

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 an independent individual patient. Any conditions for one patient would not impact the conditions of another patient.

2) Outcomes within groups are normally distributed

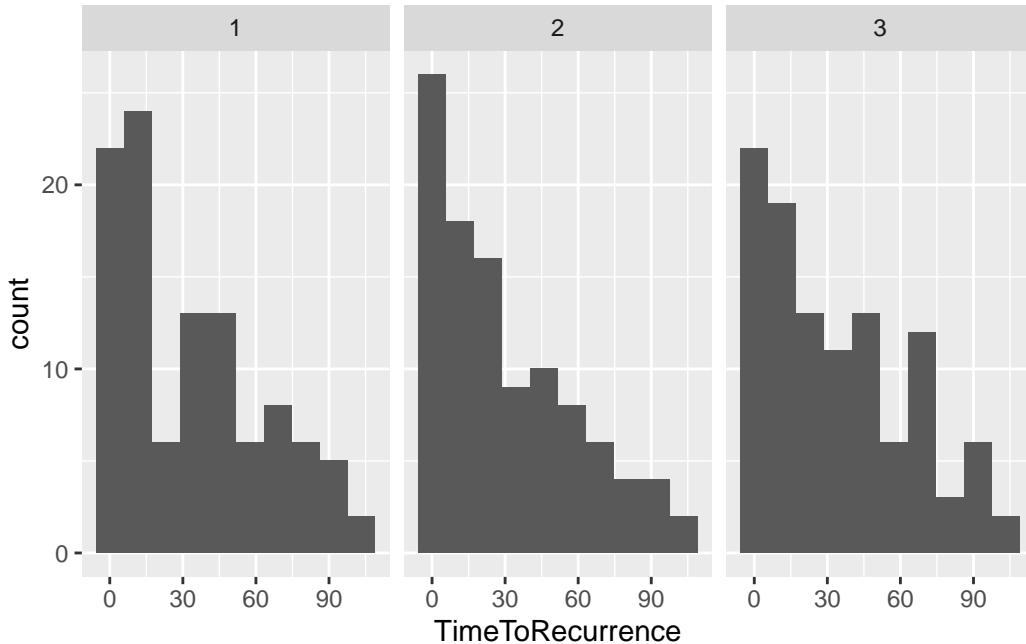
```
library(tidyverse)
```

```
-- 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()`).
```



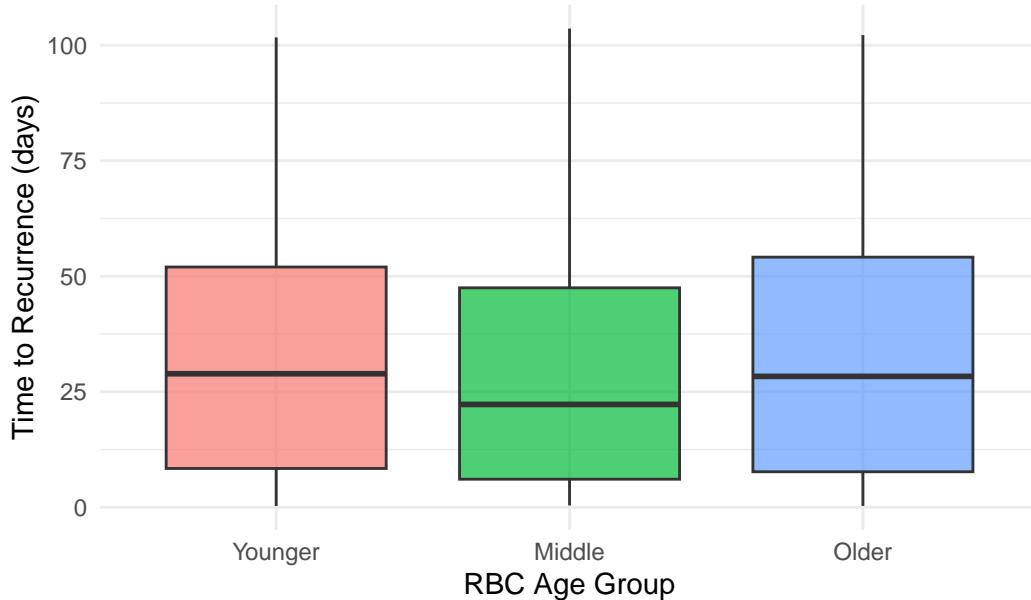
```
# Recode the numeric group variable into labeled factors
blood$RBC.Age.Group <- factor(blood$RBC.Age.Group,
                                levels = c(1, 2, 3),
                                labels = c("Younger", "Middle", "Older"))
```

