````
---
title: "Mice, stroke size"
author: "Michel de Lange"
date: "11/13/2017"
output:
  pdf_document: default
  html_document: default
fontsize: 12pt
---

```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = FALSE)
options(scipen = 1, digits = 2)

```

## Mice, stroke size


```{r echo=F}
mice <-
read.csv("/Users/micheldelange/Documents/mice/data/stroke_size2.csv",header=TRUE)


 female <- grepl('female',mice$animal_type)
PLT <- grepl('PLT',mice$animal_type)
Pound <- grepl('Pound',mice$animal_type)
type <- rep("aWT",dim(mice)[1])
type[grepl('PLT',mice$animal_type)] <- "PLT"
type[grepl('Pound',mice$animal_type)] <- "Pound"
ss <- mice$stroke_size
df <- data.frame(ss,female,type)

```

We have a total of `r dim(mice)[1]` mice with a measurement for the stroke size.
Of these, `r sum(female)` are female. There are `r dim(mice)[1]-sum(PLT)-
sum(Pound)` WT mice, `r sum(PLT)` PLT mice and `r sum(Pound)` Pound mice. The
following table shows the numbers by sex and type:

```{r }
t <- table(df$female,df$type)
print(t)
fisher.test(t)
```

Fisher's exact test tells us that we can safely say that females and males are
distributed equally among the types.

The mean stroke size is `r mean(ss)`. The median is: `r median(ss)`. The
variance is `r var(ss)`. The range is from `r range(ss)[1]` to `r range(ss)[2]`.
Let's look at our sample of stroke sizes.


```{r echo=FALSE}
{
par(mfrow=c(1,1))
hist(ss,30,xlim=c(0,8),main="Stroke size, N=108",xlab='stroke size')
}
```

Now let's look at stroke size by type and sex:
````

```{r }
{
par(mfrow=c(1,2))

boxplot(ss~type,main="stroke size by type")
grid()
boxplot(ss~female,main="stroke size by sex",xaxt='n')
axis(1, at=c(1,2), labels=c("male","female"))
grid()
}
```

We see not much difference by sex, and Pound mice clearly have smaller strokes. We can confirm that this is significant, using plain vanilla analysis of variance. First by sex:

```{r echo=FALSE}

summary(aov(ss~female))
```

We see the p-value for sex is about 0.1, so we cannot reject the null hypothesis that there is no difference by sex.

Now by type:

```{r echo=FALSE}
summary(aov(ss~type))
```

The p-value for type is <0.001, so we reject the null hypothesis that there is no difference by type. We can get an estimate of how much smaller the mean stroke size will be for a Pound mouse by fitting a linear model. I have used the log of the stroke size as a response variable, because stroke size cannot be negative.


```{r}

m <- lm(log(ss)~type*female)
summary(m)

```

We see that type Pound is the only significant factor. To interpret the coefficients above on the original scale (mm3), we need to take their exponents. For example, the mean stroke size for a Pound mouse at the log scale is `r coefficients(m)[3]` less than that of a wild type mouse of the same sex. We take the exponent, and exp(`r coefficients(m)[3]`) = `r exp(coefficients(m)[3])`. So we see that the mean stroke size for a Pound mouse will be `r exp(coefficients(m)[3])` times the mean stroke for a wild type mouse of the same sex. A 95% Confidence interval for this is from `r exp(confint(m)[3,1])` to `r exp(confint(m)[3,2])`.

Conclusion

Pound mice have smaller strokes than wild type and PLT mice. The mean stroke size for a Pound mouse will be `r exp(coefficients(m)[3])` times the mean stroke for a wild type mouse of the same sex. A 95% Confidence interval for this is from `r exp(confint(m)[3,1])` to `r exp(confint(m)[3,2])`. . There is no evidence for a difference by sex, or for any interactions between sex and type.