

# Age and Mating Success in Male Elephants

Parikshit Patil

2024-09-01

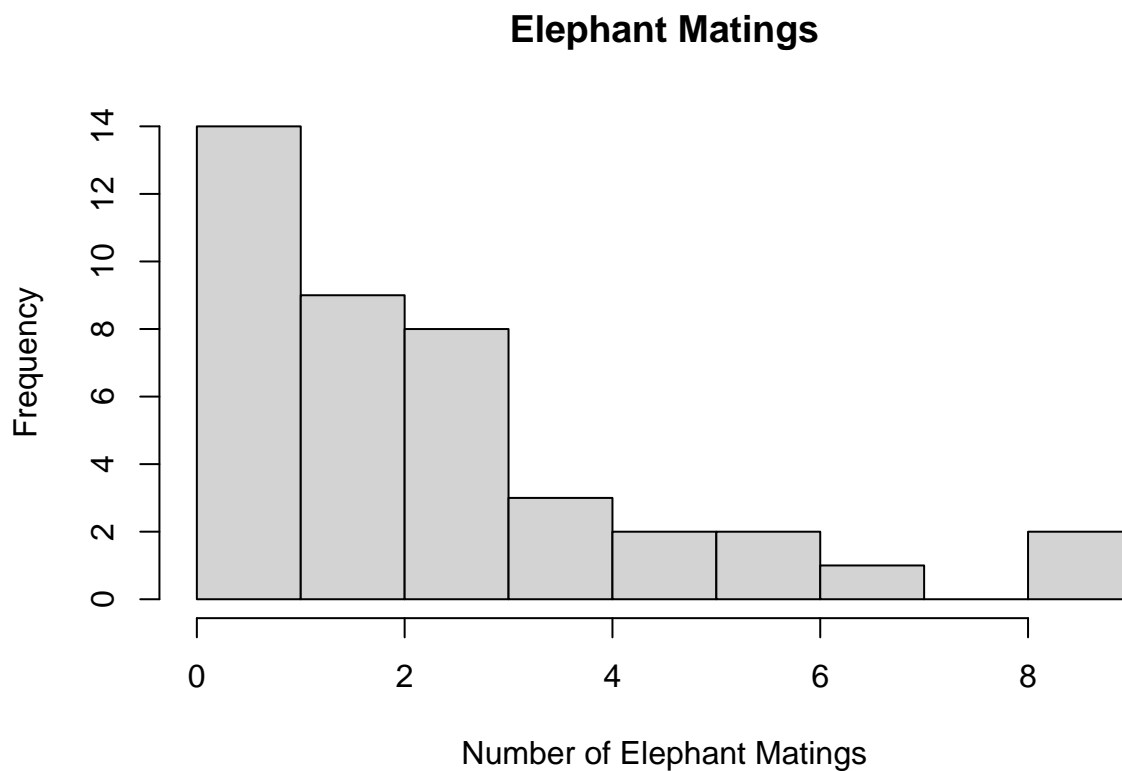
## Age and Mating Success in Male Elephants

```
elephant <- read.csv("elephant.csv")  
str(elephant)
```

```
## 'data.frame':  41 obs. of  2 variables:  
## $ AGE      : num  27 28 28 28 28 29 29 29 29 29 ...  
## $ MATINGS: num  0 1 1 1 3 0 0 0 2 2 ...
```

a)

