R09 - Poisson Regression

HCI/PSYCH 522
Iowa State University

April 14, 2022

Overview

- Linear regression (with logarithm of breaks)
  - Tension
  - + Wool Type
- Poisson regression
  - Tension
  - + Wool Type

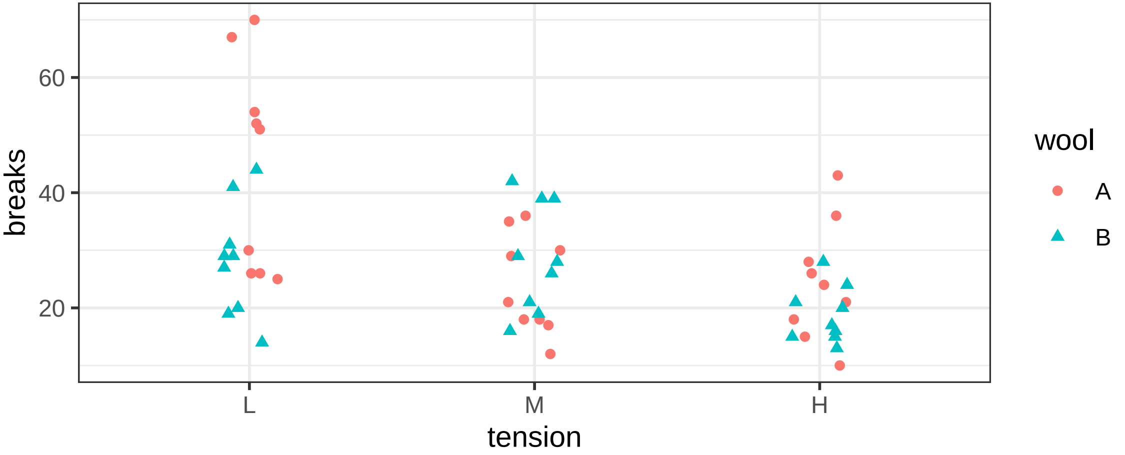
warpbreaks

```
warpbreaks

##    breaks wool tension
## 1      26    A       L
## 2      30    A       L
## 3      54    A       L
## 4      25    A       L
## 5      70    A       L
## 6      52    A       L
## 7      51    A       L
## 8      26    A       L
## 9      67    A       L
## 10     18    A       M
## 11     21    A       M
## 12     29    A       M
## 13     17    A       M
## 14     12    A       M
## 15     18    A       M
## 16     35    A       M
## 17     30    A       M
```

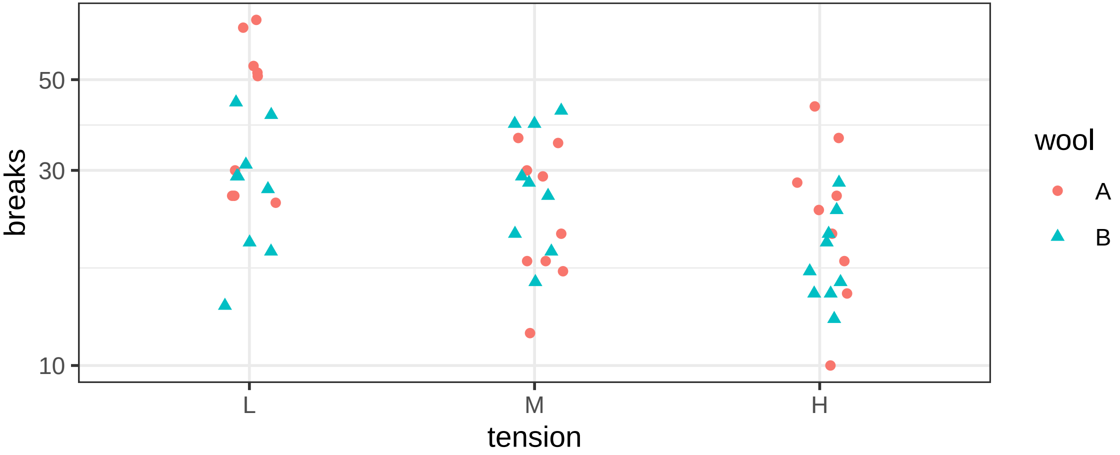
warpbreaks

```
summary(warpbreaks)

##      breaks      wool   tension
##  Min.   :10.00   A:27   L:18
##  1st Qu.:18.25   B:27   M:18
##  Median :26.00          H:18
##  Mean   :28.15
##  3rd Qu.:34.00
##  Max.   :70.00
```

