Journal of Humanistic Mathematics

Volume 7 | Issue 1 January 2017

Moneyball for Creative Writers: A Statistical Strategy for Publishing Your Work

Jon Wesick

Recommended Citation

Jon Wesick, "Moneyball for Creative Writers: A Statistical Strategy for Publishing Your Work," *Journal of Humanistic Mathematics*, Volume 7 Issue 1 (January 2017), pages 155-171. DOI: 10.5642/jhummath.201701.13. Available at: https://scholarship.claremont.edu/jhm/vol7/iss1/13

Moneyball for Creative Writers: A Statistical Strategy for Publishing Your Work

Jon Wesick

Carlsbad, CA, USA
jwesick@sbcglobal.net

Synopsis

Writers face a challenge getting their poems and stories published. Rather than following the traditional strategy I model creative writing submission as a statistical process and explore the use of numerical metrics to maximize publications.

Introduction

To publish poems or stories a writer typically selects a journal from *Poets and Writers Magazine*, the *Writer's Market* publications (such as *Poet's Market* and *Novel & Short Story Writer's Market*) or the *Duotrope* website, submits his work by mail or e-mail, and then waits for a reply. If his work is rejected, he repeats the process until a journal accepts it. How should writers choose which magazines to send their writing to?

Writing books suggest reading many journals to get a feel for editors' preferences and then target those that best fit the writer's style. I have known writers who have used this method successfully but for me detailed research would be burdensome. With hundreds of pieces to publish, carefully evaluating the best fit for each could easily take thousands of hours, hours that could be more profitably spent writing. In addition, I find editors' preferences subjective and difficult to categorize. One editor told me he seldom accepted stories written in the first person. I would have never dreamed that would matter to a reader. For me a story's literary merit has very little to do with whether it is written from the first or third person point of view.

Despite spending hundreds of dollars on sample copies and countless hours reading websites, I have only analyzed a few editors' preferences well enough to succeed in repeat acceptances of my work. Having a physics and engineering background I decided to explore a different publication strategy, one that liberal arts majors are not likely to employ. I used math.

Numerical Strategies

The numerical strategy assumes publishing poems and stories is a statistical process and uses numerical measures to optimize a writer's chance of getting his writing in front of the most readers' eyes. The numbers are only a place to start. A writer still needs to examine candidate journals to see if his writing fits there. For example, sending erotic stories to *Highlights for Children* will not get anyone published no matter what the acceptance percentages. Also writers may be disappointed to have their work appear along with other work of poor quality in journals that accept almost everything.

Many of the entries for publications in the *Poet's Market* and *Novel & Short Story Writer's Market* list press run, response time for submissions, number of submissions, and number of submissions accepted for publication. I have converted the last two into the percent of submissions accepted by taking the ratio and multiplying by one hundred.[1] Clearly there's a better chance of getting writing accepted at journals with higher percent accepted and shorter response time, a fact that surprised me quickly. I had not had a short story published in years, but when I first submitted to magazines with higher percentages, I had a few accepted within weeks.

Press run is an imperfect metric of a publication's quality. Typically the bigger a magazine's press run the better but not always. For example, some religious and horror magazines have large press runs but the writing included may be amateurish and publication in these will not necessarily convey much

[1]Since authors typically submit several poems to a journal at the same time, there is an ambiguity in the number of submission for poetry journals. The number of submissions could mean the total number of poems submitted or only the number letters or e-mails each containing several poems. A similar ambiguity exists for acceptances as well. In this essay I assume the number of submissions is the total number of poems submitted and the number of acceptances is the total number of poems accepted.

prestige for a poet or fiction writer. Also most online journals do not report press run. Even if they do report the number of hits, it is not clear whether online circulation is equivalent to that of print journals. That being so, I have excluded online journals from any analysis that involves press run.

Despite these limitations, I have combined all relevant metrics into an overall measure that I call a magazine's *Figure of Merit* (FOM), which is the percent accepted times the press run divided by the response time in months. The bigger a magazine's FOM the better the opportunity for submitting.

Taken in isolation the metrics for one journal do not mean very much. For example, is submitting a story to a journal with a press run of 200, a 1% acceptance, and response time of 9 months a good choice? To decide, a writer needs to know how this journal compares to others. In order to find out, I have extracted the numerical data from the *Poet's Market* and *Novel & Short Story Writer's Market* (about 500 journals with complete numerical data in each), entered them into spreadsheets, and calculated the averages (means) which are shown in Table 1.

	Poems	**Short Stories**
% Accepted	11.4%	4.9%
Response Time	3.4 months	3.9 months

Table 1: Average Percent Accepted and Response Time.

