

Evaluate How `motif_count` changes with trial at 40C

Michael Gilchrist

date: 2022-10-20

Goal

- Evaluate trends in `motif_count` under a (near) constant temperature.

Set up

Load libraries

```
## load libraries
library(stats)
require(MASS) # provides negative binomial fitting: glm.nb
library(RSQLite) # Don't think we need this.
library(rTPC) ##
library(nls.multstart)
library(broom)
library(tidyverse)
library(ggplot2)
library(ggpubr)
library(grid) ## provides textGrob
library(gridExtra)
library(viridisLite)

#options(ggplot2.continuous.colour="viridis",
#        ggplot2.discrete.colour="viridis",
#        ggplot2.scale_fill_discrete = scale_fill_viridis_d,
#        ggplot2.scale_fill_continuous = scale_fill_viridis_c)

library(GGally)
library(reshape2)
library(lme4)
library(nlme)
library(gnm)
library(rsample) ## provides bootstraps()

library(RVAideMemoire) # provides overdisp.glmer()
library(humidity) ## provides VPD
library(weathermetrics)
library(latex2exp)
```

Local Functions

```
kprint <- function(input, ...) {  
  print(knitr::kable(input, ...))  
  cat('\n\n<!-- -->\n\n')  
}
```

Plotting settings

```
## From: https://data-se.netlify.app/2018/12/12/changing-the-default-color-scheme-in-ggplot2/  
  
theme_set(theme_minimal(base_size = 9))  
theme_update(  
  plot.title = element_text(size = rel(1.1)),  
  plot.subtitle = element_text(size = rel(1))  
  
if(!exists("old_opts")) old_opts <- options() # save old options  
  
options(ggplot2.continuous.colour="viridis")  
options(ggplot2.continuous.fill = "viridis")  
options(ggplot2.discrete.colour="viridis")  
options(ggplot2.discrete.fill = "viridis")
```

Load Data

```
## Read in ZEFI Data sets  
## Treat 'repeatability' as round = 0  
## Add round info  
  
## Repeatability was done between round 1 and 2, female was present, but only one temp. so treating as  
  
output_dir <- "output"  
  
if(!dir.exists(output_dir)) dir.create(output_dir)  
git_root <- system("git rev-parse --show-toplevel", intern = TRUE)  
  
data_raw = list()  
  
data_raw[[1]] <- read.csv(file.path(git_root, "data", "collated", "HSPi-Round-1-Heat-Trials.csv")) %>%  
  ## Note T237 and T230 are missing numbers in the motif_count column  
  ## so we are filtering these observations out until they are found  
  filter(!is.na(motif_count))  
  
data_raw[[2]] <- read.csv(file.path(git_root, "data", "collated", "HSPi-Repeatability-Song-Count.csv"))  
  mutate(round = 2) %>%  
  ungroup()  
  
data_raw[[3]] <- read.csv(file.path(git_root, "data", "collated", "HSPi-Round-2-Heat-Trials.csv")) %>%  
  mutate(round = 3) %>%
```

```

## Deal with missing temp_mean and humidity_mean values
## in round == 3
## 2022/10/19 - code no longer needed
## group_by(temp_target) %>%
##mutate(temp = if_else((round == 3 & is.na(temp_mean)),
##                      mean(temp_mean, na.rm = TRUE),
##                      temp_mean)) %>%
##mutate(humidity = if_else((round == 3 & is.na(humidity_mean)),
##                          mean(humidity_mean, na.rm = TRUE),
##                          humidity_mean)) %>%
ungroup()

## Join data and discard empty columns
data_full <- full_join(data_raw[[1]], data_raw[[2]]) %>%
  full_join(data_raw[[3]]) %>%
  discard(~all(is.na(.) | . == "")) %>% ## get rid of columns of only NA
  mutate(trial_completed = !(is.na(motif_count)) ) %>%
  mutate(motif_count = ifelse(is.na(motif_count), 0, motif_count)) %>%
  mutate(motif_count = motif_count*1.0) %>% ## convert to a double so it's not treated as an integer
  mutate(chamber = as.factor(chamber), male = as.factor(male)) %>%
  ## create a male specific round and global trial index `trial`
  group_by(male, round) %>%
  mutate(trial_round = rank(date)) %>%
  ungroup(round) %>%
  mutate(trial = rank(date)) %>%
  ungroup() %>%
  mutate(motif_count_plus_1 = (motif_count + 1)) %>%
  mutate(log_motif_count_plus_1 = log(motif_count + 1)) %>%
  mutate(temp_target = as.numeric(temp_target)) %>%
  ## Create generic 'temp' column which is either
  ## temp_mean, if it exists, or temp_target, if it doesn't
  mutate(temp = if_else(is.na(temp_mean),
                        temp_target,
                        temp_mean)) %>%

  ## Add column with total motif_count for a given round
  group_by(male, round) %>%
  mutate(count_total_round = sum(motif_count)) %>%
  ungroup() %>%
  mutate(motif_prop = motif_count/count_total_round) %>%
  ## assuming poisson error
  ## From glm man page
  ## > Non-'NULL' 'weights' can be used to indicate that different
  ## > observations have different dispersions (with the values in
  ## > 'weights' being inversely proportional to the dispersions);
  ## add +1 to deal with single 0
  ## Interpret dispersion as ~sd() or se() not var()
  mutate(count_wt = sqrt(1/(motif_count + 1))) %>%
  ## need to rescale wts for motif_prop data
  mutate(prop_wt = count_wt * count_total_round) %>%
  ## Add vpd
  mutate(svp = SVP(t = temp_mean + 273.15, isK = TRUE), vpd = svp*(1-humidity_mean/100) ) %>%
  group_by(round) %>%

```

```
mutate(vpd_offset = vpd - mean(vpd)) %>%
ungroup() %>%
relocate(motif_count, motif_prop, vpd, temp_mean, humidity_mean, .after = male) %>%
mutate() ## Dummy function so we can comment out lines above it w/o any issues
```

```
## Joining with 'by = join_by(male, chamber, date, motif_count, counter, temp_target, temp_mean, temp_m
## Joining with 'by = join_by(male, chamber, date, motif_count, counter, test_order, temp_target, temp_m
```

Examine Data

Create Working Dataset

```
data_ind <- data_full %>%
#   filter(round %in% c(2,3)) %>%
#   filter(count_total_round >= 1) %>%
  mutate()

## copy data frame and assign `male = "combined"
data_comb <- data_ind %>% mutate(male = "combined")

data <- bind_rows(data_ind, data_comb)
```