warpbreaks



warpbreaks



Linear regression with log of dependent variable

Regression with

$$\log(Y_i) \stackrel{ind}{\sim} N(\mu_i, \sigma^2), \quad \mu_i = \beta_0 + \beta_1 X_{i,1} + \cdots + \beta_p X_{i,p}$$

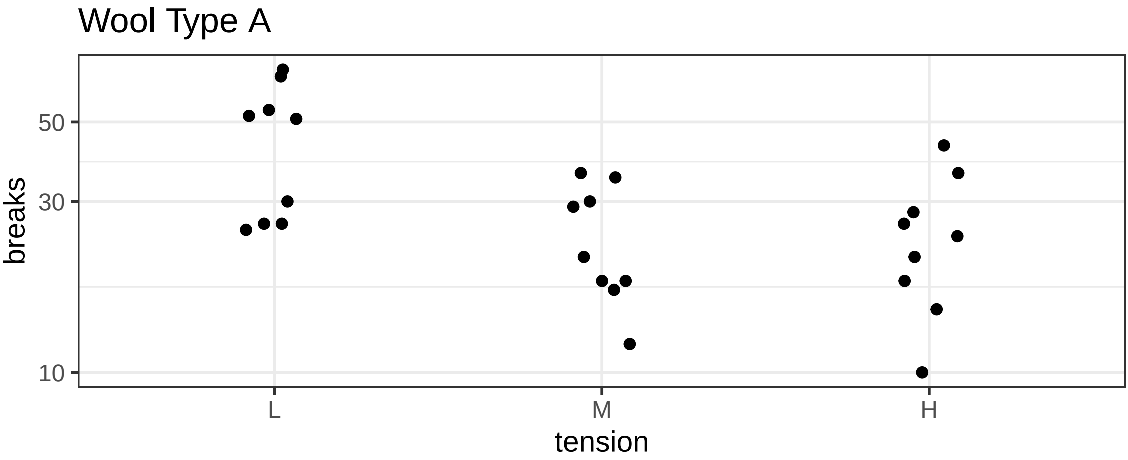
where, for observation i,

- Y_i is the dependent variable and
- $X_{i,p}$ is the p^{th} independent variable.

Interpretation

- e^{β_0} is median of the dependent variable when all independent variables (X's) are 0
- $100(e^{\beta_p} - 1)$ for $p \neq 0$, is the percent increase in the median of the dependent variable for each unit increase in the associated independent variable
- The coefficient of determination, R^2, is the proportion of variability in the logarithm of the dependent variable explained by the model.

warpbreaks



warpbreaks

```
##
## Call:
## lm(formula = log(breaks) ~ tension, data = warpbreaks %>% filter(wool ==
##     "A"))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.81504 -0.30014  0.06043  0.27774  0.64358
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.7179     0.1394  26.678  < 2e-16 ***
## tensionM     -0.6012     0.1971  -3.050  0.00550 **
## tensionH     -0.6003     0.1971  -3.046  0.00556 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4181 on 24 degrees of freedom
## Multiple R-squared:  0.3404,Adjusted R-squared:  0.2855
## F-statistic: 6.194 on 2 and 24 DF,  p-value: 0.006777
```

