

Litter size and fertility: Learning from the data

Data

Parsing (PDF \rightarrow table)

See [here](#) for PDF to table conversion and parsing.

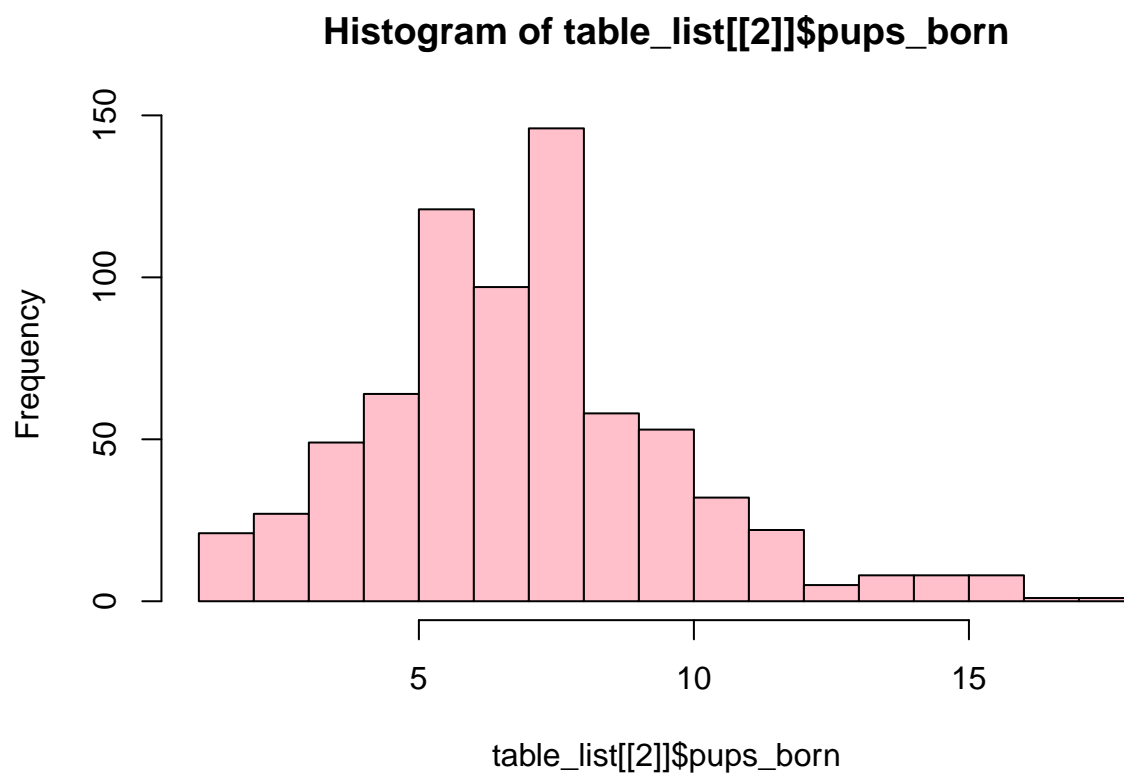
Load and clean the data

First of all, we load the litter size data (the number of pups born etc.) per strain from the already generated txt files, and then do a small quality check: Look at the distributions of the litter size (it should look like a Poisson, or at least be one-modal).