Examine How Var varies with Mean temp_target = 40

WARNING: warning() are off

```
data_40 <- data_ind %>%
  filter(temp_target == 40) %>%
  unique()

dim(data_40)
```

```
## [1] 53 31
```

```
stats_40 <- data_40 %>%
  group_by(male) %>%
  summarize(count = length(motif_count),
            mean = mean(motif_count),
            var = var(motif_count, na.rm = TRUE),
            cv = sqrt(var)/mean,
            dispersion = var/mean) %>% mutate()

print("We have 15 males, 5 of which we only have 1 observation at 40C")
```

```
## [1] "We have 15 males, 5 of which we only have 1 observation at 40C"
```

```
stats_40
```

```
## # A tibble: 15 x 6
##   male count  mean  var    cv dispersion
##   <fct> <int> <dbl> <dbl> <dbl>    <dbl>
## 1 T229     5 159.  8820.  0.592    55.6
## 2 T230     4   2.25  20.2    2         9
## 3 T231     5 106.  9770.  0.929    91.8
## 4 T234     5  53.4   688.  0.491    12.9
## 5 T235     5  78.4   890.  0.381    11.4
## 6 T236     5 184.  3233.  0.310    17.6
## 7 T237     4 145.   693.  0.181     4.77
## 8 T243     5 229   4218.  0.284    18.4
## 9 T244     5  78.6 1477.  0.489    18.8
## 10 T246     1    0    NA    NA     NA
## 11 T247     5 118.  3223.  0.483    27.4
## 12 T257     1 253    NA    NA     NA
## 13 T258     1  45    NA    NA     NA
## 14 T259     1    0    NA    NA     NA
## 15 T260     1  43    NA    NA     NA
```

```
motif_stats <- stats_40
```

```
print("Flag birds birds with a coefficient of variation `cv` > 0.5 & `mean` < 50")
```

```
## [1] "Flag birds birds with a coefficient of variation 'cv' > 0.5 & 'mean' < 50"
```

```
cv_threshold <- 0.5
```

```
mean_threshold <- 50
```

```
male_high_cv <- stats_40 %>% filter(cv > cv_threshold) %>% pull(male)
```

```
data_male_high_cv <- data_40 %>% filter(male %in% male_high_cv)
```

```
male_low_mean <- stats_40 %>% filter(mean < mean_threshold) %>% pull(male)
```

```
data_male_low_mean <- data_40 %>% filter(male %in% male_low_mean)
```

```
g0 <- ggplot(data_male_high_cv) +
```

```
  aes(x=motif_count,
```

```
      color = male,
```

```
      fill = male) +
```

```
  geom_histogram() +
```

```
  labs(title = paste0("Data for males with exceptionally high cviances", paste(male_high_cv, collapse = ", "))
```

```
for(filter_high_cv in c(FALSE, TRUE)){
```

```
  for(filter_low_mean in c(FALSE, TRUE)){
```

```
    data_tmp <- data_40
```

```
    motif_stats_tmp <- motif_stats
```

```

if(filter_high_cv){
  data_tmp <- filter(data_tmp, !(male %in% male_high_cv) )
  motif_stats_tmp <- filter(motif_stats_tmp, !(male %in% male_high_cv) )
}

if(filter_low_mean){
  data_tmp <- filter(data_tmp, !(male %in% male_low_mean) )
  motif_stats_tmp <- filter(motif_stats_tmp, !(male %in% male_low_mean) )
}

g0 <- ggplot(data_tmp) +
  aes(x=male,
      y = motif_count,
      color = male,
      fill = male) +
  geom_point() +
  geom_rug(size=0.75) +
  ## theme(axis.ticks.x=element_line(size=30))+
  geom_point(
    mapping = aes(x = male,
                  y = mean,
                  color = male),
    size = 2.5,
    shape = 24,
    data = motif_stats_tmp) +
  theme(legend.position = "none", axis.text.x = element_text(angle = -45)) +
  labs(title = "Raw Count Data")

g1 <- ggplot(data_tmp) +
  aes(x=motif_count,
      color = male,
      fill = male) +
  geom_histogram() +
  xlim(0, NA) +
  labs(title = "Raw Count Data")

formula = y ~ -1 + x
g2 <- ggplot(motif_stats_tmp) +
  aes(x=mean, y = var) +
  geom_point(aes(color = male)) +
  geom_smooth(method='lm', formula = formula, fullrange = TRUE) +
  ##stat_cor(label.y.npc = 0.95) +
  stat_regline_equation(label.y.npc = 0.9, aes(label = paste(..eq.label..., ..adj.rr.label...,
  theme(legend.position = "none") +
  xlim(0, NA) +
  labs(title = "Mean vs. var")

fit_g2 <- lm(var ~ -1 + mean, data = motif_stats_tmp)

g3 <- ggplot(motif_stats_tmp) +
  aes(x=mean, y = var) +
  geom_point(aes(color = male)) +
  geom_smooth(method='lm', formula= formula, fullrange = TRUE) +

```

```

stat_regline_equation(label.y.npc = 0.9, aes(label = paste(..eq.label.., ..adj.rr.label..,
theme(legend.position = "none") +
labs(title = "Mean vs. Var",
      subtitle = "log(var) ~ log(mean)") +
scale_x_log10() +
scale_y_log10()

formula <- y ~ -1 + x + I(x^2)
g4 <- ggplot(motif_stats_tmp) +
  aes(x=mean, y = var) +
  geom_point(aes(color = male)) +
  geom_smooth(method='lm', formula = formula, fullrange = TRUE) +
  stat_regline_equation(label.y.npc = 0.9, aes(label = paste(..eq.label.., ..adj.rr.label..,
theme(legend.position = "none") +
xlim(0, NA) +
labs(title = "Mean vs. Var",
      subtitle = "var ~ mean + mean^2")
      # scale_x_log10() +
      # scale_y_log10()
fit_g4 <- lm(var ~ -1 + mean + I(mean^2), data = motif_stats_tmp)

formula <- y ~ -1 + I(x^2)
g5 <- ggplot(motif_stats_tmp) +
  aes(x=mean, y = var) +
  geom_point(aes(color = male)) +
  geom_smooth(method='lm', formula = formula, fullrange = TRUE) +
  stat_regline_equation(label.y.npc = 0.9,
    aes(label = paste(
      ..eq.label..,
      ..adj.rr.label.., sep = "~~")),
    formula = formula, size = 2.5) +
  theme(legend.position = "none") +
  labs(title = "Mean vs. Var",
    subtitle = "var ~ mean^2")

fit_g5 <- lm(var ~ -1 + I(mean^2), data = motif_stats_tmp)
#ifelse(length(dev.list()) < 4, dev.new(), dev.next())

ga <- grid.arrange(g0, g2, g4, g5,
  ncol=2,
  top=textGrob(
    paste0("Mean vs. Var: Filter High CV = ", filter_high_cv,
    ", Low Mean = ", filter_low_mean, "\n temp_target = 40C"),
    gp=gpar(fontsize = 11))
  )
##cat("\n\n\\pagebreak\n")
##print("<P style='page-break-before: always'>") #forced new-page happens here.

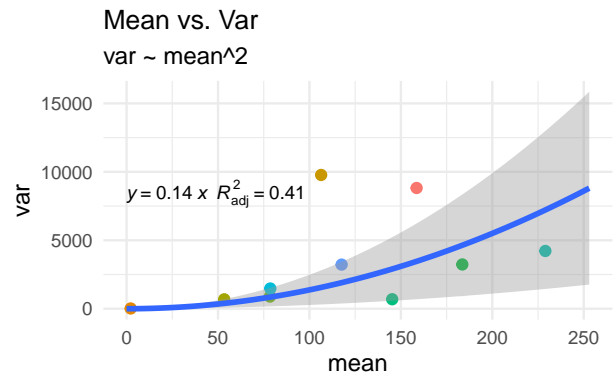
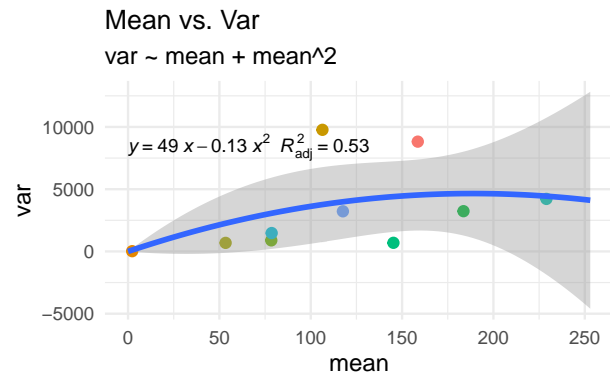
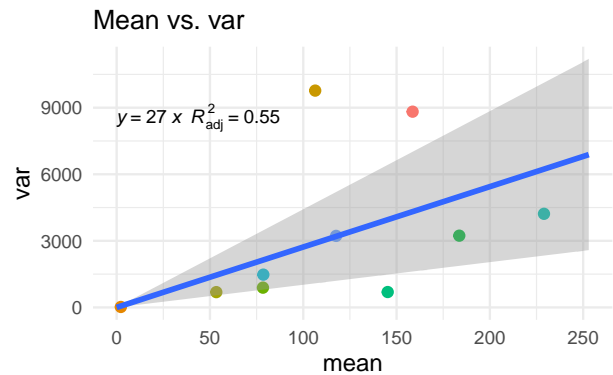
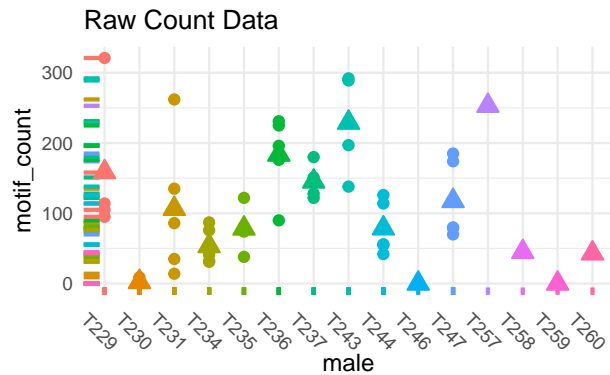
print(paste0("Mean vs Var; filter high cv: ", filter_high_cv, "; filter low mean: ", filter_low
print(summary(fit_g2))
print(summary(fit_g4))
print(summary(fit_g5))