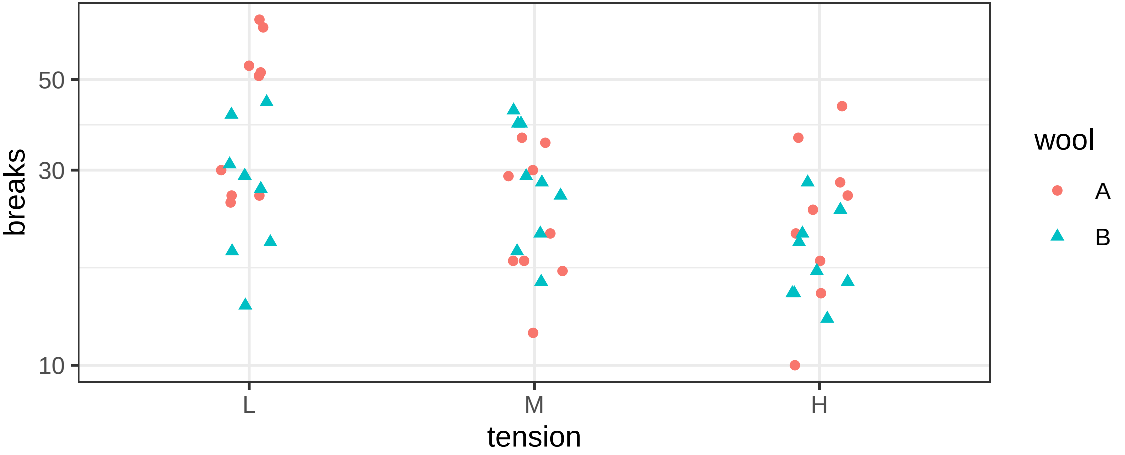
warpbreaks

```
em <- emmeans(m, pairwise ~ tension)
ci <- confint(em, type = "response", adjust = "none")
ci

## $emmeans
##  tension response   SE df lower.CL upper.CL
##  L           41.2 5.74 24     30.9     54.9
##  M           22.6 3.15 24     16.9     30.1
##  H           22.6 3.15 24     16.9     30.1
##
## Confidence level used: 0.95
## Intervals are back-transformed from the log scale
##
## $contrasts
##  contrast ratio    SE df lower.CL upper.CL
##  L / M    1.824 0.360 24    1.215     2.74
##  L / H    1.823 0.359 24    1.214     2.74
##  M / H    0.999 0.197 24    0.665     1.50
##
## Confidence level used: 0.95
## Intervals are back-transformed from the log scale
```

warpbreaks

- For wool type A, the median number of breaks when
  - tension is low is 41 (31,55),
  - tension is medium is 23 (17,30),
  - tension is high is 23 (17,30),
- For wool type A, tension
  - low compared to medium causes an 82% (21,174) increase in median number of breaks,
  - low compared to high causes an 82% (21,174) increase in median number of breaks, and
  - medium compared to high causes an 0% (-33,50) increase in median number of breaks.

warpbreaks



warpbreaks

```
##
## Call:
## lm(formula = log(breaks) ~ tension + wool, data = warpbreaks)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.80421 -0.29975 -0.01627  0.28367  0.67424
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.5762     0.1063  33.644  < 2e-16 ***
## tensionM     -0.2871     0.1302  -2.205 0.032048 *
## tensionH     -0.4893     0.1302  -3.758 0.000448 ***
## woolB        -0.1522     0.1063  -1.431 0.158540
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.3906 on 50 degrees of freedom
## Multiple R-squared:  0.246,Adjusted R-squared:  0.2008
## F-statistic: 5.438 on 3 and 50 DF,  p-value: 0.002576
```

warpbreaks

```
em_tension <- emmeans(m, pairwise ~ tension)
ci_tension <- confint(em_tension, type = "response", adjust = "none")
ci_tension

## $emmeans
##  tension response   SE df lower.CL upper.CL
##  L            33.1 3.05 50     27.5     39.8
##  M            24.9 2.29 50     20.7     29.9
##  H            20.3 1.87 50     16.9     24.4
##
## Results are averaged over the levels of: wool
## Confidence level used: 0.95
## Intervals are back-transformed from the log scale
##
## $contrasts
##  contrast ratio    SE df lower.CL upper.CL
##  L / M     1.33 0.173 50    1.026     1.73
##  L / H     1.63 0.212 50    1.256     2.12
##  M / H     1.22 0.159 50    0.942     1.59
##
## Results are averaged over the levels of: wool
## Confidence level used: 0.95
```