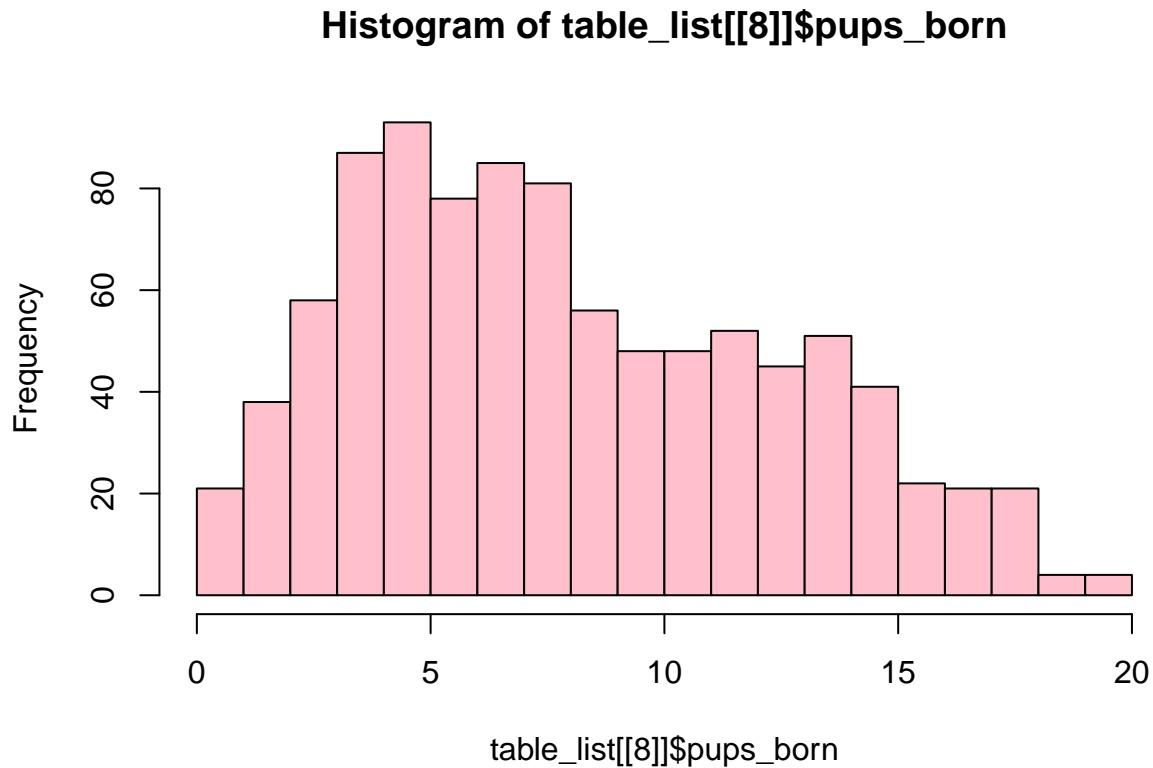
```
names(table_list)
```

```
## [1] "B6cBrd"      "B6D2F1"      "B6J_Cr1F"    "B6J_Fue"     "Balbc"
## [6] "DBA2_J_Fue"  "NMRI"        "CD1_99_10"   "CD1_10_20"
```

```
hist(table_list[[2]]$pups_born, breaks = 20, col = "pink")
```



```
hist(table_list[[8]]$pups_born, breaks = 25, col = "pink")
```



Quite often, only pups of one particular gender are required for the experiment. Therefore, all pups of the other gender are killed - and this is usually not stated in the data. Therefore, I filter all these cases out. Here is the number of litters per strain before (left) and after cleaning (right).

```
nlitters
```

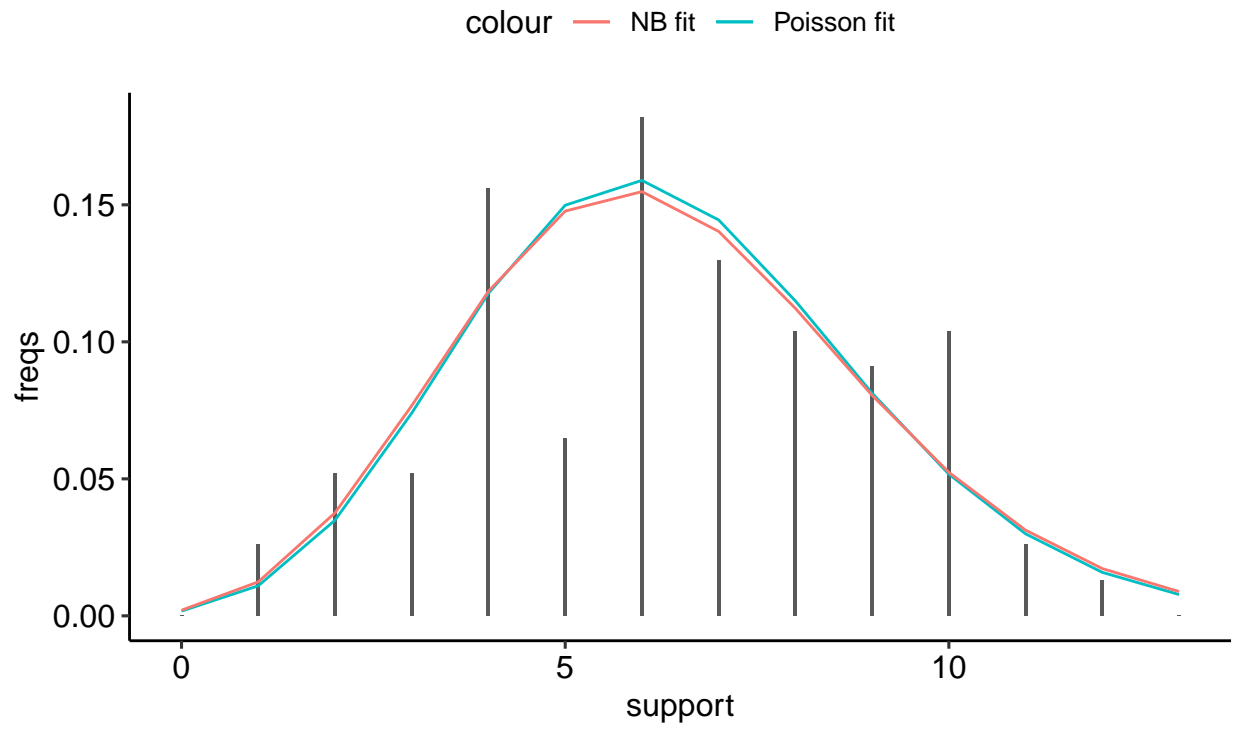
##	before	after
## B6cBrd	94	77
## B6D2F1	721	565
## B6J_CrlF	483	444
## B6J_Fue	1371	1219
## Balbc	139	116
## DBA2_J_Fue	204	178
## NMRI	387	375
## CD1_99_10	954	593
## CD1_10_20	992	280

