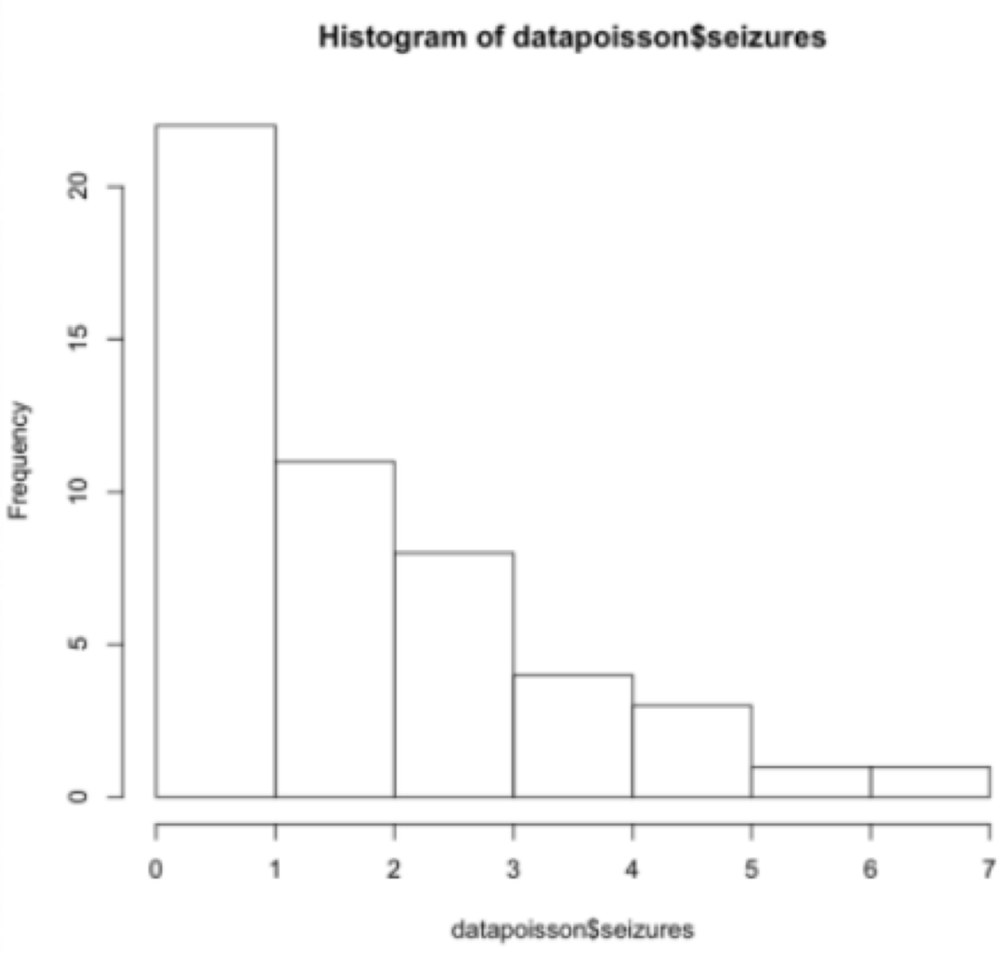
Exercise resolution

- A researcher conducted a clinical trial of 50 patients with epilepsy, 25 of whom were randomized to receive the new anticonvulsant drug (experimental group) and 25 to receive the old one (control group). The group (experimental or control group), age and outcome were recorded. The outcome variable was seizure counts that occurred during the eight-week follow-up period.

We want to know whether the new drug increases or decreases the number of seizures when adjusted for age.

- Let's see the distribution of the number of seizures with a histogram

```
hist(datapoisson$seizures)
```



- Outcome: count → Poisson Regression

```
poisson <- glm(datapoisson$seizures ~ datapoisson$age + datapoisson$group, data = datapoisson, family = "poisson")
summary(poisson)
```

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.3308  -0.4922  -0.4563   0.3505   2.2198

Coefficients:
                            Estimate Std. Error z value Pr(>|z|)
(Intercept)                 0.347514   0.678376   0.512  0.60846
datapoisson$age             0.003112   0.021099   0.148  0.88273
datapoisson$groupnew drug 0.561495   0.204292   2.748  0.00599 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 61.808  on 49  degrees of freedom
Residual deviance: 53.827  on 47  degrees of freedom
AIC: 177.48

Number of Fisher Scoring iterations: 5
```

$$\ln(\mu_y) = 0.347514 + 0.003112\, age + \ 0.561495\, new\ group$$

- Outcome: count ⟶ Poisson Regression

```
exp(cbind(RR = coef(poisson), confint(poisson)))
```

	RR	2.5 %	97.5 %
(Intercept)	1.415545	0.3667683	5.258693
datapoisson$age	1.003117	0.9626480	1.045765
datapoisson$groupnew drug	1.753292	1.1809467	2.637163

The new drug increases the number of seizures by 1.8 times even adjusted for age.

But we can even predict the average number of seizures over an eight-year period weeks in a patient, for example, 30 years old with epilepsy treated with the new drug:

$$\ln(\mu_y) = 0.347514 + 0.003112\, age + 0.561495\, new\ group$$

$$e^{\ln(\mu_y)} = \mu_y = e^{0.347514 + 0.003112 \times age + 0.561495 \times new\ drug\ group}$$

$$\mu_y = e^{0.347514 + 0.003112 \times 30 + 0.561495 \times 1}$$

$$\mu_y = e^{0.561495} = 2{,}7$$

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.3308  -0.4922  -0.4563   0.3505   2.2198

Coefficients:
                         Estimate Std. Error z value Pr(>|z|)
(Intercept)              0.347514   0.678376   0.512  0.60846
datapoisson$age          0.003112   0.021099   0.148  0.88273
datapoisson$groupnew drug 0.561495   0.204292   2.748  0.00599 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 61.808  on 49  degrees of freedom
Residual deviance: 53.827  on 47  degrees of freedom
AIC: 177.48

Number of Fisher Scoring iterations: 5

                               RR     2.5 %   97.5 %
(Intercept)              1.415545 0.3667683 5.258693
datapoisson$age          1.003117 0.9626480 1.045765
datapoisson$groupnew drug 1.753292 1.1809467 2.637163
```