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()`).
```

Boxplot of Time to Recurrence across RBC Age Groups



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: The independence condition appears to be met, as each observation in the Blood Storage dataset represents a separate patient, and there's no indication that the measurements influence one another. Therefore, the residuals can reasonably be assumed to be independent.

Checking linearity and equal variance

```

blood$AA <- factor(blood$AA, labels = c("Non-AA", "AA"))
blood$FamHx <- factor(blood$FamHx, labels = c("No", "Yes"))
blood$T.Stage <- factor(blood$T.Stage)
blood$bGS <- factor(blood$bGS)
blood$AnyAdjTherapy <- factor(blood$AnyAdjTherapy, labels = c("No", "Yes"))
blood$AdjRadTherapy <- factor(blood$AdjRadTherapy, labels = c("No", "Yes"))

lm_mod_full <- lm(TimeToRecurrence ~ RBC.Age.Group + Age + AA + FamHx +
                    PVol + TVol + T.Stage + bGS + PreopPSA + Units +
                    AnyAdjTherapy + AdjRadTherapy, data = blood)

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

```

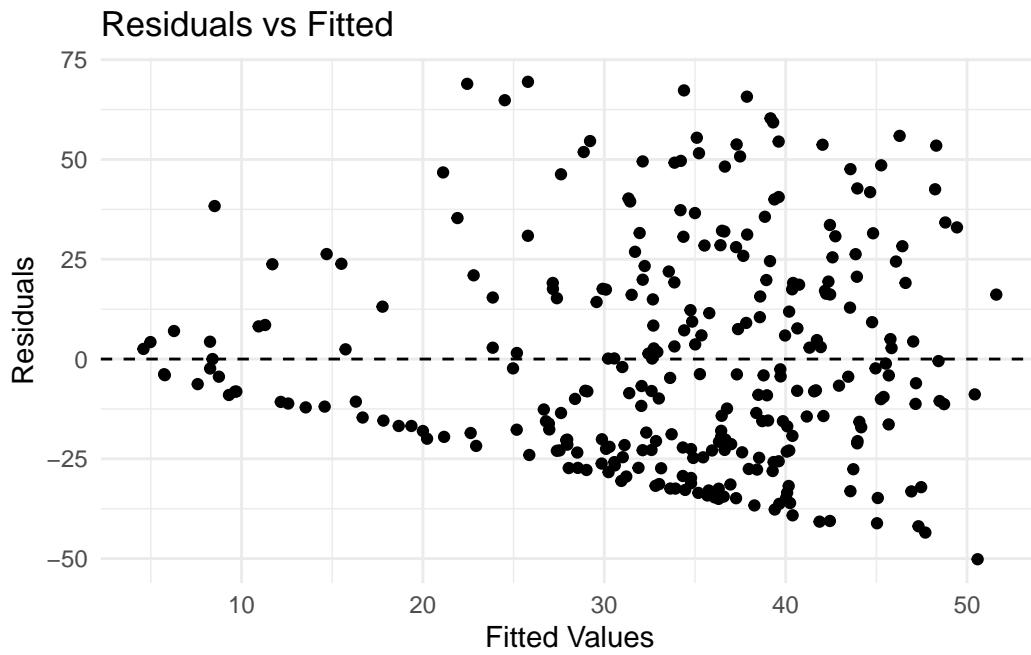
ggplot(lm_mod_full, aes(x = .fitted, y = .resid)) +
  geom_point() +
  geom_hline(yintercept = 0, linetype = "dashed") +
  labs(
    title = "Residuals vs Fitted",

```

```

x = "Fitted Values",
y = "Residuals"
) +
theme_minimal()