Values for individual magazines have a lot of spread around the averages. Standard deviations are roughly the size of the means and the distributions are skewed to the right. I have plotted these distributions in Appendix A.[2] Table 1 does not include averages for press run and FOM because outliers with huge press runs drastically change the values when they are not included.

Writer's Digest Books did not respond when I asked about the source of the metrics listed in the *Poet's Market* and *Novel & Short Story Writer's Market*. I believe the metrics are what magazine editors report. As a cross-check on their accuracy I compared them to values reported on the *Duotrope* website. *Duotrope* collects submission statistics from writers. *Duotrope* gives overall averages of roughly 6% to 7% acceptances and 65 days for response times

[2]The plots of press run and FOM in Appendix A are not true histograms since the bins are not of constant size.

for both fiction and poetry. These are in the ballpark of the numbers in Table 1. *Duotrope* also mentions that 8% to 12% of journals never respond to submissions, something an editor would be reluctant to admit about his magazine.

I have extended this comparison a bit further by plotting percent accepted and response times from *Duotrope* (with 30 or more data points) versus those from the *Poet's Market* or *Novel & Short Story Writer's Market*. As an example, Figure 1 compares response times for fiction submissions. If there were good agreement the data points would cluster around a line with slope one. Instead the points form a jumble. Both stories and poems, response times and acceptance rates show the same disagreement. There may be several explanations for this. *Duotrope* mentions that writers who submit data to them tend to underreport rejections. Writers choose when to list a publication as not responding on *Duotrope*, which may bias the response time toward shorter values. Writers who share data with *Duotrope* may not be not a reliable sample. This may be especially true now that *Duotrope* has starting charging users fees. And finally, editors who report values to the *Poet's Market* or *Novel & Short Story Writer's Market* may be poor at estimating what their statistics really are.

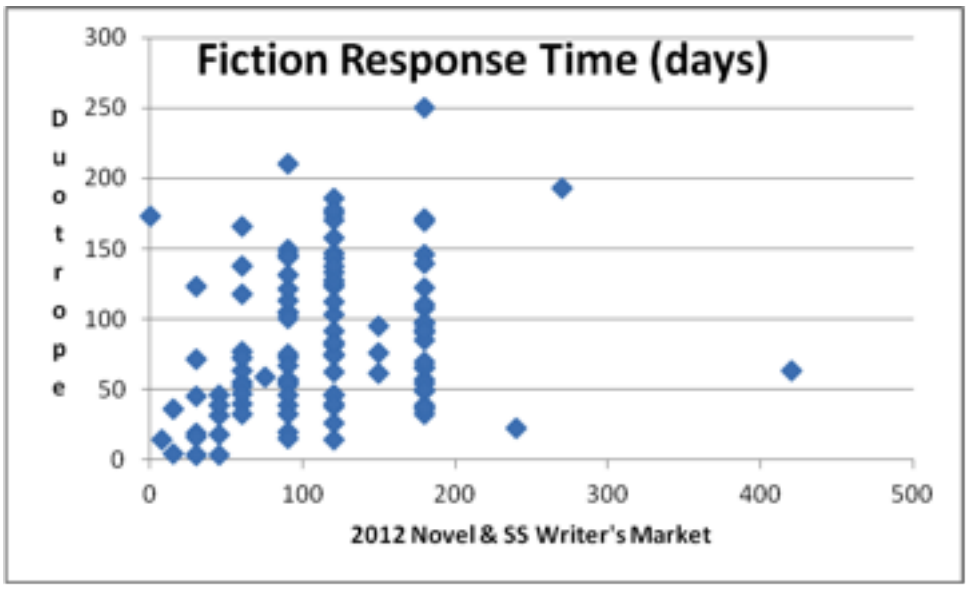


Figure 1: Journal Response Time for Fiction Duotrope vs. NSSWM.

As for the example question about submitting to the sample journal, the 1% acceptance percentage is smaller than average, the 9-month response time longer, and a press run of 200 is not very large. Taking these into account a writer would have a better chance submitting to a different journal.

Another lesson to take away from the percent accepted is not to be discouraged. Acceptances are rare. On average a writer must submit a piece many times before it gets published. If we make the rough approximation that all journals have the same acceptance rate p (I will use the mean acceptance percentage divided by 100), then the chance, C, of getting accepted after n tries ($n-1$ rejections followed by one acceptance) is given by the geometric distribution

$$C = p(1-p)^{n-1}.$$

Figure 2 is a histogram of how many rejections 211 of my poems received before getting accepted.