```

```
ga
```

```
dev.print(device = pdf, file = file.path(output_dir, paste0("mean.vs.var_filter.high-", filter_l
}
}
```

Mean vs. Var: Filter High CV = FALSE, Low Mean = FALSE
temp_target = 40C



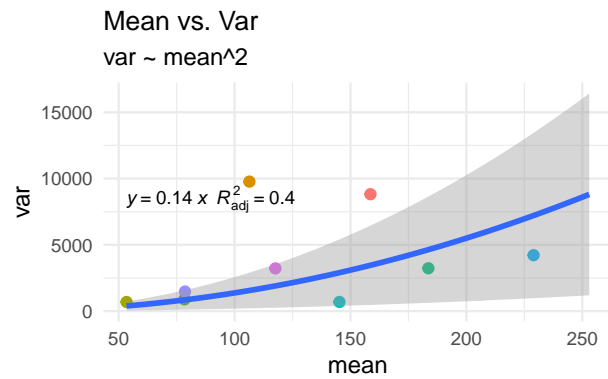
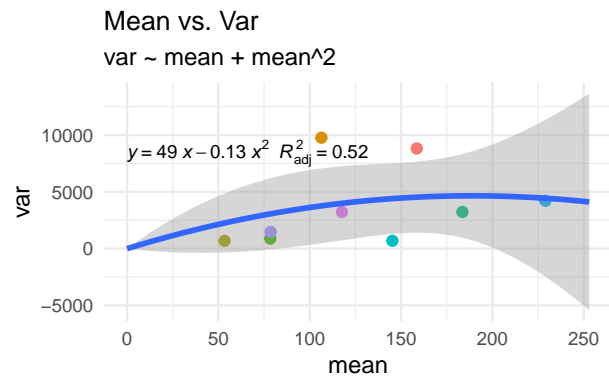
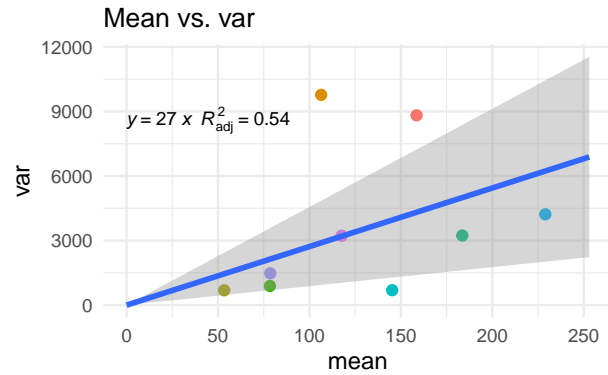
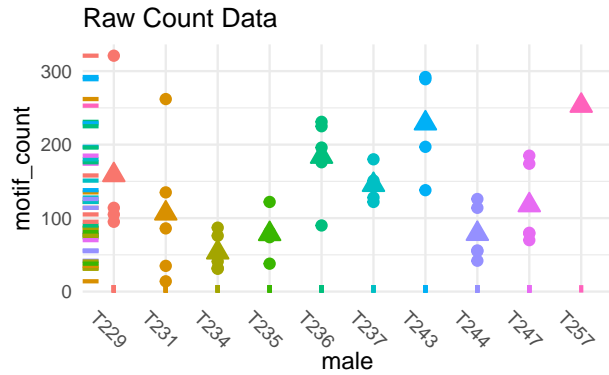
```
## [1] "Mean vs Var; filter high cv: FALSE; filter low mean: FALSE"
##
## Call:
## lm(formula = var ~ -1 + mean, data = motif_stats_tmp)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3259.7 -1633.0  -713.5    7.1  6874.9
##
## Coefficients:
##      Estimate Std. Error t value Pr(>|t|)
## mean      27.21      7.52   3.619  0.00558 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3125 on 9 degrees of freedom
## (5 observations deleted due to missingness)
## Multiple R-squared:  0.5927, Adjusted R-squared:  0.5474
## F-statistic: 13.1 on 1 and 9 DF, p-value: 0.005583
```



```
##
##
## Call:
## lm(formula = var ~ -1 + mean + I(mean^2), data = motif_stats_tmp)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3711.3 -1585.6 -1090.7  -122.8  6004.4
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## mean          49.2854    26.9567   1.828   0.105
## I(mean^2)    -0.1306     0.1529  -0.854   0.418
##
## Residual standard error: 3174 on 8 degrees of freedom
## (5 observations deleted due to missingness)
## Multiple R-squared:  0.6267, Adjusted R-squared:  0.5334
## F-statistic: 6.715 on 2 and 8 DF,  p-value: 0.01942
##
##
## Call:
## lm(formula = var ~ -1 + I(mean^2), data = motif_stats_tmp)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2995.9 -1048.2   170.4  1147.3  8212.9
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## I(mean^2)   0.13757    0.04863   2.829  0.0198 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3563 on 9 degrees of freedom
## (5 observations deleted due to missingness)
## Multiple R-squared:  0.4707, Adjusted R-squared:  0.4119
## F-statistic: 8.004 on 1 and 9 DF,  p-value: 0.01975
```