But we can even predict the average number of seizures over an eight-year period weeks in a patient, for example, 30 years old with epilepsy treated with the old drug:

$$e^{\ln(\mu_y)} = \mu_y = e^{0.347514 + 0.003112 \times age + 0.561495 \times new\ drug\ group}$$

$$\mu_y = e^{0.347514 + 0.003112 \times 30 + 0.561495 \times 0}$$

$$\mu_y = e^{0.440874} = 1{,}5$$

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.3308  -0.4922  -0.4563   0.3505   2.2198

Coefficients:
                          Estimate Std. Error z value Pr(>|z|)
(Intercept)               0.347514   0.678376   0.512  0.60846
datapoisson$age           0.003112   0.021099   0.148  0.88273
datapoisson$groupnew drug 0.561495   0.204292   2.748  0.00599 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 61.808  on 49  degrees of freedom
Residual deviance: 53.827  on 47  degrees of freedom
AIC: 177.48

Number of Fisher Scoring iterations: 5

                                RR     2.5 %   97.5 %
(Intercept)               1.415545 0.3667683 5.258693
datapoisson$age           1.003117 0.9626480 1.045765
datapoisson$groupnew drug 1.753292 1.1809467 2.637163
```

Assumptions

- The observations are independente
- The rate is constant
- Mean=variance (right skewed)

Assumptions

- The observations are independente

If we assume that having a seizure doesn't increase the chances of having moreseizures the hypothesis is fulfilled; otherwise we could not use the Poisson model.

- The rate is constant

- Mean=variance (right skewed)

Assumptions

- The observations are independente

If we assume that having a seizure doesn't increase the chances of having moreseizures the hypothesis is fulfilled; otherwise we could not use the Poisson model.

- The rate is constant

If the likelihood of seizures decreased or increased over time,we could not use the Poisson model.

- Mean=variance (right skewed)

Assumptions

- The observations are independente

If we assume that having a seizure doesn't increase the chances of having moreseizures the hypothesis is fulfilled; otherwise we could not use the Poisson model.

- The rate is constant

If the likelihood of seizures decreased or increased over time,we could not use the Poisson model.

- Mean=variance (right skewed)

A common problem with Poisson regression is the problem of overdispersion i.e. variance > mean.

The problem of oversdispersion

- A common problem with Poisson regression is the problem of oversdispersion i.e. variance > mean.

- If there is more variability than expected, the estimated standard errors will be very low (exaggerated levels of accuracy) but the coefficients estimates are OK. The problema is only with the inference (confidence intervals and p values)

The problem of oversdispersion

- We can compute the dispersion

```
DISPERSION <- (sum(residuals(poisson,type="pearson")^2))/poisson$df.residual

DISPERSION
```

[1] 1.095602

- Dispersion above one is overdispersion.
- Underdispersion (below one) is less common, but it can exist.
- We want a dispersion close to one.

As rule we can say that a dispersion above 2 is when this problem of overdispersion becomes a real problem.

The problem of oversdispersion

- In this exercise overdispersion was not a problem but when it exists, a possible solution is to try to include more relevant independent variables in the model as this problem may be related to the fact that we failed to include many of the important independent variables for the model.

- If the problema was not the lack of important independent variables and overdispersion is still a problem, we can try to adjust a model using the quasi-likelihood function. In this way, the estimates will be the same but the standard errors will be adjusted (will be larger).

The problem of oversdispersion

quasipoisson <- glm(datapoisson$seizures ~ datapoisson$age + datapoisson$group, data = datapoisson, family = "quasipoisson")

poisson model

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.3308  -0.4922  -0.4563   0.3505   2.2198

Coefficients:
                           Estimate Std. Error z value Pr(>|z|)
(Intercept)                0.347514   0.678376   0.512  0.60846
datapoisson$age            0.003112   0.021099   0.148  0.88273
datapoisson$groupnew drug 0.561495   0.204292   2.748  0.00599 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 61.808  on 49  degrees of freedom
Residual deviance: 53.827  on 47  degrees of freedom
AIC: 177.48

Number of Fisher Scoring iterations: 5
```

quasi-poisson modle

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.3308  -0.4922  -0.4563   0.3505   2.2198

Coefficients:
                           Estimate Std. Error t value Pr(>|t|)
(Intercept)                0.347514   0.710063   0.489   0.6268
datapoisson$age            0.003112   0.022084   0.141   0.8885
datapoisson$groupnew drug 0.561495   0.213835   2.626   0.0116 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for quasipoisson family taken to be 1.095602)

    Null deviance: 61.808  on 49  degrees of freedom
Residual deviance: 53.827  on 47  degrees of freedom
AIC: NA

Number of Fisher Scoring iterations: 5
```