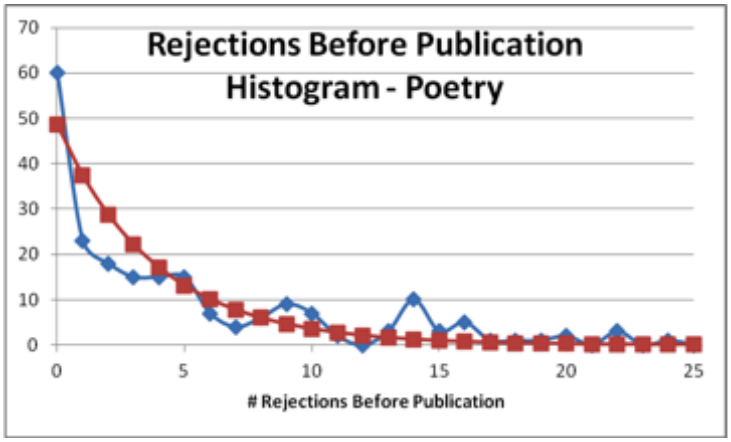


Figure 2: Number of Rejections Before Publishing Histogram.

The jagged curve is my actual data while the smooth curve is a calculation that uses the above equation with $p = 0.23$ (which gives the best fit by minimizing the chi square).[3] Both curves show reasonable agreement.[4]

[3]My probability of acceptance is somewhat high here due to several publications that accept a lot of my poems.

[4]Just for completeness the Negative Binomial Distribution gives the chance of getting r success after n tries. I find that you'd have to have roughly 500 tries to have an 80% chance

To show what this means in practice, I also provide, in Figure 3, a histogram that shows how long it took me to publish poems after completing them.

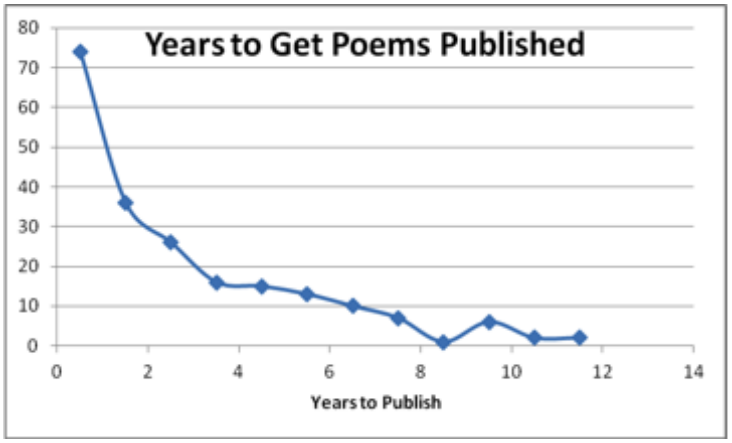


Figure 3: Number of Years Between Poems' Completion and Publication.

A modification of the above formula can help determine when a writer should stop submitting a piece that gets continually rejected. I'd consider giving up on a piece (or radically revising it) when the chance of n rejections is 5%, that is when $(1-p)^n = .05$. Using the acceptance rates in the table gives 25 rejections for poems and 60 rejections for stories.

If my experience is representative of that of other writers, there is another reason not to be discouraged. I have found that my better pieces got more rejections before eventual publication than my mediocre ones. Table 2 compares the mean number of rejections before publication for my good and average poems and stories. SD is the standard deviation and N is the number of pieces accepted. I have excluded pieces published in local journals (and from one blog that publishes a large number of my political poems) from the analysis. Clearly the number of rejections is greater for what I consider my better work. In order to determine this is not due to random variations, I performed hypothesis tests for both poems and stories. These calculations show the results to be statistically significant with a z value of 3.82 for poems and a t value of 2.70 for stories.

---

of publishing 20 stories. See "A Probabilistic Approach to Determining the Number of Units to Build in a Yield-Constrained Process" by Timothy P. Anderson (*Journal of Cost*

	Average Rejections	SD	N
Best Poems	8.75	7.03	68
Average Poems	4.82	4.39	57
Best Stories	9.56	6.94	27
Average Stories	4.09	4.68	23

Table 2: Rejections Before Publications for Best and the Rest of Pieces.

Which Strategy is Best?

Several numerical publication strategies are possible. A writer could submit to magazines that respond quickest (fast) or accept the most (easy) to get the maximum number of publications. He could submit to journals with the largest press run (circ) to maximize the number of eyes that could see his writing. Or he could try to maximize both journals and press run by submitting to journals with the largest Figure of Merit (FOM). To find which is best I performed Monte Carlo simulations using VBA in Microsoft Excel.