Histograms with a Pois and NB fit for each strain

```
hists_list
```

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

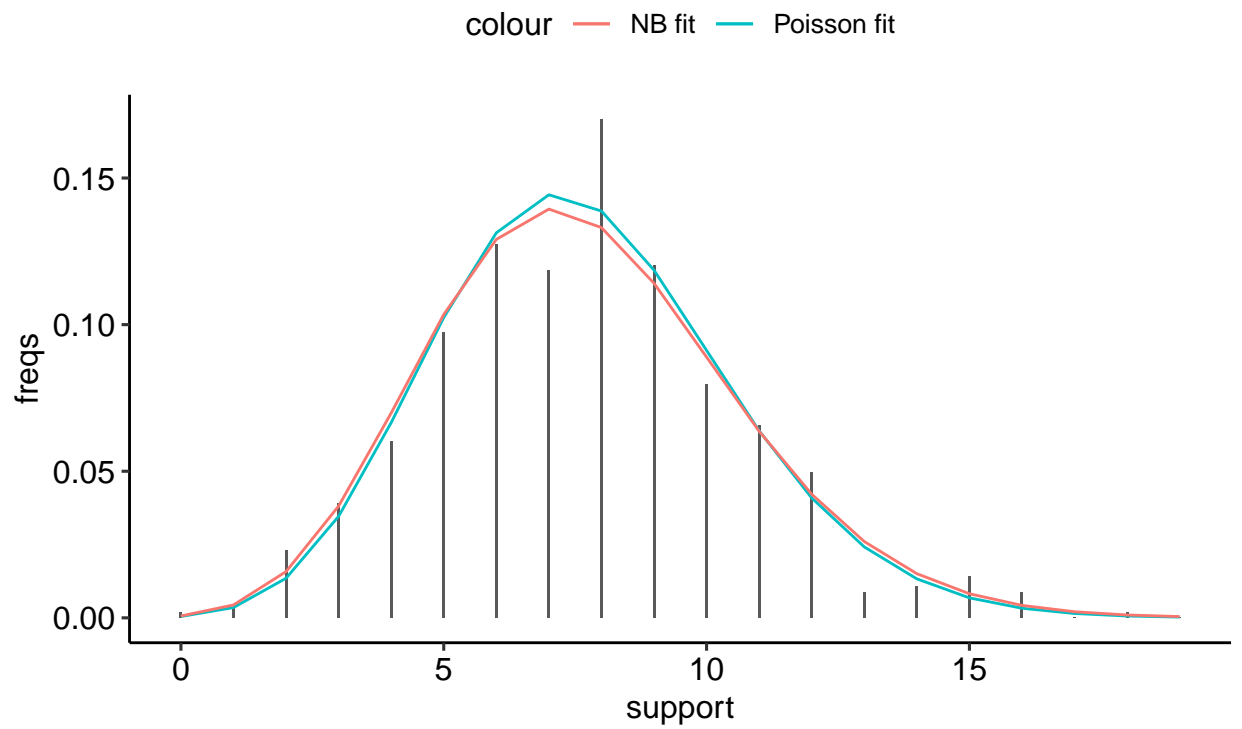
Strain: B6cBrd



Number of litters: 77

```
##  
## [[2]]
```

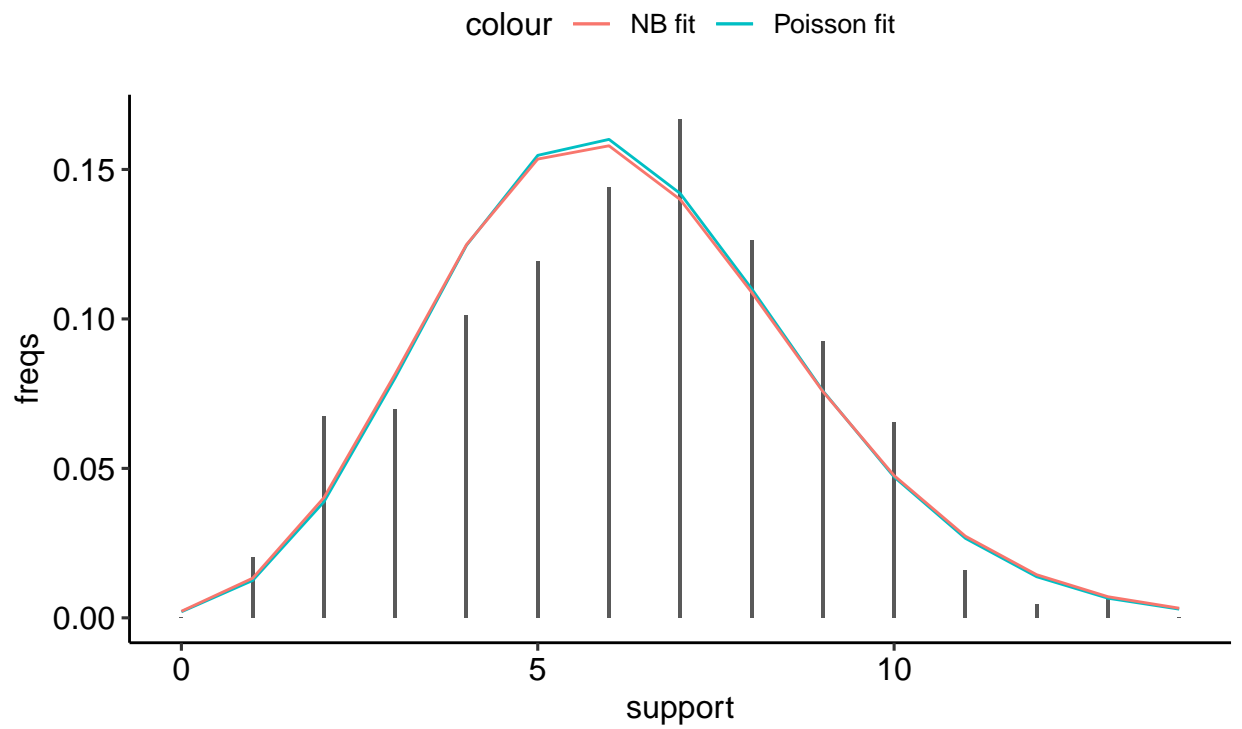
Strain: B6D2F1



Number of litters: 565

```
##  
## [[3]]
```

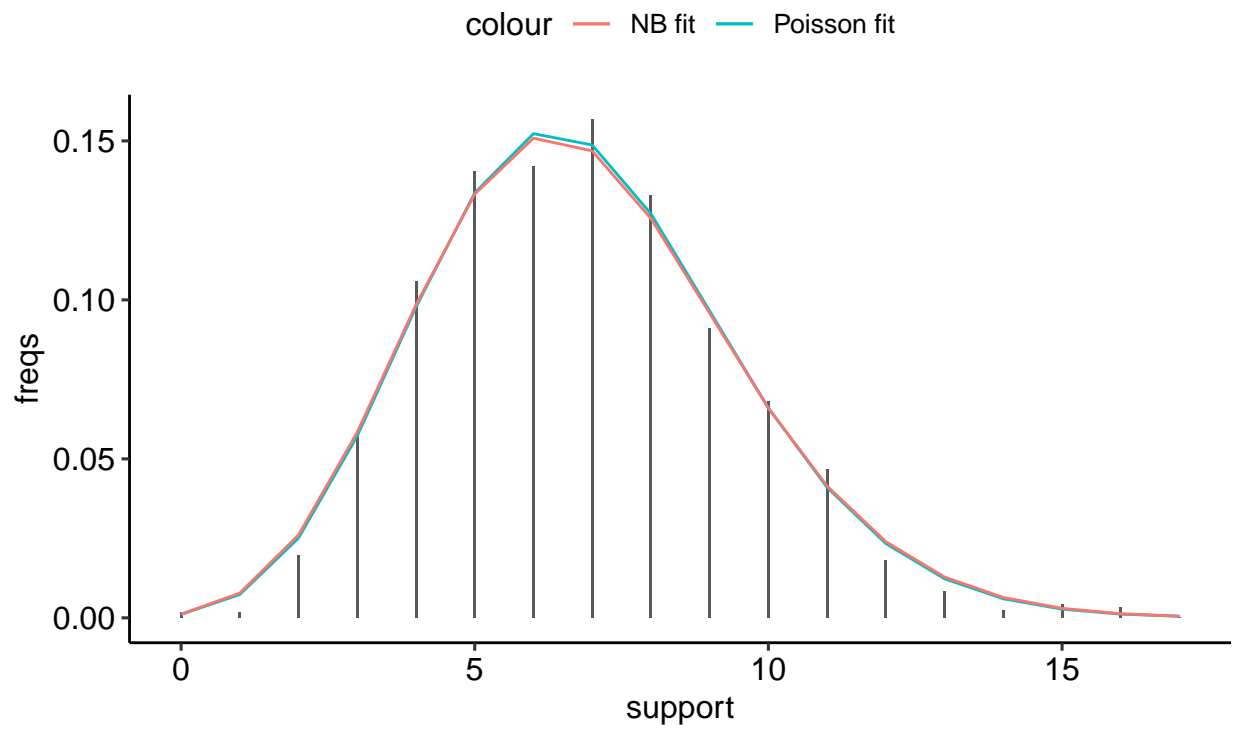
Strain: B6J_CrIF



Number of litters: 444

```
##  
## [[4]]
```

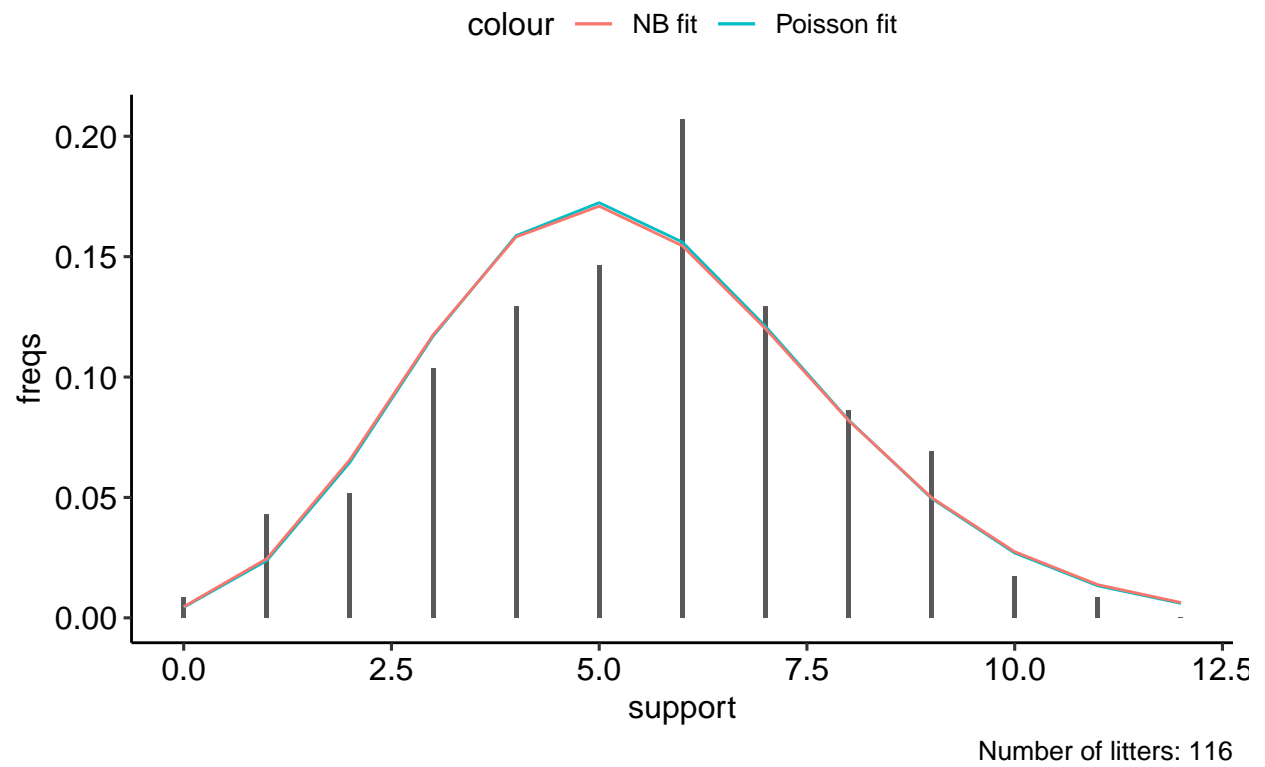
Strain: B6J_Fue



Number of litters: 1219

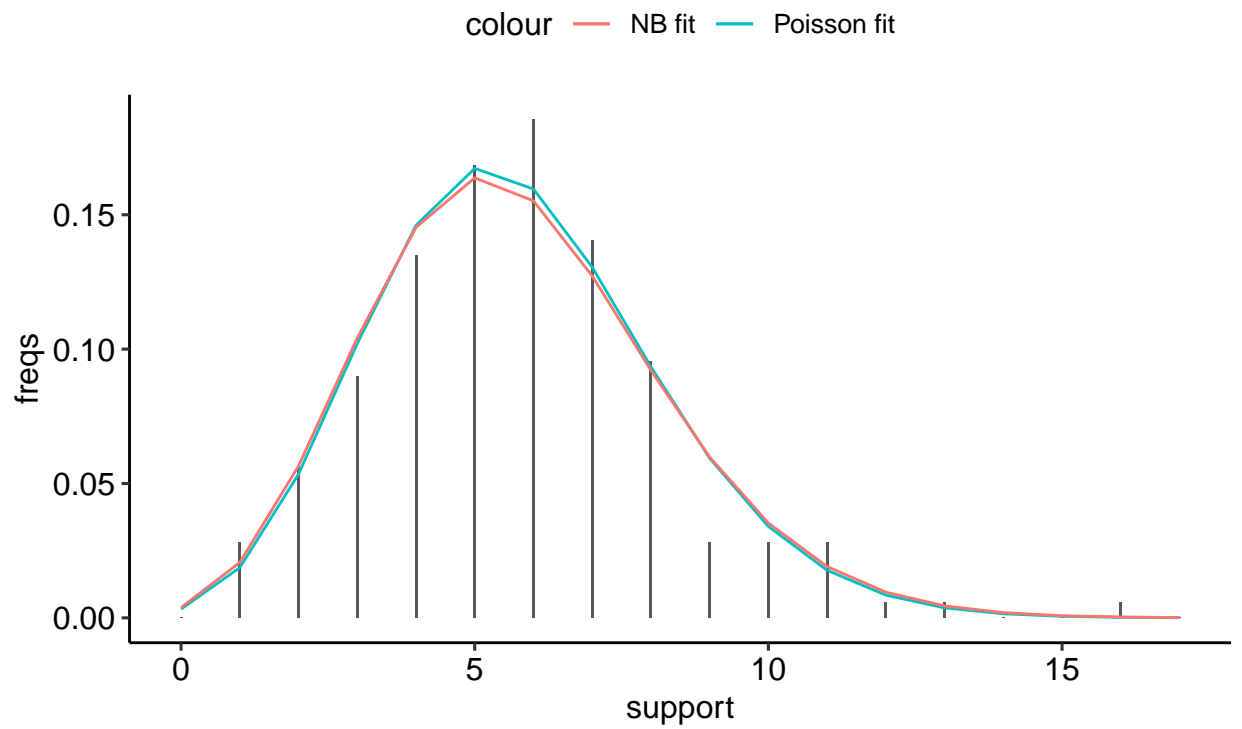
```