warpbreaks

```
em_wool <- emmeans(m, pairwise ~ wool)
ci_wool <- confint(em_wool, type = "response", adjust = "none")
ci_wool

## $emmeans
##  wool response   SE df lower.CL upper.CL
##  A        27.6 2.07 50     23.7     32.1
##  B        23.7 1.78 50     20.4     27.6
##
## Results are averaged over the levels of: tension
## Confidence level used: 0.95
## Intervals are back-transformed from the log scale
##
## $contrasts
##  contrast ratio    SE df lower.CL upper.CL
##  A / B     1.16 0.124 50     0.94     1.44
##
## Results are averaged over the levels of: tension
## Confidence level used: 0.95
## Intervals are back-transformed from the log scale
```

warpbreaks

```
em_tension_by_wool <- emmeans(m, pairwise ~ tension | wool)
ci_tension_by_wool <- confint(em_tension_by_wool, type = "response", adjust = "none")
ci_tension_by_wool$emmeans

## wool = A:
##  tension response   SE df lower.CL upper.CL
##  L           35.7 3.80 50     28.9     44.2
##  M           26.8 2.85 50     21.7     33.2
##  H           21.9 2.33 50     17.7     27.1
##
## wool = B:
##  tension response   SE df lower.CL upper.CL
##  L           30.7 3.26 50     24.8     38.0
##  M           23.0 2.45 50     18.6     28.5
##  H           18.8 2.00 50     15.2     23.3
##
## Confidence level used: 0.95
## Intervals are back-transformed from the log scale
```

warpbreaks

```
ci_tension_by_wool$contrasts

## wool = A:
##  contrast ratio    SE df lower.CL upper.CL
##  L / M     1.33 0.173 50    1.026     1.73
##  L / H     1.63 0.212 50    1.256     2.12
##  M / H     1.22 0.159 50    0.942     1.59
##
## wool = B:
##  contrast ratio    SE df lower.CL upper.CL
##  L / M     1.33 0.173 50    1.026     1.73
##  L / H     1.63 0.212 50    1.256     2.12
##  M / H     1.22 0.159 50    0.942     1.59
##
## Confidence level used: 0.95
## Intervals are back-transformed from the log scale
```

warpbreaks

```
em_wool_by_tension <- emmeans(m, pairwise ~ wool | tension)
ci_wool_by_tension <- confint(em_wool_by_tension, type = "response", adjust = "none")
ci_wool_by_tension$emmeans

## tension = L:
##  wool response   SE df lower.CL upper.CL
##  A        35.7 3.80 50     28.9     44.2
##  B        30.7 3.26 50     24.8     38.0
##
## tension = M:
##  wool response   SE df lower.CL upper.CL
##  A        26.8 2.85 50     21.7     33.2
##  B        23.0 2.45 50     18.6     28.5
##
## tension = H:
##  wool response   SE df lower.CL upper.CL
##  A        21.9 2.33 50     17.7     27.1
##  B        18.8 2.00 50     15.2     23.3
##
## Confidence level used: 0.95
## Intervals are back-transformed from the log scale
```

warpbreaks

```
ci_wool_by_tension$contrasts

## tension = L:
##  contrast ratio    SE df lower.CL upper.CL
##  A / B     1.16 0.124 50     0.94     1.44
##
## tension = M:
##  contrast ratio    SE df lower.CL upper.CL
##  A / B     1.16 0.124 50     0.94     1.44
##
## tension = H:
##  contrast ratio    SE df lower.CL upper.CL
##  A / B     1.16 0.124 50     0.94     1.44
##
## Confidence level used: 0.95
## Intervals are back-transformed from the log scale
```