```



Interpretation: INSERT SMTH HERE

Checking normality of residuals

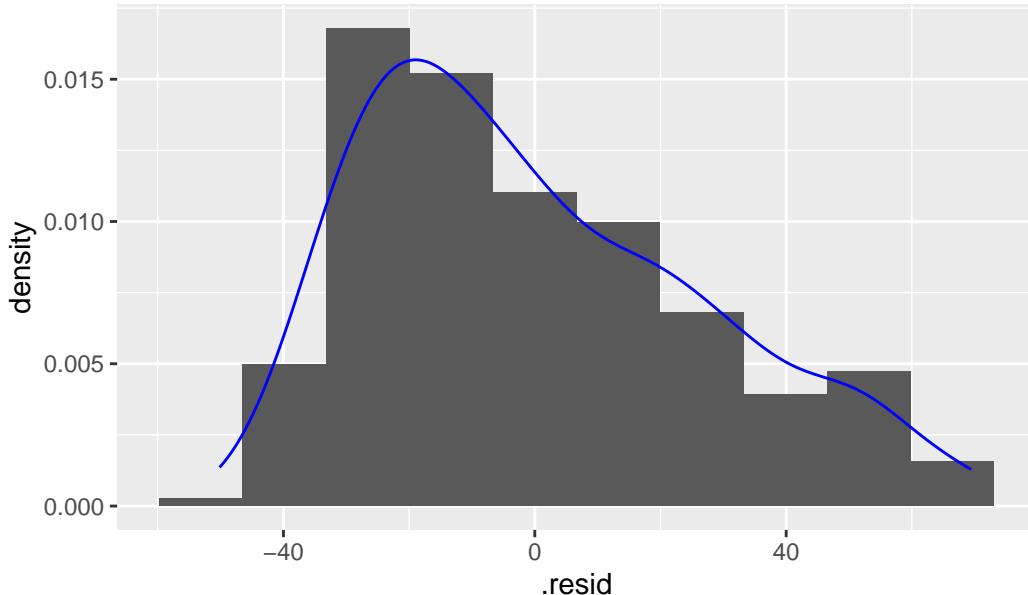
```

ggplot(lm_mod_full, 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.

Histogram of Residuals



Interpretation: Normality is not met, as the plot is right skewed with most residuals clustered near 0 and a longer tail extending to the right. This suggests that the residuals are not normally distributed.

Running the regression

```
summary(lm_mod_full)
```

Call:

```
lm(formula = TimeToRecurrence ~ RBC.Age.Group + Age + AA + FamHx +
    PVol + TVol + T.Stage + bGS + PreopPSA + Units + AnyAdjTherapy +
    AdjRadTherapy, data = blood)
```

Residuals:

Min	1Q	Median	3Q	Max
-50.154	-21.657	-6.057	18.845	69.459

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	34.05392	15.91447	2.140	0.0333 *
RBC.Age.GroupMiddle	-5.15855	4.18655	-1.232	0.2189
RBC.Age.GroupOlder	0.41949	4.05473	0.103	0.9177

Age	0.18619	0.23964	0.777	0.4379
AAAA	-7.95593	4.44940	-1.788	0.0749 .
FamHxYes	-6.59899	4.04019	-1.633	0.1036
PVol	0.07948	0.06419	1.238	0.2167
TVol	-0.84280	2.68914	-0.313	0.7542
T.Stage2	-4.66279	6.15841	-0.757	0.4496
bGS2	-7.34024	4.04438	-1.815	0.0706 .
bGS3	-24.30547	6.15624	-3.948	0.0001 ***
PreopPSA	-0.13934	0.31418	-0.443	0.6578
Units	-1.46836	0.93400	-1.572	0.1171
AnyAdjTherapyYes	10.73943	11.27707	0.952	0.3418
AdjRadTherapyYes	-8.58889	28.65729	-0.300	0.7646

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

Residual standard error: 27.77 on 272 degrees of freedom

(29 observations deleted due to missingness)

Multiple R-squared: 0.1213, Adjusted R-squared: 0.0761

F-statistic: 2.683 on 14 and 272 DF, p-value: 0.001064