Mean vs. Var: Filter High CV = FALSE, Low Mean = TRUE

temp_target = 40C



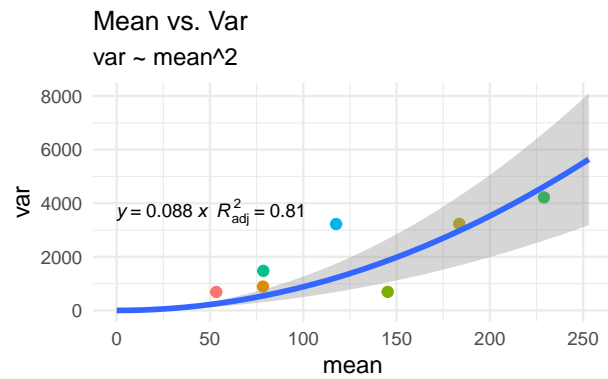
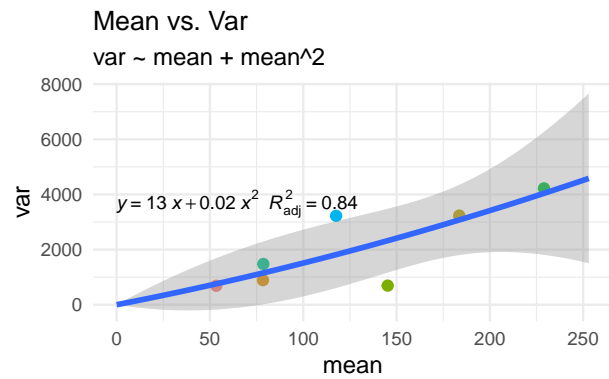
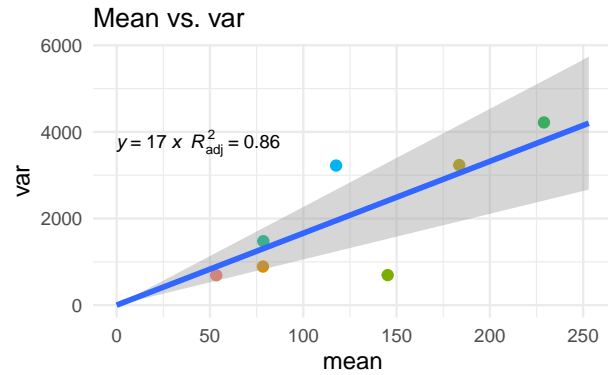
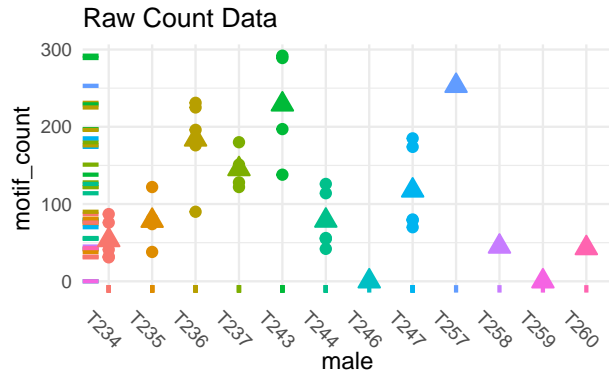
```
## [1] "Mean vs Var; filter high cv: FALSE; filter low mean: TRUE"
##
## Call:
## lm(formula = var ~ -1 + mean, data = motif_stats_tmp)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3259.8 -1763.0  -764.9   23.0  6874.8
##
## Coefficients:
##      Estimate Std. Error t value Pr(>|t|)
## mean    27.213     7.976   3.412  0.0092 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3315 on 8 degrees of freedom
## (1 observation deleted due to missingness)
## Multiple R-squared:  0.5927, Adjusted R-squared:  0.5418
## F-statistic: 11.64 on 1 and 8 DF, p-value: 0.009198
##
## Call:
## lm(formula = var ~ -1 + mean + I(mean^2), data = motif_stats_tmp)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3711.7 -1591.1 -1414.4  -220.3  6003.8
```

```

##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## mean          49.2998    28.8216   1.711   0.131
## I(mean^2)    -0.1306     0.1635  -0.799   0.451
##
## Residual standard error: 3392 on 7 degrees of freedom
## (1 observation deleted due to missingness)
## Multiple R-squared:  0.6267, Adjusted R-squared:  0.5201
## F-statistic: 5.876 on 2 and 7 DF,  p-value: 0.03178
##
##
## Call:
## lm(formula = var ~ -1 + I(mean^2), data = motif_stats_tmp)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2996  -1404    296   1321   8213
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## I(mean^2)  0.13757    0.05158   2.667   0.0285 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3779 on 8 degrees of freedom
## (1 observation deleted due to missingness)
## Multiple R-squared:  0.4707, Adjusted R-squared:  0.4045
## F-statistic: 7.114 on 1 and 8 DF,  p-value: 0.02848

