

Blood Storage

One way anova test Assumptions 1) The samples are independent.

This assumption is reasonably met by the nature of the data. Each sample represents and independent individual patient. Any conditions for one patient would not impact the conditions of another patient.

2) Outcomes within groups are normally distributed

```
library(tidyverse)
```

```
-- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
v dplyr      1.1.4      v readr      2.1.5
v forcats    1.0.0      v stringr    1.5.1
v ggplot2    3.5.2      v tibble     3.3.0
v lubridate  1.9.4      v tidyr      1.3.1
v purrr      1.1.0
```

```
-- Conflicts ----- tidyverse_conflicts() --
```

```
x dplyr::filter() masks stats::filter()
```

```
x dplyr::lag()     masks stats::lag()
```

```
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become
```

```
blood <- read.csv("BloodStorage.csv")
```

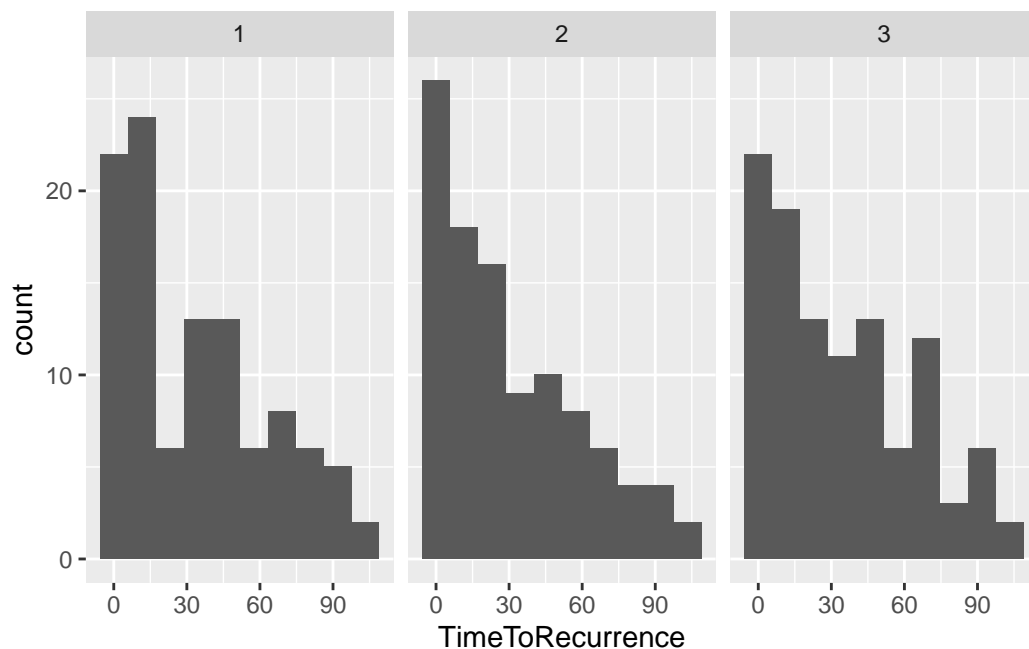
```
library(ggplot2)
```

```
ggplot(data = blood, aes(x = TimeToRecurrence)) +
```

```
  geom_histogram(bins = 10) +
```

```
  facet_grid(~ RBC.Age.Group)
```

```
Warning: Removed 1 row containing non-finite outside the scale range
(`stat_bin()`).
```