warpbreaks

The estimated median number of breaks is

```
pm <- ci_tension_by_wool$emmeans %>%
  as.data.frame() %>%
  mutate(mean_with_ci = paste0(
    round(response), " (", round(lower.CL), ", ", round(upper.CL), ")")
  ) %>%
  tidyr::pivot_wider(id_cols = tension, names_from = wool, values_from = mean_with_ci)
pm

## # A tibble: 3 x 3
##   tension A            B
##   <fct>   <chr>        <chr>
## 1 L       36 (29, 44) 31 (25, 38)
## 2 M       27 (22, 33) 23 (19, 29)
## 3 H       22 (18, 27) 19 (15, 23)
```

warpbreaks

While holding wool type constant, the percent change in median number of breaks is

```
ci_tension$contrasts %>% as.data.frame %>%
  mutate(
    change = 100*(ratio-1),
    lower  = 100*(lower.CL-1),
    upper  = 100*(upper.CL-1),

    change_with_ci = paste0(
      round(change), " (", round(lower), ", ", round(upper), ")"
    )) %>%
  select(contrast, change_with_ci)

##   contrast change_with_ci
## 1    L / M     33 (3, 73)
## 2    L / H   63 (26, 112)
## 3    M / H    22 (-6, 59)
```

Poisson regression

Poisson regression with

$$Y_i \overset{ind}{\sim} Po(\lambda_i), \quad \log(\lambda_i) = \beta_0 + \beta_1 X_{i,1} + \cdots + \beta_p X_{i,p}$$

where, for observation i,

- Y_i is the dependent variable and
- $X_{i,p}$ is the p^{th} independent variable.

Interpretation

- e^{β_0} is mean of the dependent variable when all independent variables (X's) are 0
- $100(e^{\beta_p} - 1)$ for $p \neq 0$, is the percent increase in the mean of the dependent variable for each unit increase in the associated independent variable.

warpbreaks

```
##
## Call:
## glm(formula = breaks ~ tension, family = poisson, data = warpbreaks %>%
##     filter(wool == "A"))
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.3383  -1.7940  -0.1125   1.2736   3.5153
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  3.79674    0.04994  76.030  < 2e-16 ***
## tensionM    -0.61868    0.08440  -7.330 2.30e-13 ***
## tensionH    -0.59580    0.08378  -7.112 1.15e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 194.97  on 26  degrees of freedom
## Residual deviance: 119.62  on 24  degrees of freedom
## AIC: 264.99
```

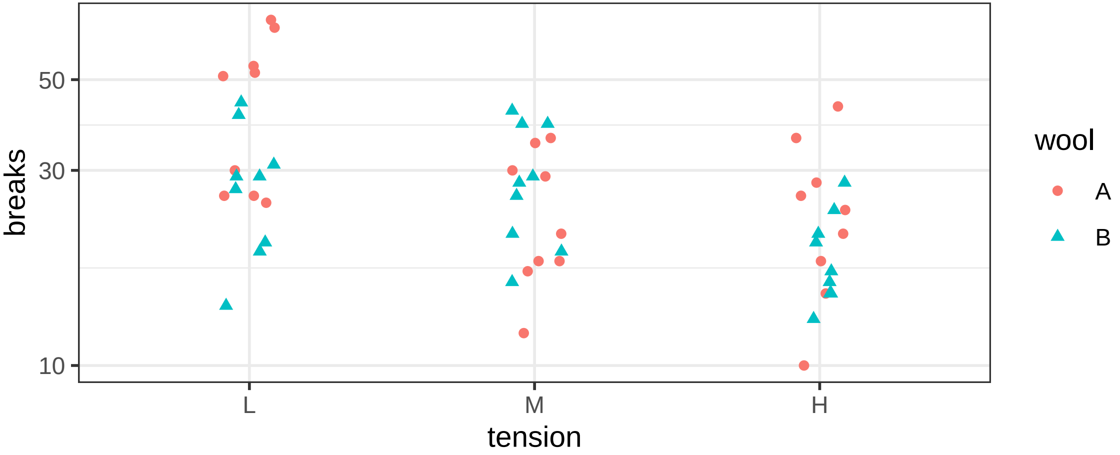
warpbreaks

```
em <- emmeans(m, pairwise ~ tension)
ci <- confint(em, type = "response", adjust = "none")
ci

## $emmeans
##  tension rate   SE  df asymp.LCL asymp.UCL
##  L       44.6 2.22 Inf      40.4      49.1
##  M       24.0 1.63 Inf      21.0      27.4
##  H       24.6 1.65 Inf      21.5      28.0
##
## Confidence level used: 0.95
## Intervals are back-transformed from the log scale
##
## $contrasts
##  contrast ratio     SE  df asymp.LCL asymp.UCL
##  L / M    1.856 0.1567 Inf      1.57      2.19
##  L / H    1.814 0.1520 Inf      1.54      2.14
##  M / H    0.977 0.0935 Inf      0.81      1.18
##
## Confidence level used: 0.95
## Intervals are back-transformed from the log scale
```