```

Mean vs. Var: Filter High CV = TRUE, Low Mean = FALSE
temp_target = 40C



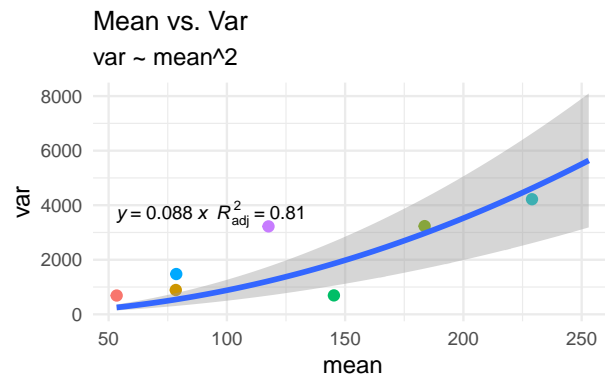
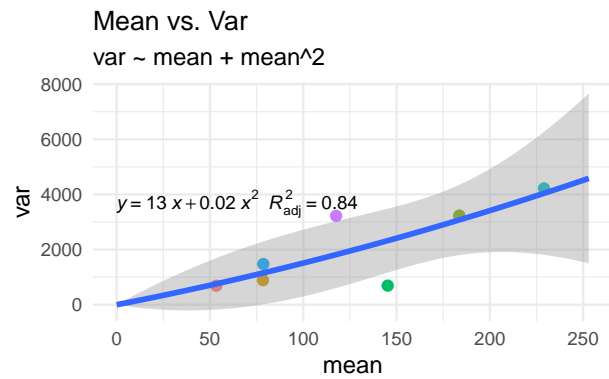
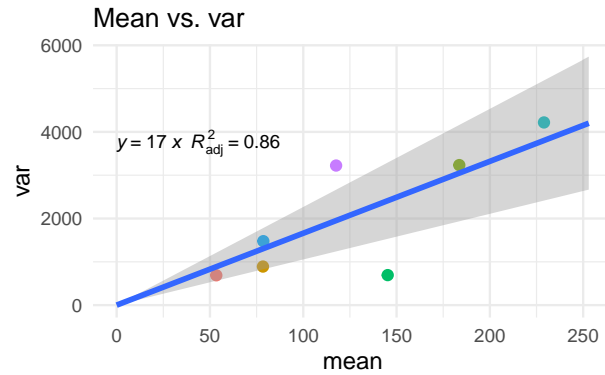
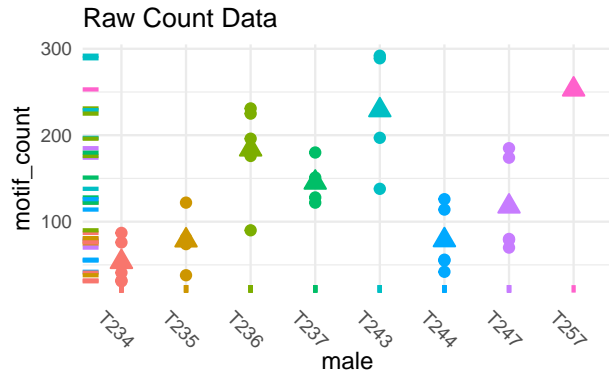
```
## [1] "Mean vs Var; filter high cv: TRUE; filter low mean: FALSE"
##
## Call:
## lm(formula = var ~ -1 + mean, data = motif_stats_tmp)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1718.5  -304.7   171.9   301.0  1271.0
##
## Coefficients:
##      Estimate Std. Error t value Pr(>|t|)
## mean    16.601      2.477   6.703 0.000535 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 914.2 on 6 degrees of freedom
## (5 observations deleted due to missingness)
## Multiple R-squared:  0.8822, Adjusted R-squared:  0.8626
## F-statistic: 44.93 on 1 and 6 DF, p-value: 0.0005351
##
## Call:
## lm(formula = var ~ -1 + mean + I(mean^2), data = motif_stats_tmp)
##
## Residuals:
##      1       2       3       4       5       6       8
## -69.43 -260.67  158.60 -1630.15  179.33  322.59 1406.31
```

```

##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## mean       13.14106    9.35311   1.405   0.219
## I(mean^2)  0.01964    0.05087   0.386   0.715
##
## Residual standard error: 986.8 on 5 degrees of freedom
## (5 observations deleted due to missingness)
## Multiple R-squared:  0.8856, Adjusted R-squared:  0.8399
## F-statistic: 19.36 on 2 and 5 DF, p-value: 0.004426
##
##
## Call:
## lm(formula = var ~ -1 + I(mean^2), data = motif_stats_tmp)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1166.28   -70.03   348.64   684.69  2004.57
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## I(mean^2)  0.08812    0.01568   5.622  0.00135 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1064 on 6 degrees of freedom
## (5 observations deleted due to missingness)
## Multiple R-squared:  0.8404, Adjusted R-squared:  0.8139
## F-statistic: 31.61 on 1 and 6 DF, p-value: 0.001353

```

Mean vs. Var: Filter High CV = TRUE, Low Mean = TRUE
temp_target = 40C



```
## [1] "Mean vs Var; filter high cv: TRUE; filter low mean: TRUE"
##
## Call:
## lm(formula = var ~ -1 + mean, data = motif_stats_tmp)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1718.5  -304.7   171.9   301.0  1271.0
##
## Coefficients:
##      Estimate Std. Error t value Pr(>|t|)
## mean    16.601      2.477   6.703 0.000535 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 914.2 on 6 degrees of freedom
## (1 observation deleted due to missingness)
## Multiple R-squared:  0.8822, Adjusted R-squared:  0.8626
## F-statistic: 44.93 on 1 and 6 DF, p-value: 0.0005351
##
## Call:
## lm(formula = var ~ -1 + mean + I(mean^2), data = motif_stats_tmp)
##
## Residuals:
##      1       2       3       4       5       6       7
## -69.43 -260.67  158.60 -1630.15  179.33  322.59 1406.31
```

```
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## mean       13.14106    9.35311   1.405   0.219
## I(mean^2)   0.01964    0.05087   0.386   0.715
##
## Residual standard error: 986.8 on 5 degrees of freedom
## (1 observation deleted due to missingness)
## Multiple R-squared:  0.8856, Adjusted R-squared:  0.8399
## F-statistic: 19.36 on 2 and 5 DF,  p-value: 0.004426
##
##
## Call:
## lm(formula = var ~ -1 + I(mean^2), data = motif_stats_tmp)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1166.28   -70.03   348.64   684.69  2004.57
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## I(mean^2)  0.08812    0.01568   5.622  0.00135 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1064 on 6 degrees of freedom
## (1 observation deleted due to missingness)
## Multiple R-squared:  0.8404, Adjusted R-squared:  0.8139
## F-statistic: 31.61 on 1 and 6 DF,  p-value: 0.001353
```

Conclusion

- Excluding males with very high variances indicates that for the remaining males `var ~ mean`, but overdispersed relative to the poisson.
- Using multiple trials at one temp could be an effective way to identify 'reliable' males.

Compare motif_count vs trial at 40C

WARNING: warning() are off

```
for(filter_high_cv in c(FALSE, TRUE)){
  for(filter_low_mean in c(FALSE, TRUE)){

    data_tmp <- data_40 %>% group_by(male) %>%
      arrange(trial, .by_group = TRUE) %>%
      mutate(cummean = cummean(motif_count)) %>%
      select(male, trial, motif_count, cummean)

    if(filter_high_cv){
      data_tmp <- filter(data_tmp, !(male %in% male_high_cv) )
    }
  }
}
```

```

if(filter_low_mean){
  data_tmp <- filter(data_tmp, !(male %in% male_low_mean) )
}

g1 <- ggplot(data_tmp) +
  aes(x = trial, y = motif_count) +
  geom_point(aes(color = male), position = "jitter") +
  geom_line(aes(x = trial, y = cummean, color = male)) +
  theme(legend.position="bottom")

legend <- get_legend(g1)

## Apply a linear regression

formula <- y ~ 1 + x
g0 <- ggplot(data_tmp, aes(x = trial, y = motif_count)) +
  geom_point(aes(color = male))

g2 <- g0 +
  geom_smooth(method = 'glm', formula = formula, fullrange = TRUE) +
  stat_regline_equation(label.y.npc = 0.9, aes(label = paste(..eq.label..., ..adj.rr.label...,
  theme(legend.position = "none") +
  xlim(1, NA) +
  labs(title = "Linear")
##g2

fit_g2 <- glm( motif_count ~ 1 + trial, data = data_tmp)

g3 <- g0 + geom_smooth(method = 'glm',
  formula= formula,
  ## Can't use link = identity because it leads to negative expectations
  method.args = list(family = quasipoisson(link = "log"))) +
  stat_regline_equation(label.y.npc = 0.9, aes(label = paste(..eq.label..., ..adj.rr.label...,
  theme(legend.position = "none") +
  xlim(1, NA) +
  labs(title = "log-linear: family = qpoisson(link = log)")
##g3

fit_g3 <- glm( motif_count ~ 1 + trial, data = data_tmp,
  family = quasipoisson(link = "log"))
formula <- y ~ 1 + x + I(x^2)
g4 <- g0 + geom_smooth(method='glm',
  formula= formula,
  ## Can't use link = identity because it leads to negative expectations
  method.args = list(family = quasipoisson(link = "log"))) +
  stat_regline_equation(label.y.npc = 0.9, aes(label = paste(..eq.label..., ..adj.rr.label...,
  theme(legend.position = "none") +
  xlim(1, NA) +
  labs(title = "log-quad: family = qpoisson(link = log)")
##g4

