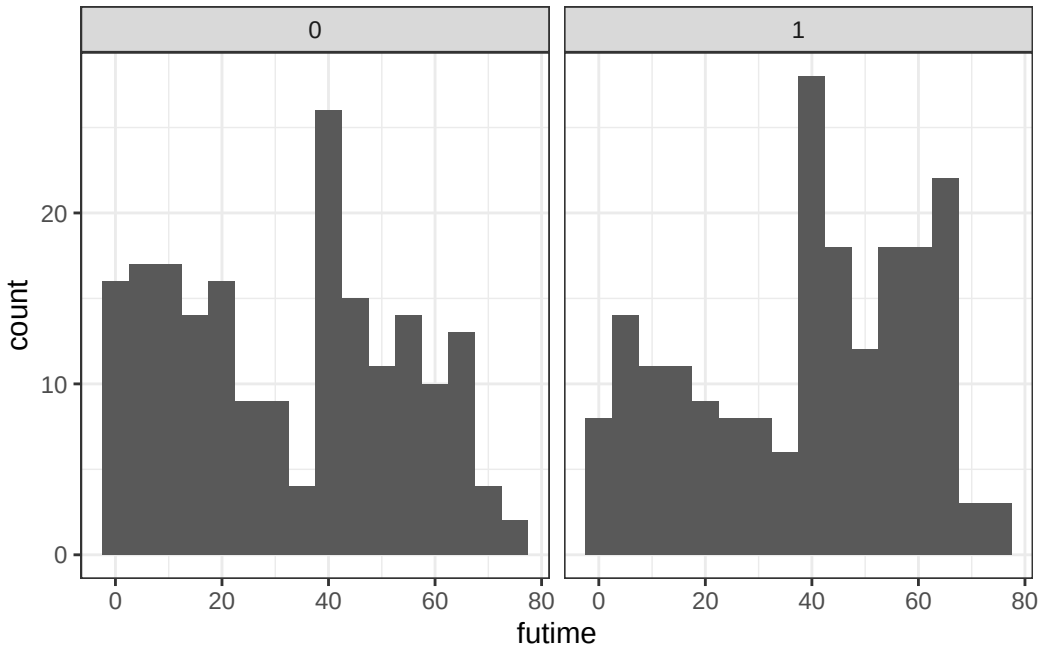
Nowacek HW 5-4

Data import

```
data <- read.csv("https://www2.stat.duke.edu/courses/Fall24/sta490.01/data/hw5.csv")

data |>
  ggplot(aes(x = futime)) +
  geom_histogram(binwidth = 5) +
  facet_wrap(~ trt) +
  theme_bw()
```



Is there evidence that laser treatment is beneficial in addressing blindness among patients with high risk diabetic retinopathy, controlling for age at diabetes diagnosis, laser type, and risk score?

To answer this question, fit a parametric AFT model assuming that time-to-blindness has an **exponential** distribution, specifically interpret a relevant estimate from your AFT model, and conduct an appropriate hypothesis test at the $\alpha = 0.05$ level (provide the null and alternative hypotheses, the distribution of the relevant test statistic under the null hypothesis, and your conclusion in context of the problem).

```
fit <- survreg(Surv(futime, status) ~ trt + risk + age + laser,
               data = data, dist = "exponential")
summary(fit)
```

```
Call:
survreg(formula = Surv(futime, status) ~ trt + risk + age + laser,
    data = data, dist = "exponential")
              Value Std. Error     z       p
(Intercept)  5.64856    0.58104  9.72 < 2e-16
trt          0.83400    0.16910  4.93 8.1e-07
risk        -0.15267    0.05548 -2.75  0.0059
age         -0.00625    0.00548 -1.14  0.2544
laserxenon   0.20855    0.16133  1.29  0.1961

Scale fixed at 1

Exponential distribution
Loglik(model)= -835.8   Loglik(intercept only)= -853.2
    Chisq= 34.8 on 4 degrees of freedom, p= 5.1e-07
Number of Newton-Raphson Iterations: 5
n= 394
```

Null hypothesis: We **do not** expect to see a difference in log expected survival time for those patients who received treatment and those who did not.

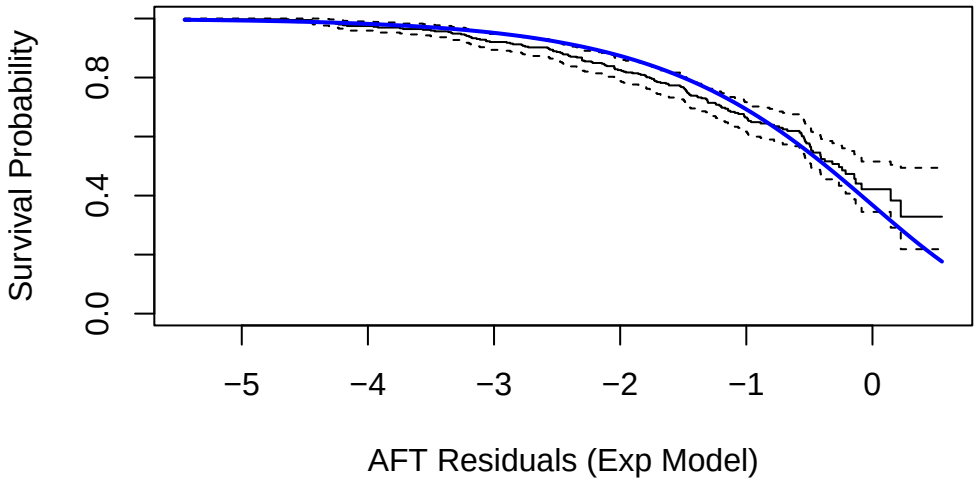
Alternative hypothesis: We **do** expect to see a difference in log expected survival time for those patients who received treatment and those who did not.

Under the null hypothesis, the distribution of survival times is expected to be exponential.

Since the coefficient of the treatment variable is significant at the $\alpha = 0.05$ level, we reject the null hypothesis, there is an association between treatment and survival times.

```
resids <- (log(data$futime) - fit$linear.predictors) / 1
m1 <- survfit(Surv(resids, status) ~ 1, data = data)
plot(m1, xlab = "AFT Residuals (Exp Model)", ylab = "Survival Probability")
```

```r
exp.x <- seq(min(resids), max(resids), length = 100)
exp.y <- exp(-exp(exp.x)) # F(t)
lines(exp.x, exp.y, col = "blue", lwd = 2)
```



Due to the deviation of the residual plot from the Kaplan_Meier estimate, we can not say that the expectation of an exponential distribution of the failure times is met.