warpbreaks

- For wool type A, the mean number of breaks when
  - tension is low is 45 (40,49),
  - tension is medium is 24 (21,27),
  - tension is high is 25 (22,28),
- For wool type A, tension
  - low compared to medium causes an 86% (57,119) increase in mean number of breaks,
  - low compared to high causes an 81% (54,114) increase in mean number of breaks, and
  - medium compared to high causes an -2% (-19,18) increase in mean number of breaks.

warpbreaks



warpbreaks

```
##
## Call:
## glm(formula = breaks ~ tension + wool, family = poisson, data = warpbreaks)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -3.6871  -1.6503  -0.4269   1.1902   4.2616
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  3.69196    0.04541  81.302  < 2e-16 ***
## tensionM    -0.32132    0.06027  -5.332 9.73e-08 ***
## tensionH    -0.51849    0.06396  -8.107 5.21e-16 ***
## woolB       -0.20599    0.05157  -3.994 6.49e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for poisson family taken to be 1)
##
##     Null deviance: 297.37  on 53  degrees of freedom
## Residual deviance: 210.39  on 50  degrees of freedom
## AIC: 493.06
```

warpbreaks

```
em_tension <- emmeans(m, pairwise ~ tension)
ci_tension <- confint(em_tension, type = "response", adjust = "none")
ci_tension

## $emmeans
##  tension rate   SE  df asymp.LCL asymp.UCL
##  L       36.2 1.42 Inf      33.5      39.1
##  M       26.2 1.21 Inf      24.0      28.7
##  H       21.6 1.09 Inf      19.5      23.8
##
## Results are averaged over the levels of: wool
## Confidence level used: 0.95
## Intervals are back-transformed from the log scale
##
## $contrasts
##  contrast ratio     SE  df asymp.LCL asymp.UCL
##  L / M     1.38 0.0831 Inf      1.23      1.55
##  L / H     1.68 0.1074 Inf      1.48      1.90
##  M / H     1.22 0.0832 Inf      1.07      1.39
##
## Results are averaged over the levels of: wool
## Confidence level used: 0.95
```

warpbreaks

```
em_wool <- emmeans(m, pairwise ~ wool)
ci_wool <- confint(em_wool, type = "response", adjust = "none")
ci_wool

## $emmeans
##  wool rate    SE  df asymp.LCL asymp.UCL
##  A    30.3 1.061 Inf      28.3      32.5
##  B    24.7 0.955 Inf      22.9      26.6
##
## Results are averaged over the levels of: tension
## Confidence level used: 0.95
## Intervals are back-transformed from the log scale
##
## $contrasts
##  contrast ratio     SE  df asymp.LCL asymp.UCL
##  A / B     1.23 0.0634 Inf      1.11      1.36
##
## Results are averaged over the levels of: tension
## Confidence level used: 0.95
## Intervals are back-transformed from the log scale
```

warpbreaks

```
em_tension_by_wool <- emmeans(m, pairwise ~ tension | wool)
ci_tension_by_wool <- confint(em_tension_by_wool, type = "response", adjust = "none")
ci_tension_by_wool$emmeans

## wool = A:
##  tension rate   SE  df asymp.LCL asymp.UCL
##  L       40.1 1.82 Inf      36.7      43.9
##  M       29.1 1.50 Inf      26.3      32.2
##  H       23.9 1.33 Inf      21.4      26.6
##
## wool = B:
##  tension rate   SE  df asymp.LCL asymp.UCL
##  L       32.7 1.58 Inf      29.7      35.9
##  M       23.7 1.28 Inf      21.3      26.3
##  H       19.4 1.13 Inf      17.4      21.8
##
## Confidence level used: 0.95
## Intervals are back-transformed from the log scale
```