##  
## [[5]]
```

Strain: Balbc



```
##  
## [[6]]
```

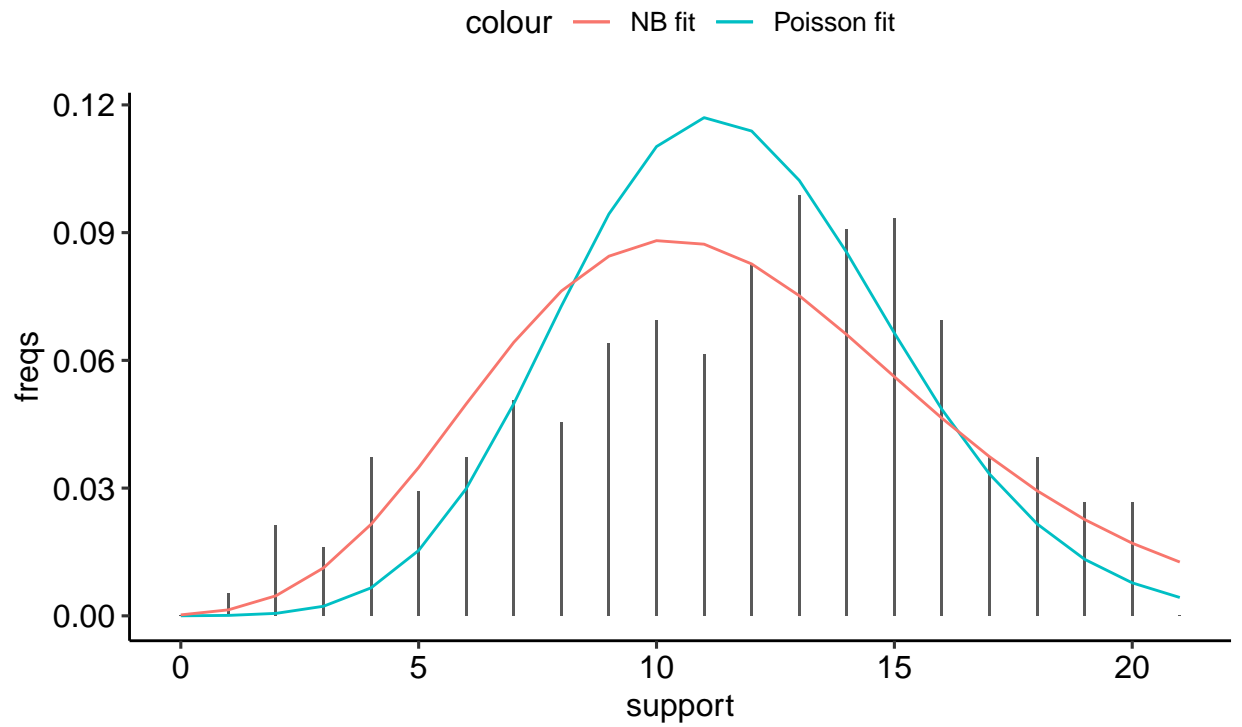

Strain: DBA2_J_Fue



Number of litters: 178

```
##  
## [[7]]
```

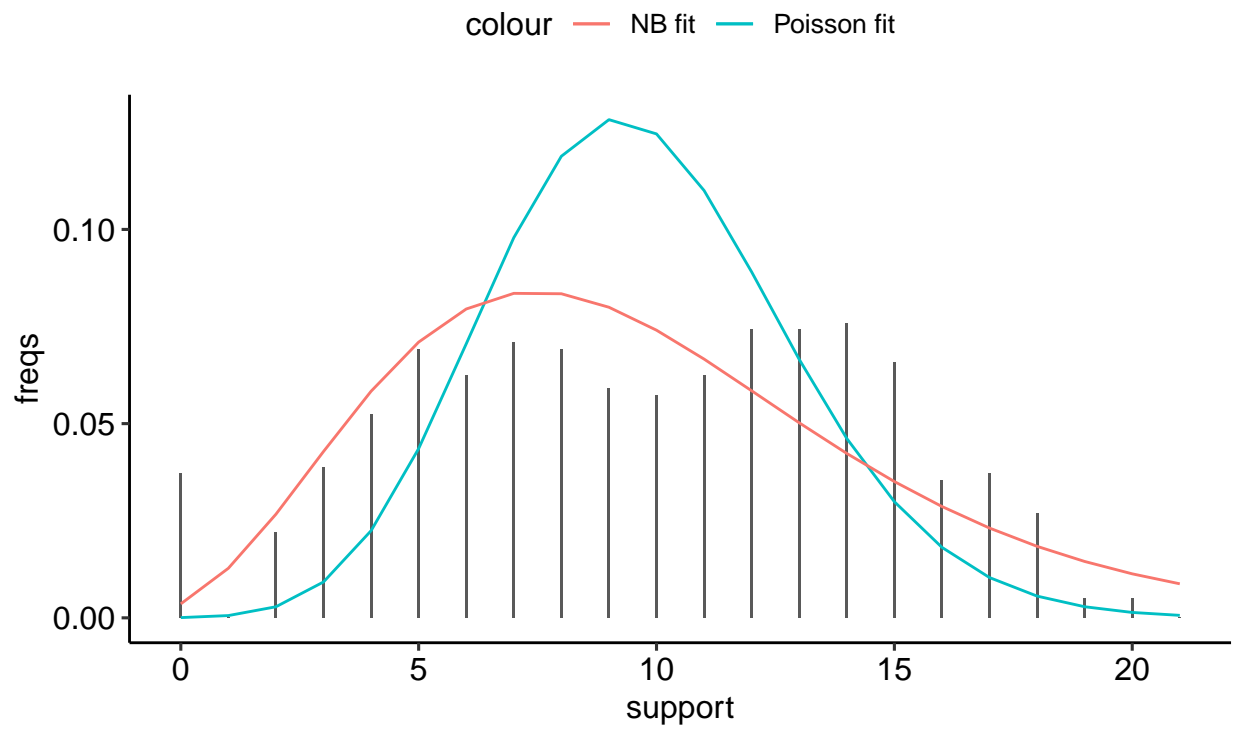
Strain: NMRI



Number of litters: 375

```
##  
## [[8]]
```

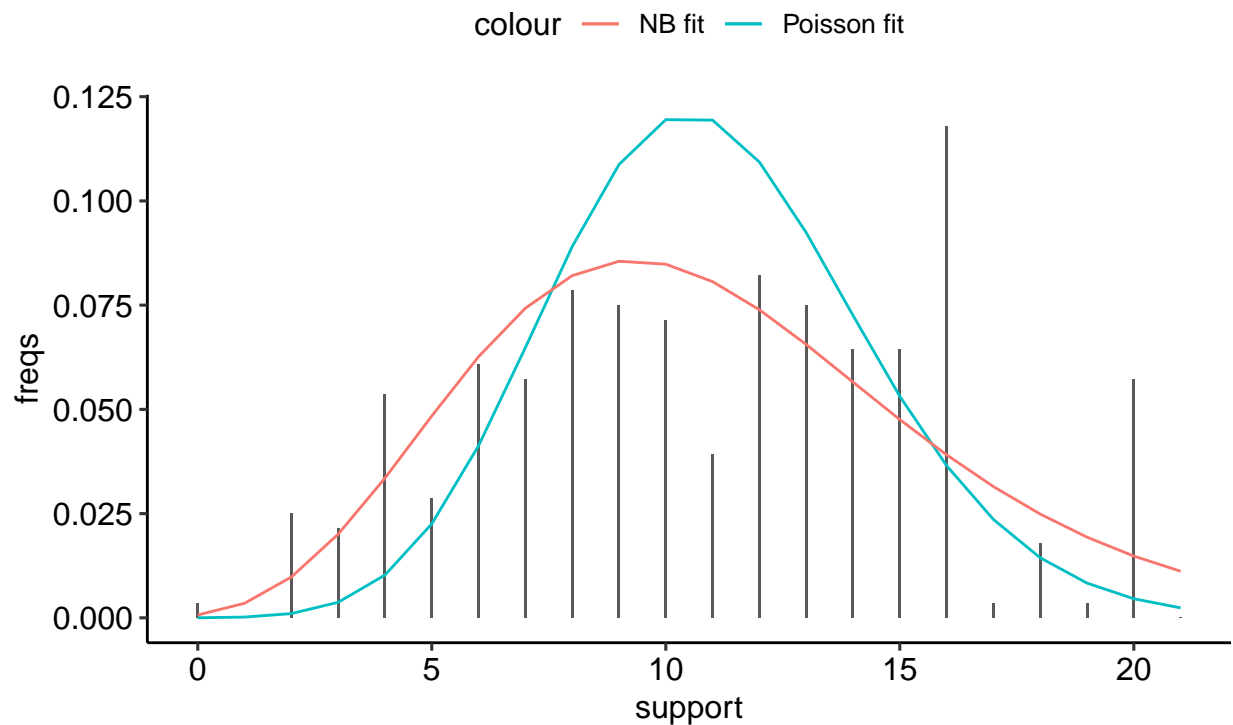
Strain: CD1_99_10



Number of litters: 593

```
##  
## [[9]]
```

Strain: CD1_10_20



Number of litters: 280

Fertility

```
load("../external/mouse_data/mouse3/fertility/nfemales.Rdata")
pre_tab_list <- lapply(
  names(table_list_clean), FUN = function(df_name){