```



```

fit_g4 <- glm(motif_count ~ 1 + trial + I(trial^2), data = data_tmp,
             family = quasipoisson(link = "log"))

ga <- grid.arrange(g2, g3, g4, as_ggplot(legend),
                 ncol=2,
                 top=textGrob(
                   paste0("motif_count vs. trial: Filter High CV = ", filter_high_cv, ", Low Mean = ", filter_low_mean),
                   gp=gpar(fontsize = 11))
                 )

##cat("\n\n\\pagebreak\n")
##print("<P style='page-break-before: always'>") #forced new-page happens here.

print(paste0("motif_count vs. trial; filter.high: ", filter_high_cv, "; filter.low.mean: ", filter_low_mean))

print(summary(fit_g2))
print(summary(fit_g3))
print(summary(fit_g4))

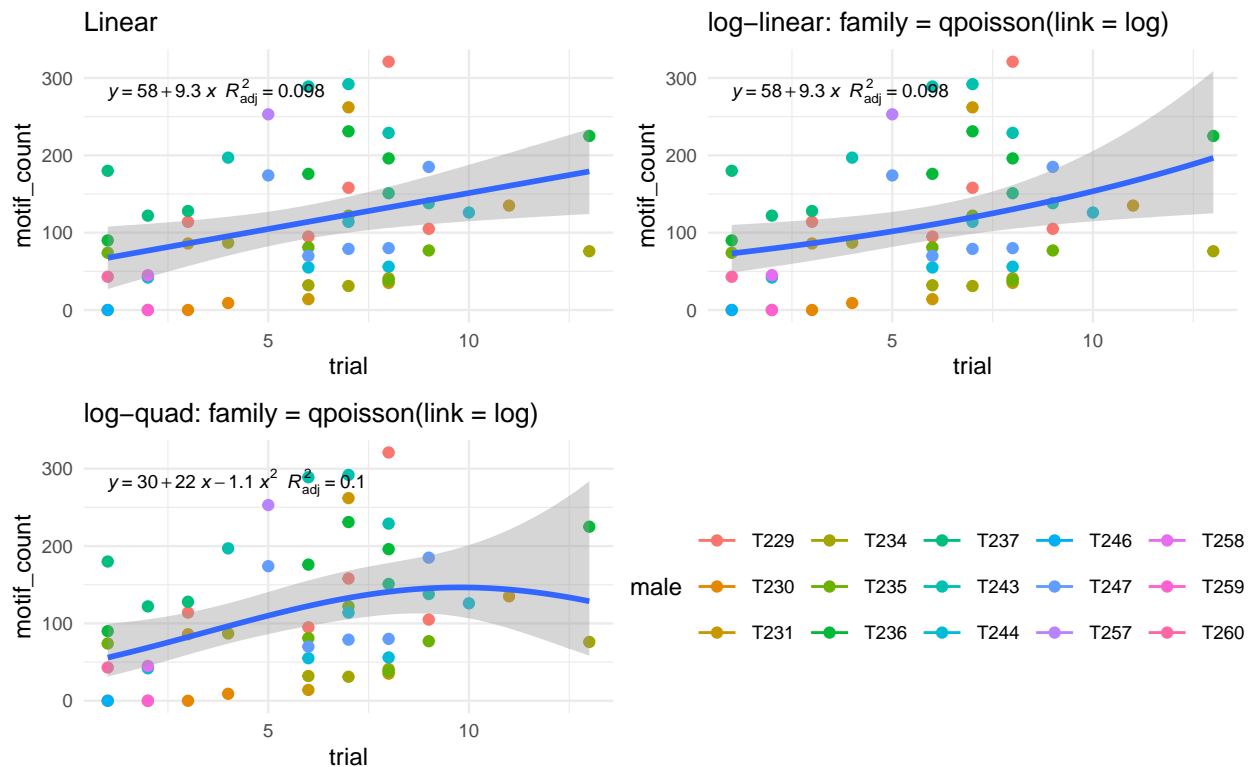
ga

dev.print(device = pdf, file = file.path(output_dir, paste0("motif_count.vs.trial_filter.high-", filter_high_cv, ".pdf")))
}

