STAT550 Elderly Falls Analysis Project Report

Hailey Ernest

2026-03-14

Problems Encountered

Two main problems during the analysis of fall counts in elderly participants I found were overdispersion and many zeros.

Overdispersion

Initial Poisson models showed evidence of overdispersion, since the variance of fall counts was higher than the mean in both exercise groups. The variance to mean ratios of the groups are 2.28 and 1.57. A Poisson model assumes $\mathrm{Var}(Y) = E(Y)$. It isn't met, which leads to underestimated standard errors and test results look more significance. To fix this, I used a Negative Binomial model to compare to Poisson because it has a dispersion parameter θ that allows the variance to be higher than the mean.

Many zero counts

Thirty nine elderly participants (39%) reported zero falls. This showed that the data possibly has more zeros than expected under a Poisson model. Thus, Zero-inflated (ZIP, ZINB) and hurdle models were also fit. Zero-inflated and hurdle models resulted in slightly lower AIC values, but the lowest BIC value was the Negative Binomial model because the other models need many more parameters. The sample size ($n = 100$) is not very large, so the simpler NB model explained the data more efficiently.

Variable and interaction selection

I used Likelihood ratio tests were to check for possible interaction terms involving exercise with balance, strength, and gender. All interactions had weak evidence of significance at the $\alpha = 0.10$ level. Using backward elimination and Type II likelihood ratio tests to evaluate the importance of each predictor, Gender had weak evidence of significant ($p = 0.499$) and was removed from the model. The final model retained exercise, balance, and strength as predictors.

Final Model

The final model is a Negative Binomial GLM.

Random component

$$Y_i \sim \text{NB}(\mu_i,\, \theta), \quad \text{Var}(Y_i) = \mu_i + \frac{\mu_i^2}{\theta}$$

The estimated dispersion parameter $\hat{\theta} = 4.75$, shows overdispersion.

Link function

Log link was used: $\log(\mu_i) = \eta_i$.

Linear predictor

$$\eta_i = \beta_0 + \beta_1(\text{exercise}) + \beta_2(\text{balance}) + \beta_3(\text{strength})$$

Interaction terms and predictor Gender weren't retained in the final model.

Interpretation of Model Coefficients

Intercept:

The intercept represents the expected fall count for a non-exercising participant with balance = 0 and strength = 0. These below the observed data range, so the intercept isn't interpretable in context.

Exercise **($p < 0.001$)**:

The estimated rate ratio for exercise is 0.153. Participants who exercise are expected to have 0.153 times the fall rate of non-exercisers with the same balance and strength scores, or a decrease of about 84.7% in expected falls, after adjusting for Balance and Strength.

Balance **($p = 0.002$) :**

The rate ratio for balance is 1.014. Each one unit increase in balance score multiplies the expected number of falls by 1.014, or an increase of about 1.4% in expected falls, after adjusting for Exercise and Strength.

Strength **($p = 0.003$):**

The rate ratio for strength is 1.020. Each one unit increase in strength score multiplies the expected number of falls by 1.020, or an increase of about 2.0% in expected falls, after adjusting for Balance and Exercise.

Remaining Issues

Two minor problems remain in the final model.

Zeros counts

The Negative Binomial model predicts about 33 (32.7) zero-fall observations compared to the 39 observed, showing only a small difference that is mostly explained by the variance of NB.

Moderate influence

A few observations were over the standard Cook's distance threshold of $4/n$. No observation was over the extreme influence threshold of 1. Observation 92 showed moderately high leverage and Cook's distance, so I used a sensitivity check to check it and got identical results, so it was retained. Therefore, all observations were retained in the final model.

Predictions

Predicted fall counts were generated for all combinations of the predictors:

- **Exercise:** 0 (no), 1 (yes)
- **Balance:** min = 13, mean = 53, max = 98
- **Strength:** min = 18, mean = 61, max = 90

This creates $2 \times 3 \times 3 = 18$ prediction combinations.

Overall, exercise looks to have lower predicted fall counts across all combinations of balance and strength values. The table of predicted counts and confidence intervals is provided below in Table 1.

Tables

Table 1. Predicted Fall Counts

Table 1: Predicted fall counts from the final model.

Exercise	Balance	Strength	Predicted Falls	95% CI Lower	95% CI Upper
No	13 (min)	18 (min)	0.94	0.46	1.92
No	13 (min)	61 (mean)	2.16	1.42	3.30
No	13 (min)	90 (max)	3.79	2.21	6.49
No	53 (mean)	18 (min)	1.63	0.90	2.96
No	53 (mean)	61 (mean)	3.75	3.09	4.55
No	53 (mean)	90 (max)	6.57	4.42	9.77
No	98 (max)	18 (min)	3.03	1.52	6.07
No	98 (max)	61 (mean)	6.97	4.57	10.63
No	98 (max)	90 (max)	12.21	7.01	21.26
Yes	13 (min)	18 (min)	0.14	0.06	0.33
Yes	13 (min)	61 (mean)	0.33	0.19	0.57
Yes	13 (min)	90 (max)	0.58	0.31	1.09
Yes	53 (mean)	18 (min)	0.25	0.12	0.51
Yes	53 (mean)	61 (mean)	0.57	0.39	0.84
Yes	53 (mean)	90 (max)	1.01	0.61	1.65
Yes	98 (max)	18 (min)	0.46	0.21	1.01
Yes	98 (max)	61 (mean)	1.07	0.64	1.79
Yes	98 (max)	90 (max)	1.87	1.01	3.45

Table 2. Final Model Output

Table 2: Final model output: falls   exercise + balance + strength.

Predictor	Estimate	SE	Rate Ratio	CI Lower	CI Upper	p-value
Intercept	-0.5866	0.4909	0.5562	0.2104	1.4213	0.2321
Exercise	-1.8768	0.2165	0.1531	0.0986	0.2310	0.0000
Balance	0.0138	0.0045	1.0139	1.0052	1.0227	0.0023
Strength	0.0193	0.0064	1.0195	1.0071	1.0324	0.0027