warpbreaks

```
ci_tension_by_wool$contrasts

## wool = A:
##  contrast ratio     SE  df asymp.LCL asymp.UCL
##  L / M     1.38 0.0831 Inf      1.23      1.55
##  L / H     1.68 0.1074 Inf      1.48      1.90
##  M / H     1.22 0.0832 Inf      1.07      1.39
##
## wool = B:
##  contrast ratio     SE  df asymp.LCL asymp.UCL
##  L / M     1.38 0.0831 Inf      1.23      1.55
##  L / H     1.68 0.1074 Inf      1.48      1.90
##  M / H     1.22 0.0832 Inf      1.07      1.39
##
## Confidence level used: 0.95
## Intervals are back-transformed from the log scale
```

warpbreaks

```
em_wool_by_tension <- emmeans(m, pairwise ~ wool | tension)
ci_wool_by_tension <- confint(em_wool_by_tension, type = "response", adjust = "none")
ci_wool_by_tension$emmeans

## tension = L:
##  wool rate   SE  df asymp.LCL asymp.UCL
##  A    40.1 1.82 Inf      36.7      43.9
##  B    32.7 1.58 Inf      29.7      35.9
##
## tension = M:
##  wool rate   SE  df asymp.LCL asymp.UCL
##  A    29.1 1.50 Inf      26.3      32.2
##  B    23.7 1.28 Inf      21.3      26.3
##
## tension = H:
##  wool rate   SE  df asymp.LCL asymp.UCL
##  A    23.9 1.33 Inf      21.4      26.6
##  B    19.4 1.13 Inf      17.4      21.8
##
## Confidence level used: 0.95
## Intervals are back-transformed from the log scale
```

warpbreaks

```
ci_wool_by_tension$contrasts

## tension = L:
##  contrast ratio     SE  df asymp.LCL asymp.UCL
##  A / B     1.23 0.0634 Inf      1.11      1.36
##
## tension = M:
##  contrast ratio     SE  df asymp.LCL asymp.UCL
##  A / B     1.23 0.0634 Inf      1.11      1.36
##
## tension = H:
##  contrast ratio     SE  df asymp.LCL asymp.UCL
##  A / B     1.23 0.0634 Inf      1.11      1.36
##
## Confidence level used: 0.95
## Intervals are back-transformed from the log scale
```

warpbreaks

The estimated mean number of breaks is

```
pm <- ci_tension_by_wool$emmeans %>%
  as.data.frame() %>%
  mutate(mean_with_ci = paste0(
    round(rate), " (", round(asymp.LCL), ", ", round(asymp.UCL), ")")
  ) %>%
  tidyr::pivot_wider(id_cols = tension, names_from = wool, values_from = mean_with_ci)
pm

## # A tibble: 3 x 3
##   tension A            B
##   <fct>   <chr>        <chr>
## 1 L       40 (37, 44) 33 (30, 36)
## 2 M       29 (26, 32) 24 (21, 26)
## 3 H       24 (21, 27) 19 (17, 22)
```

warpbreaks

While holding wool type constant, the percent change in mean number of breaks is

```
ci_tension$contrasts %>% as.data.frame %>%
  mutate(
    change = 100*(ratio-1),
    lower  = 100*(asymp.LCL-1),
    upper  = 100*(asymp.UCL-1),

    change_with_ci = paste0(
      round(change), " (", round(lower), ", ", round(upper), ")"
    )) %>%
  select(contrast, change_with_ci)

##   contrast change_with_ci
## 1    L / M    38 (23, 55)
## 2    L / H    68 (48, 90)
## 3    M / H     22 (7, 39)
```

Summary

With count data (with no upper limit), a Poisson regression model is appropriate.