

Test Normality Assumption for ANOVA

Tidy Data

8/14/2022

The following is to generate the data.

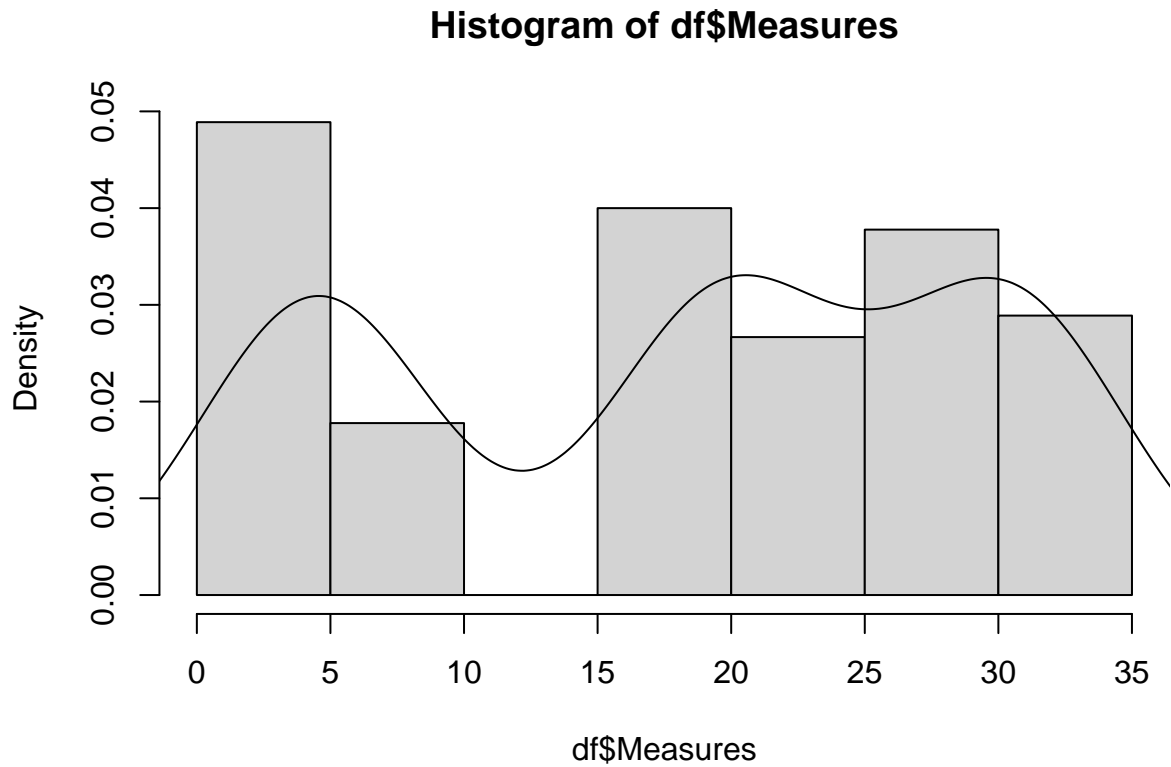
```
set.seed(12)
number_x=30

df <- data.frame(
  Levels=factor(rep(c("X_1", "X_2", "X_3"), each=number_x)),
  Measures=round(c(rnorm(number_x, mean=5, sd=2),
                    rnorm(number_x, mean=20, sd=2),
                    rnorm(number_x, mean=30, sd=2)))
)
head(df)
```

```
##   Levels Measures
## 1    X_1         2
## 2    X_1         8
## 3    X_1         3
## 4    X_1         3
## 5    X_1         1
## 6    X_1         4
```

We can plot to see how it looks like.

```
hist(df$Measures, freq = FALSE)
lines(density(df$Measures))
```



We can test whether it is normally distributed.

```
shapiro.test(df$Measures)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  df$Measures
## W = 0.88042, p-value = 5.956e-07
```

How to test normality for one-way ANOVA

$$\bar{x} = \sum_{i=1}^k \sum_{j=1}^{n_i} x_{ij}$$

$$\bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij}$$

$$\epsilon_{ij} = x_{ij} - \bar{x}_i$$

In the following, we can use the `aov()` function to calculate the residual epsilon. Then, use Shapiro test to test the normality.

```
model=aov(df$Measures~df$Levels)
res_epsilon=model$residuals
shapiro.test(res_epsilon)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  res_epsilon
## W = 0.98538, p-value = 0.4124
```

We can also plot the residual epsilon.

```
hist(res_epsilon, freq =FALSE)
lines(density(res_epsilon))
```