```

motif_count vs. trial: Filter High CV = FALSE, Low Mean = FALSE

temp_target = 40C



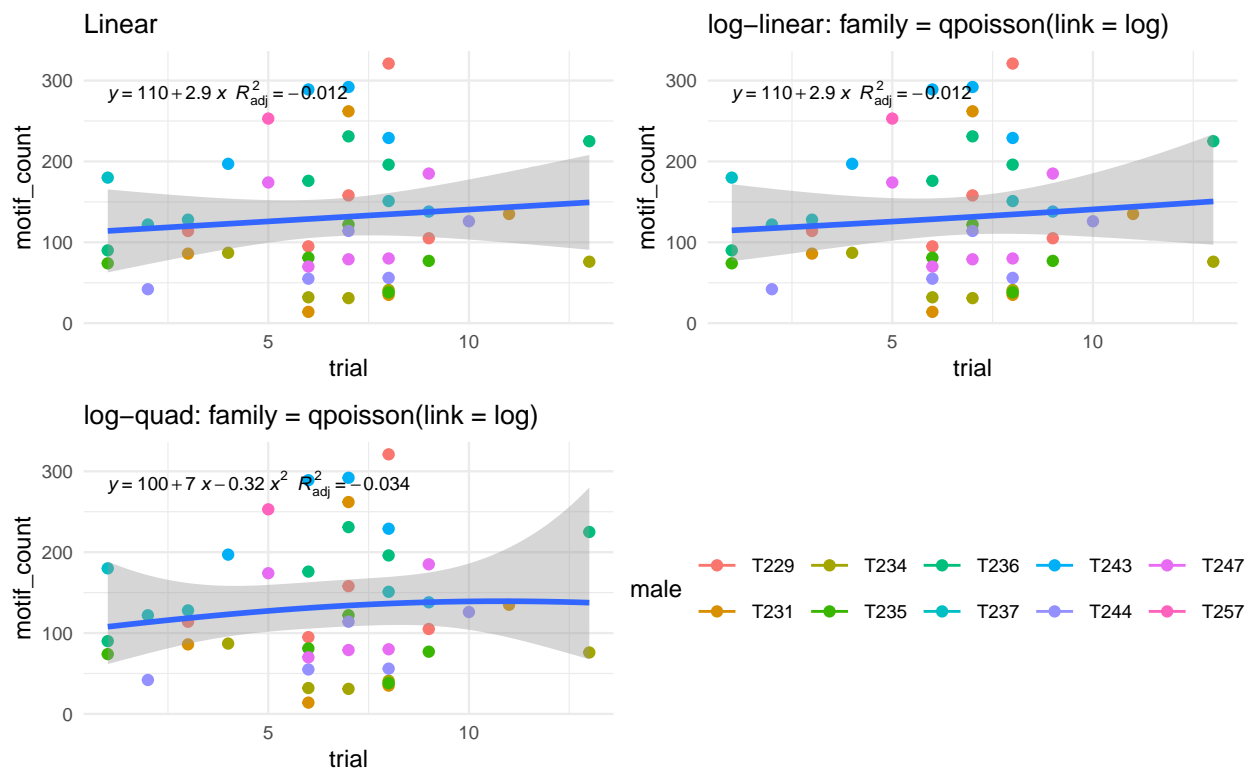
```
## [1] "motif_count vs. trial; filter.high: FALSE; filter.low.mean: FALSE"
```

```
##
## Call:
## glm(formula = motif_count ~ 1 + trial, data = data_tmp)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -103.13   -67.50   -19.01    45.20   188.38
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   58.197     23.711   2.454  0.0176 *
## trial         9.303       3.600   2.584  0.0127 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 6447.535)
##
##      Null deviance: 371877  on 52  degrees of freedom
## Residual deviance: 328824  on 51  degrees of freedom
## AIC: 619.26
##
## Number of Fisher Scoring iterations: 2
##
## Call:
## glm(formula = motif_count ~ 1 + trial, family = quasipoisson(link = "log"),
##      data = data_tmp)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
##  -13.134   -7.343   -1.505    4.191   14.104
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.21013    0.23716  17.752 <2e-16 ***
## trial        0.08238    0.03255   2.531  0.0145 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasipoisson family taken to be 59.54193)
##
##      Null deviance: 3611.6  on 52  degrees of freedom
## Residual deviance: 3230.1  on 51  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 5
##
## Call:
## glm(formula = motif_count ~ 1 + trial + I(trial^2), family = quasipoisson(link = "log"),
##      data = data_tmp)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
```

```
## -12.838   -6.853   -1.692    4.530   13.160
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.789536   0.392501   9.655 5.17e-13 ***
## trial        0.244958   0.120186   2.038  0.0468 *
## I(trial^2)   -0.012524   0.008951  -1.399  0.1679
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasipoisson family taken to be 57.86452)
##
## Null deviance: 3611.6 on 52 degrees of freedom
## Residual deviance: 3106.8 on 50 degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 5
```

motif_count vs. trial: Filter High CV = FALSE, Low Mean = TRUE

temp_target = 40C



```
## [1] "motif_count vs. trial; filter.high: FALSE; filter.low.mean: TRUE"
##
## Call:
## glm(formula = motif_count ~ 1 + trial, data = data_tmp)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -114.76  -58.76  -14.53   48.18  186.35
```

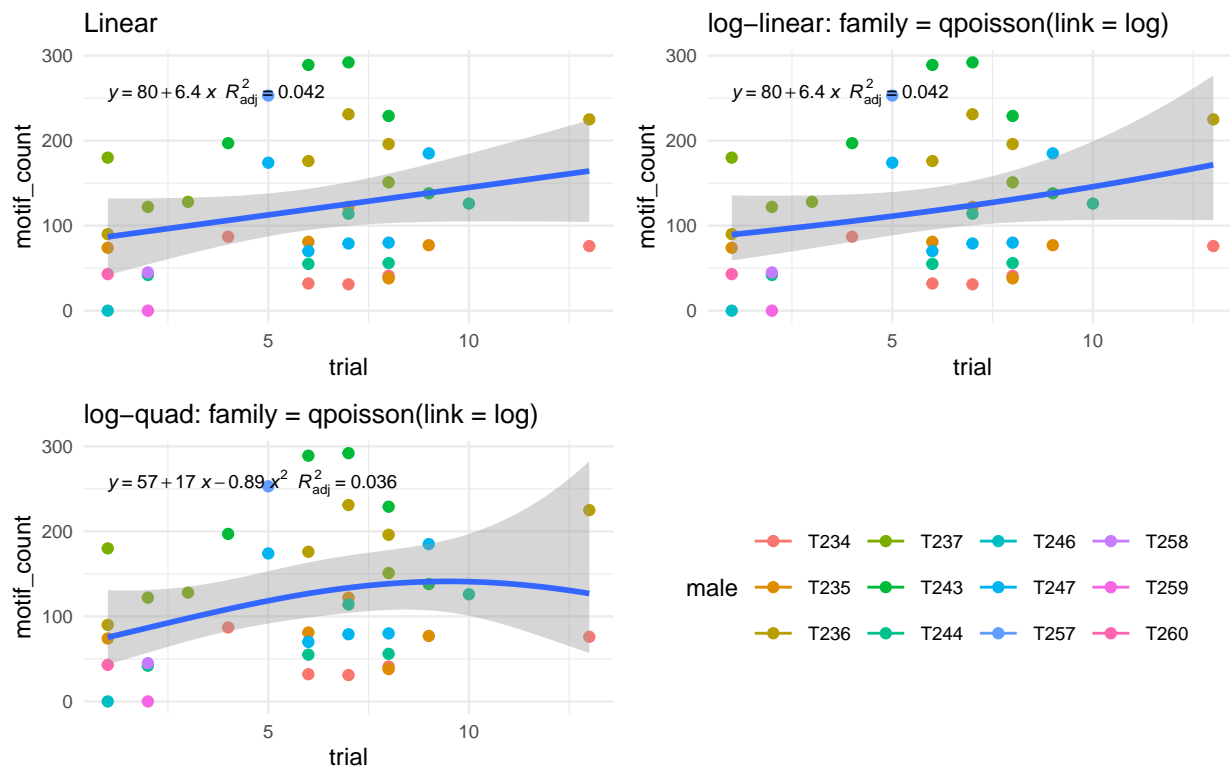
```

##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  111.120     30.026   3.701 0.000607 ***
## trial         2.941       4.238   0.694 0.491489
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 6308.269)
##
## Null deviance: 274293  on 44  degrees of freedom
## Residual deviance: 271256  on 43  degrees of freedom
## AIC: 525.39
##
## Number of Fisher Scoring iterations: 2
##
## Call:
## glm(formula = motif_count ~ 1 + trial, family = quasipoisson(link = "log"),
##      data = data_tmp)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -12.921   -5.634   -1.261    4.075   13.622
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.72022     0.23567  20.029  <2e-16 ***
## trial         0.02263     0.03253   0.696    0.49
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasipoisson family taken to be 48.16706)
##
## Null deviance: 2096.1  on 44  degrees of freedom
## Residual deviance: 2072.8  on 43  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 5
##
## Call:
## glm(formula = motif_count ~ 1 + trial + I(trial^2), family = quasipoisson(link = "log"),
##      data = data_tmp)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -13.094   -5.682   -1.131    3.911   13.419
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.625391   0.378350  12.225 2.01e-15 ***
## trial         0.057720   0.112630   0.512   0.611
## I(trial^2)  -0.002672   0.008209  -0.326   0.746