    df <- table_list_clean[[df_name]]
    df_clean <- as.data.frame(
      apply(data.frame(df[,2:3]),
        c(1,2),
        as.numeric))
    rowmax <- apply(df_clean, 1, max)
    c(
      round(mean(rowmax),1),
      median(rowmax),
      max(rowmax),
      length(rowmax),
      nfemales[df_name],
      df_name)
  })

pre_tab_df <- data.frame(do.call(rbind, pre_tab_list))
```

```

colnames(pre_tab_df) <- c("mean", "median", "max", "n_breedings", "n_fertile", "strain")

fertility <- round(
  as.numeric(pre_tab_df$n_fertile)/as.numeric(pre_tab_df$n_breedings)*100,
  digits = 1)

strain_fertility_litter_df <- data.frame(
  strain = pre_tab_df[,c("strain")],
  fertility = fertility,
  litter_mean = as.numeric(pre_tab_df[,c("mean")]),
  litter_median = as.numeric(pre_tab_df[,c("median")]),
  litter_max = as.numeric(pre_tab_df[,c("max")]),
  n_breedings = as.numeric(pre_tab_df[,c("n_breedings")]),
  n_fertile = as.numeric(pre_tab_df[,c("n_fertile")])
)

save(strain_fertility_litter_df, file = "strain_fertility_litter_df.RData")
write.table(strain_fertility_litter_df, file = "strain_fertility_litter_df.txt",
  quote = FALSE, sep = "\t", row.names = F, col.names = T)

strain_fertility_litter_df

```

```

##      strain fertility litter_mean litter_median litter_max n_breedings
## 1   B6cBrd    93.5         6.8           7.0         12          77
## 2   B6D2F1    33.8         8.1           8.0         18         565
## 3   B6J_Cr1F   26.8         6.4           7.0         13         444
## 4   B6J_Fue    46.6         7.4           7.0         19        1219
## 5    Balbc    73.3         5.8           6.0         11         116
## 6 DBA2_J_Fue   57.9         5.9           6.0         16         178
## 7    NMRI     15.5        11.8          12.0         20         375
## 8 CD1_99_10   120.1        10.2          10.0         20         593
## 9 CD1_10_20   135.4        12.0          12.5         20         280
##      n_fertile
## 1          72
## 2         191
## 3         119
## 4         568
## 5          85
## 6         103
## 7          58
## 8         712
## 9         379

```

```

sort(strain_fertility_litter_df$fertility)

```

```

## [1] 15.5 26.8 33.8 46.6 57.9 73.3 93.5 120.1 135.4

```

```

mean(strain_fertility_litter_df$litter_mean)

```

```

## [1] 8.266667

```