log-transformed) progesterone receptor levels are associated with significantly better survival (Cox Ph Model)
- Variables like tumor size, tumor_grade, nodes, and log_prog_recp showed good fit
- The Log-normal model has the lowest AIC and BIC, suggesting it provides the best balance between fit and complexity and Weibull indicating slightly better predictive ability