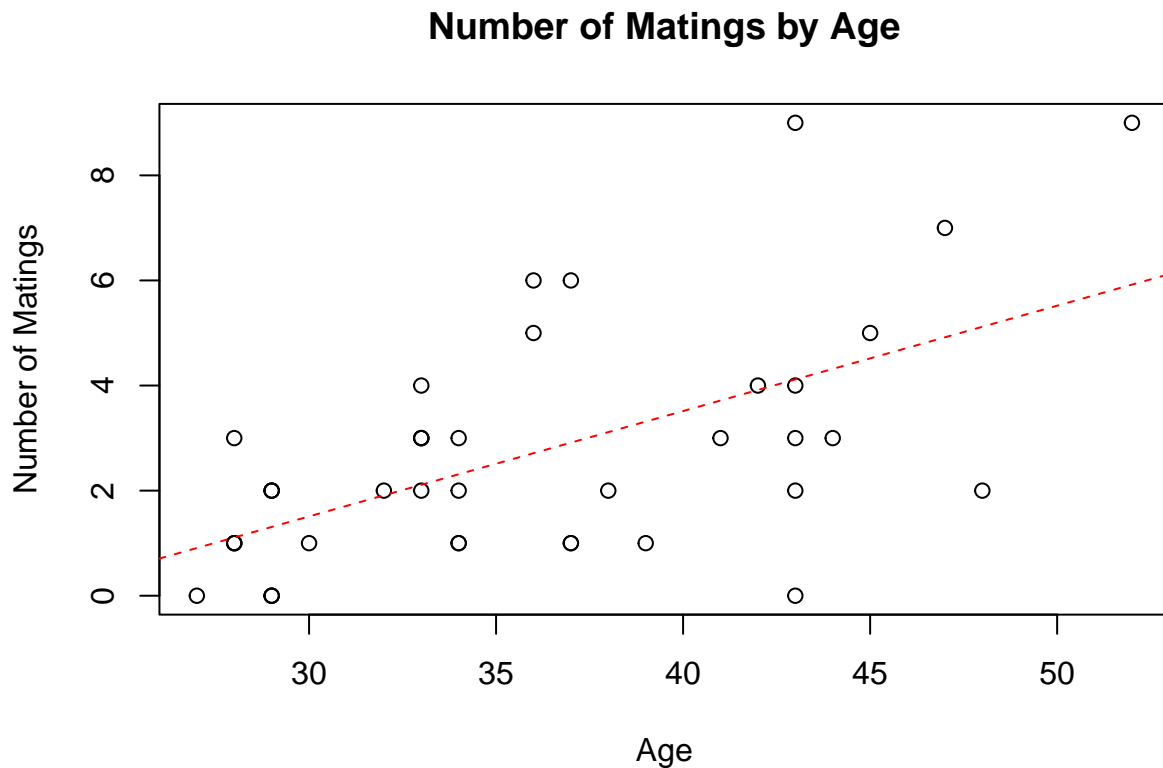
```
#histogram for elephant matings  
hist(elephant$MATINGS, main = "Elephant Matings", xlab = "Number of Elephant Matings")
```



By observing the histogram, it resembles a poisson distribution with a peak at 0 and right skewed tail, without too much variability to this trend. This suggests that this data can be modeled well as a poisson response.

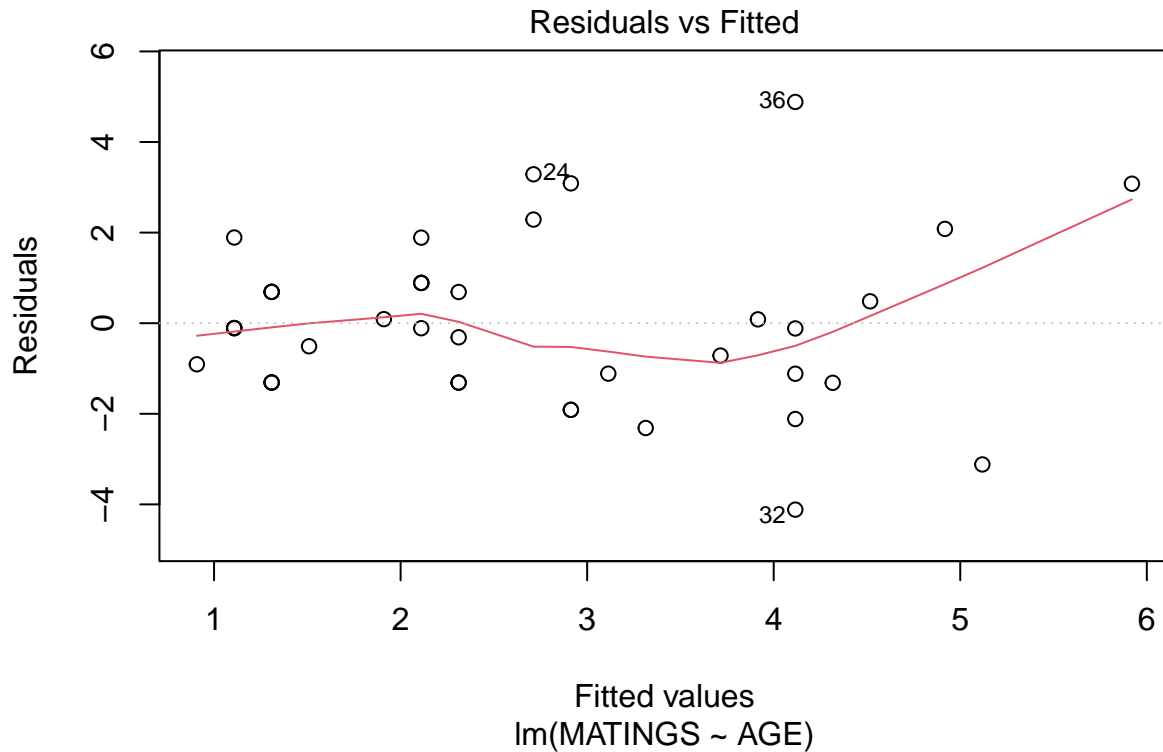
b)

```
#matings by age
plot(elephant$AGE, elephant$MATINGS, xlab = "Age", ylab = "Number of Matings", main = "Number of Matings by Age")
#least squares line
abline(lm(MATINGS ~ AGE, data = elephant), col = "red", lty = 2)
```



```
#Linear regression model
lm_model_elephant = lm(MATINGS ~ AGE, data = elephant)

# Residual Plot
plot(lm_model_elephant, which = 1) # Residuals vs Fitted
```