```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasipoisson family taken to be 48.74887)
##
## Null deviance: 2096.1  on 44  degrees of freedom
## Residual deviance: 2067.5  on 42  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 5
```

motif_count vs. trial: Filter High CV = TRUE, Low Mean = FALSE
temp_target = 40C



```
## [1] "motif_count vs. trial; filter.high: TRUE; filter.low.mean: FALSE"
##
## Call:
## glm(formula = motif_count ~ 1 + trial, data = data_tmp)
##
## Deviance Residuals:
##   Min       1Q   Median       3Q      Max
## -94.53  -56.70  -12.85   58.85  169.91
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   80.403     26.449   3.040  0.00433 **
## trial          6.447       3.947   1.633  0.11089
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

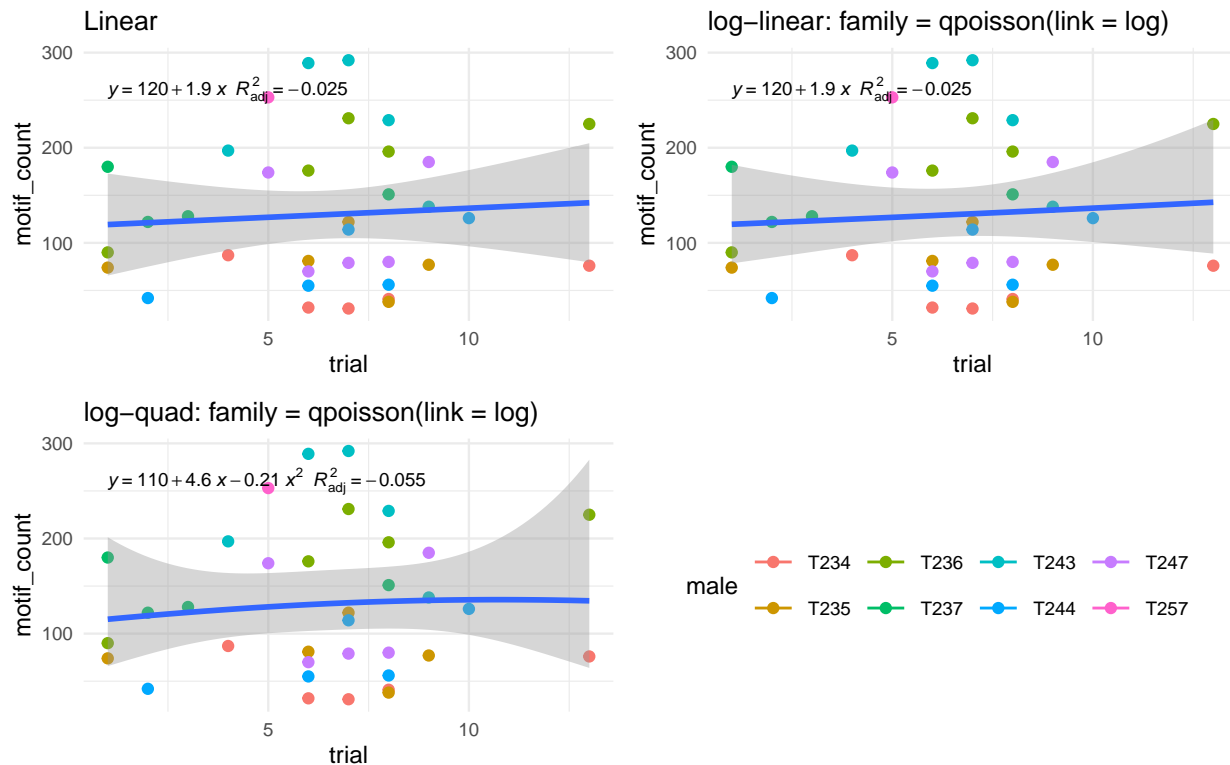
```

##
## (Dispersion parameter for gaussian family taken to be 5963.929)
##
##      Null deviance: 236575  on 38  degrees of freedom
## Residual deviance: 220665  on 37  degrees of freedom
## AIC: 453.67
##
## Number of Fisher Scoring iterations: 2
##
## Call:
## glm(formula = motif_count ~ 1 + trial, family = quasipoisson(link = "log"),
##      data = data_tmp)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -13.746   -5.875   -1.681    4.462   13.325
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.43988    0.24117  18.410  <2e-16 ***
## trial        0.05425    0.03360   1.615   0.115
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasipoisson family taken to be 51.37537)
##
##      Null deviance: 2119.6  on 38  degrees of freedom
## Residual deviance: 1985.7  on 37  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 5
##
## Call:
## glm(formula = motif_count ~ 1 + trial + I(trial^2), family = quasipoisson(link = "log"),
##      data = data_tmp)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -13.164   -5.460   -1.274    4.359   12.309
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.166886    0.373094  11.168 2.99e-13 ***
## trial        0.164773    0.115602   1.425   0.163
## I(trial^2)  -0.008669    0.008713  -0.995   0.326
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasipoisson family taken to be 50.82185)
##
##      Null deviance: 2119.6  on 38  degrees of freedom
## Residual deviance: 1932.2  on 36  degrees of freedom

```

```
## AIC: NA
##
## Number of Fisher Scoring iterations: 5
```

motif_count vs. trial: Filter High CV = TRUE, Low Mean = TRUE
temp_target = 40C



```
## [1] "motif_count vs. trial; filter.high: TRUE; filter.low.mean: TRUE"
##
## Call:
## glm(formula = motif_count ~ 1 + trial, data = data_tmp)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -99.72  -58.17  -10.45   55.60  161.28
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  117.351     31.343   3.744 0.000691 ***
## trial         1.910       4.444   0.430 0.670174
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 5817.569)
##
##      Null deviance: 193054  on 34  degrees of freedom
## Residual deviance: 191980  on 33  degrees of freedom
## AIC: 406.67
##
```

```

## Number of Fisher Scoring iterations: 2
##
##
## Call:
## glm(formula = motif_count ~ 1 + trial, family = quasipoisson(link = "log"),
##      data = data_tmp)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -10.4902   -5.5325   -0.9118    4.6282   12.1277
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.76896    0.24555   19.42  <2e-16 ***
## trial        0.01475    0.03427    0.43    0.67
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasipoisson family taken to be 44.77057)
##
##      Null deviance: 1486.5  on 34  degrees of freedom
## Residual deviance: 1478.2  on 33  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 5
##
##
## Call:
## glm(formula = motif_count ~ 1 + trial + I(trial^2), family = quasipoisson(link = "log"),
##      data = data_tmp)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -10.6221   -5.4656   -0.9185    4.5441   11.9429
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  4.710533    0.377257   12.486 7.64e-14 ***
## trial        0.037176    0.113337    0.328   0.745
## I(trial^2)  -0.001729    0.008319   -0.208   0.837
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for quasipoisson family taken to be 45.88031)
##
##      Null deviance: 1486.5  on 34  degrees of freedom
## Residual deviance: 1476.2  on 32  degrees of freedom
## AIC: NA
##
## Number of Fisher Scoring iterations: 5

```


Conclusion

When filtering out low mean (which have few trials and, thus, bias the analysis towards a small intercept) results in no evidence that `motif_count` increases with `trial`

End

```
knitr::knit_exit()
```