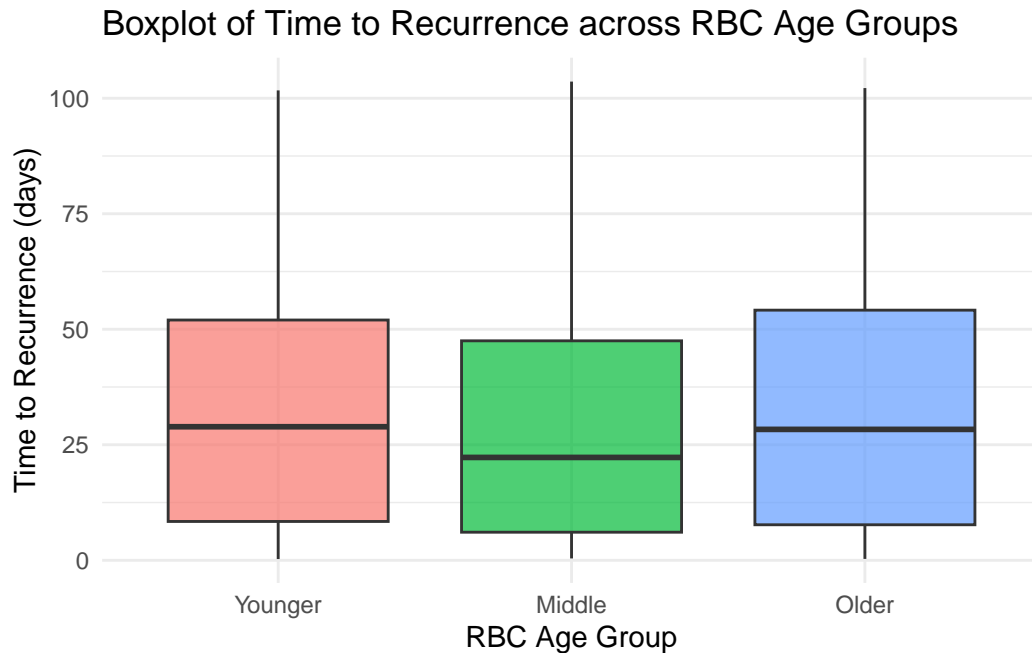


```
# Recode the numeric group variable into labeled factors
blood$RBC.Age.Group <- factor(blood$RBC.Age.Group,
                              levels = c(1, 2, 3),
                              labels = c("Younger", "Middle", "Older"))
```

This assumption is not met. Outcomes between groups of Blood Age (Younger, Middle, Older) are not normally distributed. Rather, they are right skewed in each group.

3) The variance is the same within all groups

```
Warning: Removed 1 row containing non-finite outside the scale range
(`stat_boxplot()`).
```



This assumption of anova is met. The variance across groups is nearly identical.

```
anova_result <- aov(TimeToRecurrence ~ RBC.Age.Group, data = blood)
summary(anova_result)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
RBC.Age.Group	2	771	385.5	0.471	0.625
Residuals	312	255329	818.4		

1 observation deleted due to missingness

Linear regression

Conditions:

Independence:

Checking linearity and equal variance

```
lm_mod <- lm(TimeToRecurrence ~ RBC.Age.Group, data = blood)
summary(lm_mod)
```

Call:

```
lm(formula = TimeToRecurrence ~ RBC.Age.Group, data = blood)
```

Residuals:

Min	1Q	Median	3Q	Max
-33.766	-25.563	-7.366	18.143	72.925

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	33.9482	2.7918	12.160	<2e-16 ***
RBC.Age.GroupMiddle	-3.2736	3.9673	-0.825	0.410
RBC.Age.GroupOlder	0.1181	3.9297	0.030	0.976

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

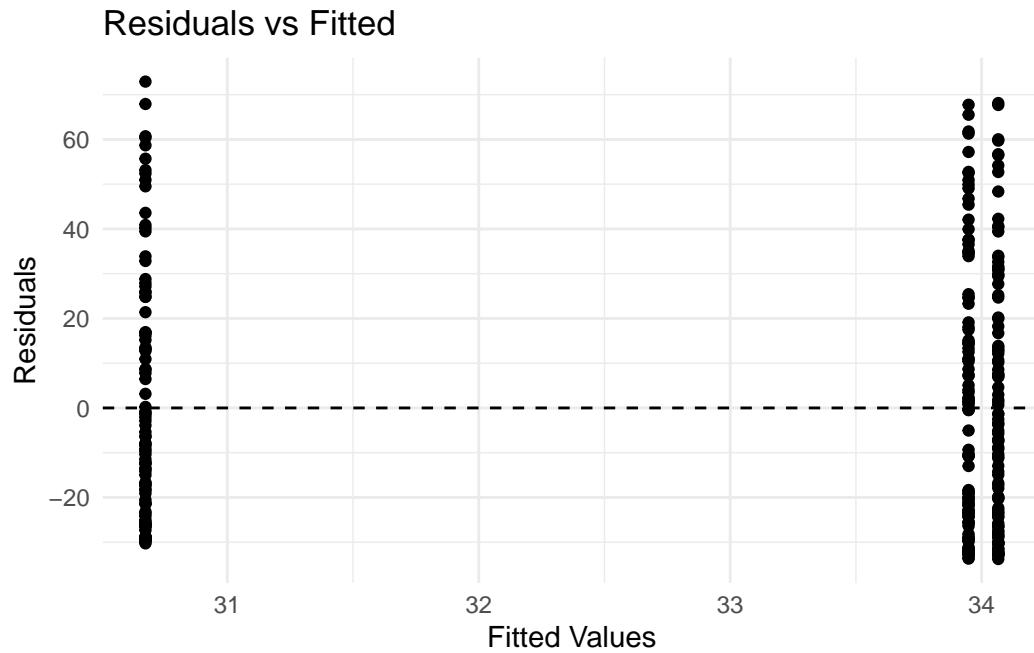
Residual standard error: 28.61 on 312 degrees of freedom

(1 observation deleted due to missingness)

Multiple R-squared: 0.00301, Adjusted R-squared: -0.003381

F-statistic: 0.471 on 2 and 312 DF, p-value: 0.6248

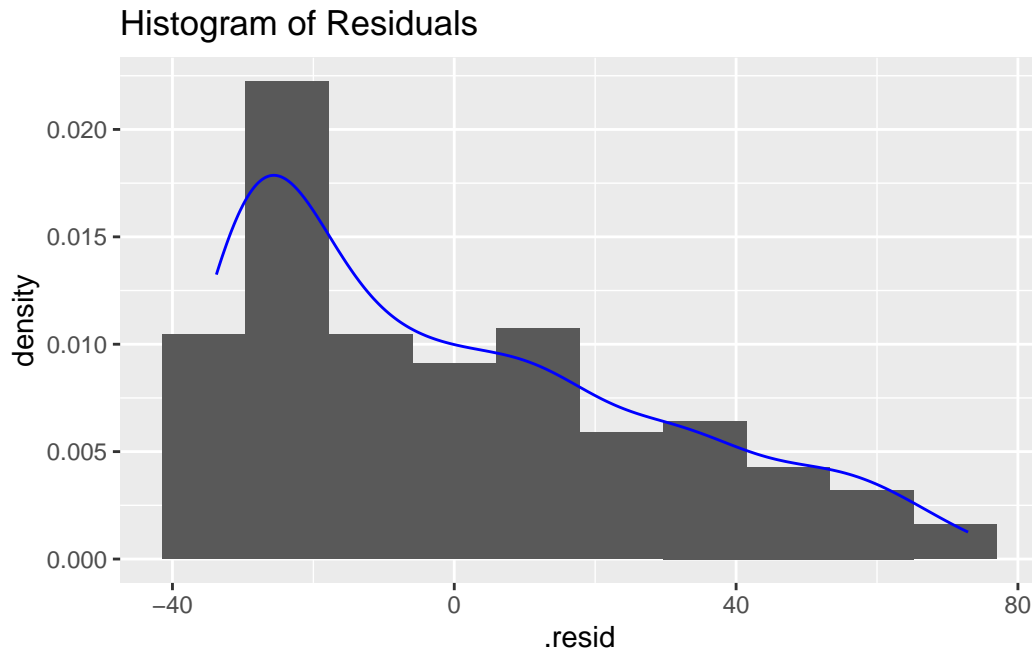
```
ggplot(lm_mod, aes(x = .fitted, y = .resid)) +  
  geom_point() +  
  geom_hline(yintercept = 0, linetype = "dashed") +  
  labs(  
    title = "Residuals vs Fitted",  
    x = "Fitted Values",  
    y = "Residuals"  
  ) +  
  theme_minimal()
```



Checking normality of residuals

```
ggplot(lm_mod, aes(.resid)) +  
  geom_histogram(bins = 10, aes(y = ..density..)) +  
  geom_density(color = "blue") +  
  labs(title = "Histogram of Residuals")
```

Warning: The dot-dot notation (`..density..`) was deprecated in ggplot2 3.4.0.
i Please use `after_stat(density)` instead.



Running the regression

```
# Make sure you have clean data
blood_clean <- blood %>% drop_na(TimeToRecurrence, RBC.Age.Group, Age)

lm_mod <- lm(TimeToRecurrence ~ RBC.Age.Group + Age, data = blood_clean)

# View results
summary(lm_mod)
```

Call:

```
lm(formula = TimeToRecurrence ~ RBC.Age.Group + Age, data = blood_clean)
```

Residuals:

Min	1Q	Median	3Q	Max
-34.503	-25.434	-6.864	18.513	72.294

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	20.7682	14.0103	1.482	0.139
RBC.Age.GroupMiddle	-3.2836	3.9678	-0.828	0.409

RBC.Age.GroupOlder	0.3474	3.9374	0.088	0.930
Age	0.2143	0.2232	0.960	0.338

Residual standard error: 28.61 on 311 degrees of freedom

Multiple R-squared: 0.005956, Adjusted R-squared: -0.003633

F-statistic: 0.6211 on 3 and 311 DF, p-value: 0.6018