Although the linearity assumptions seem to be valid, it appears that the equal variance assumption is violated. This suggests that a linear regression for modelling mating counts with age might not be appropriate.

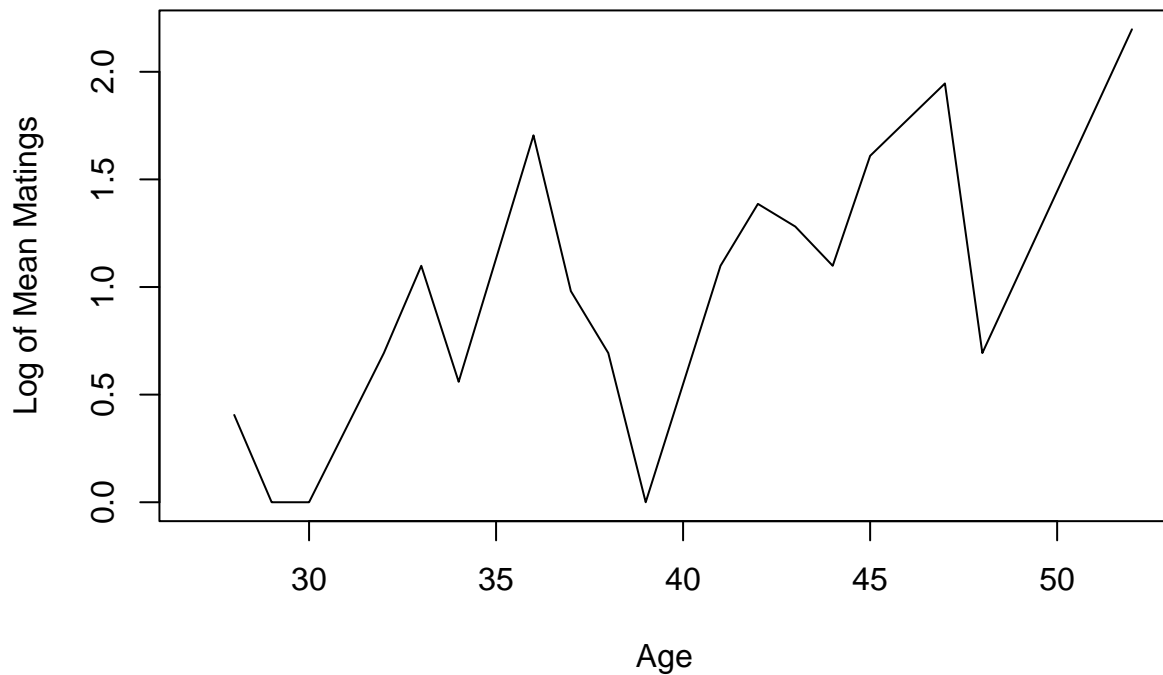
c)

```
# Calculate the mean number of matings for each age
mean_matings = tapply(elephant$MATINGS, elephant$AGE, mean)

# Take the log of each mean
log_mean_matings = log(mean_matings)

# Plot log of mean matings by AGE
plot(names(log_mean_matings), log_mean_matings, type = "l", xlab = "Age", ylab = "Log of Mean Matings",
```

## Log of Mean Matings by Age



The assumption being assessed with this assumption is the linearity assumption between age and the log mean number of matings.

There is no evidence of a quadratic trend in this graph, since there are only straight lines.

d)

```
#poisson model
elephant_model_poisson = glm(MATINGS ~ AGE, data = elephant, family = poisson)

summary(elephant_model_poisson)
```

```
##
## Call:
## glm(formula = MATINGS ~ AGE, family = poisson, data = elephant)
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.58201    0.54462  -2.905  0.00368 **
## AGE          0.06869    0.01375   4.997 5.81e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
## Null deviance: 75.372  on 40  degrees of freedom
```

```
## Residual deviance: 51.012  on 39  degrees of freedom
## AIC: 156.46
##
## Number of Fisher Scoring iterations: 5
```

The coefficient for AGE = 0.06869. This means that for each unit increase in AGE, the log of MATINGS increase by 0.0686.

e)

```
#confidence model
CI = confint(elephant_model_poisson, "AGE")
```

```
## Waiting for profiling to be done...
```

```
CI
```

```
##      2.5 %      97.5 %
## 0.04167776 0.09563762
```

The 95% confidence interval for the slope (AGE) is 0.0416, 0.0956. This means that we are 95% sure that one year increase in AGE will result in an increase of log of MATINGS by a value between 0.0416 and 0.0956.

f)

```
#make prediction
predict.glm(elephant_model_poisson, newdata = data.frame(AGE = 31), type = "response")
```

```
##      1
## 1.728872
```

```
#exponent of confidence intervals
exp(CI)
```

```
##      2.5 %      97.5 %
## 1.042558 1.100360
```

The predicted mating rate for a 31 year old elephant is 1.7288 times a year.

The 95% confidence interval is 1.0425, 1.1003

g)

```
#drop-in deviance test for relation between matings and age
null_model <- glm(MATINGS ~ 1, family = poisson, data = elephant)
anova(elephant_model_poisson, null_model, test='Chisq')
```

```
## Analysis of Deviance Table
##
## Model 1: MATINGS ~ AGE
## Model 2: MATINGS ~ 1
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1         39      51.012
## 2         40      75.372 -1    -24.36 7.991e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Since p-value is quite small, we reject the null hypothesis meaning that there is a significant relationship between the number of matings and age.

h)

```
#poisson model with quadratic term
elephant_model_poisson_quad = glm(MATINGS ~ AGE + I(AGE^2), data = elephant, family = poisson)

anova(elephant_model_poisson_quad, elephant_model_poisson, test='Chisq')
```

```
## Analysis of Deviance Table
##
## Model 1: MATINGS ~ AGE + I(AGE^2)
## Model 2: MATINGS ~ AGE
##   Resid. Df Resid. Dev Df Deviance Pr(>Chi)
## 1         38      50.826
## 2         39      51.012 -1    -0.18544  0.6667
```

The p-value for the quadratic model is worse than that of the linear model. Therefore the quadratic model is not preferred.

I)

```
#pchisq test
pchisq(elephant_model_poisson$deviance, df.residual(elephant_model_poisson), lower = FALSE)
```

```
## [1] 0.09426231
```

For the model with a linear term for AGE as the only predictor, the model does not fit the data well as indicated by the high deviance value.

j)

```
#quassi poisson model
elephant_model_quasi_poisson = glm(MATINGS ~ AGE, data = elephant, family = quasipoisson)
summary(elephant_model_quasi_poisson)
```

```
##
## Call:
## glm(formula = MATINGS ~ AGE, family = quasipoisson, data = elephant)
```

```
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.58201    0.58590  -2.700   0.0102 *
## AGE          0.06869    0.01479   4.645 3.81e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasipoisson family taken to be 1.157334)
##
## Null deviance: 75.372  on 40  degrees of freedom
## Residual deviance: 51.012  on 39  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 5
```

The estimated coefficients of the quasipoisson model remains the same as the poisson model.

The standard errors for the quasipoisson model are slightly higher than that of the poisson model.

The estimated dispersion parameter = 1.157334.

When adjusting for overdispersion, it is less likely to obtain a significant result when testing coefficients. This is due to the increased uncertainty associated with the coefficients which reduces the power of the tests to detect significant effects.