In the Monte Carlo technique, the outcome of a statistical event is determined by a computer-generated random number. By repeating the event thousands of times, one builds up a distribution of probabilities associated with that event. One can then find averages, standard deviations, and so on. In my calculation I simulated submitting 50 stories to magazines with parameters I extracted from the *Novel & Short Story Writer's Market* for a period of 36 months and counted the number of stories that would be published using the percent accepted values. The simulation then repeated this 36-month period for a thousand trials. Each trial resulted in a slightly different number of publications. Figure 4 shows one of the results in the form of a cumulative probability distribution. One interprets this as meaning there is a 20% chance of getting at least 32 publications, a 50% chance of getting at least 29, and an 80% chance of getting at least 26.

I ran calculations for the four different strategies discussed as well as for a random strategy in which any publication could be chosen regardless of numbers to serve as a control. In addition to the number of publications I also kept track of cumulative press run, which is the number of acceptances weighted by the accepting journal's circulation or the total number of printed

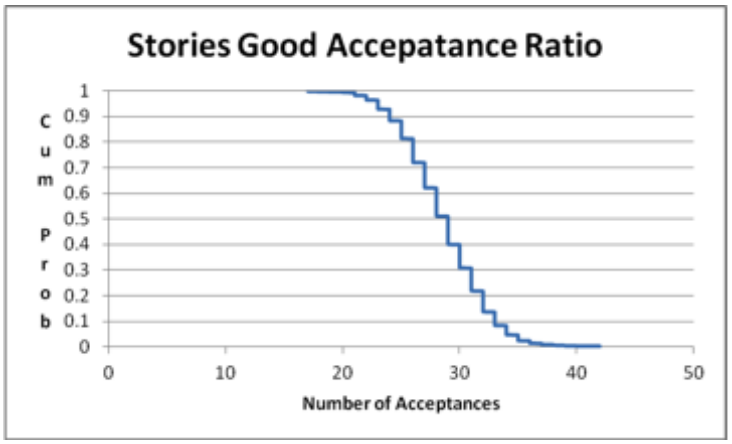


Figure 4: Cumulative Distribution of Publications for Easy Strategy. awkward caption on figure!

copies that contain the accepted stories. For example, if a story is published in magazine A with a press run of 500 and another story is published in magazine B with a press run of 250, the cumulative press run would be 750. Similar calculations were done for poems with the added complication that more than one poem is submitted at a time to the same magazine. In the poetry simulations, poems were submitted in batches of five. Simultaneous submissions (one piece sent to two or more journals at the same time) were not included in the model. Appendix B contains more details about the calculations.

Figures 5 through 8 on the following pages show the results. Both for stories and poems, submitting to the fastest responders yields the greatest number of publications. Submitting to those with the highest acceptance rate comes in second. Interestingly, submitting to magazines with the largest press run yields the fewest publications. This makes sense. A writer who only sends work to the New Yorker is not likely to get many acceptances.

As for maximizing cumulative circulation, the FOM strategy does best followed by the circulation strategy. Random does worst for stories and the fastest responders do worst for poems.

The best strategy depends on a writer's goal. Submitting to the fastest responders yields the most publications probably because the quick turnaround outweighs any advantage due to journals with high acceptance rates. Sub-

mitting to journals with the largest FOM yields the largest cumulative press run, even more than submitting to the journals with the biggest press runs probably because the latter are so hard to get into.

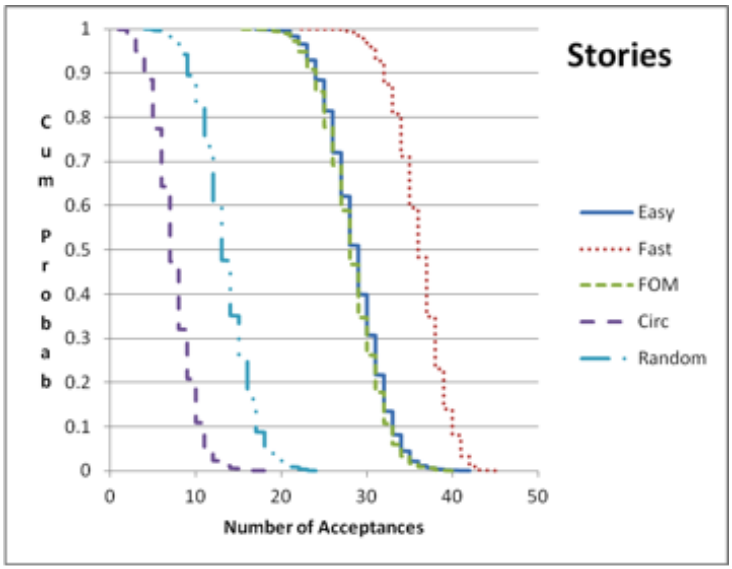


Figure 5: Number of Story Publications from Monte Carlo.

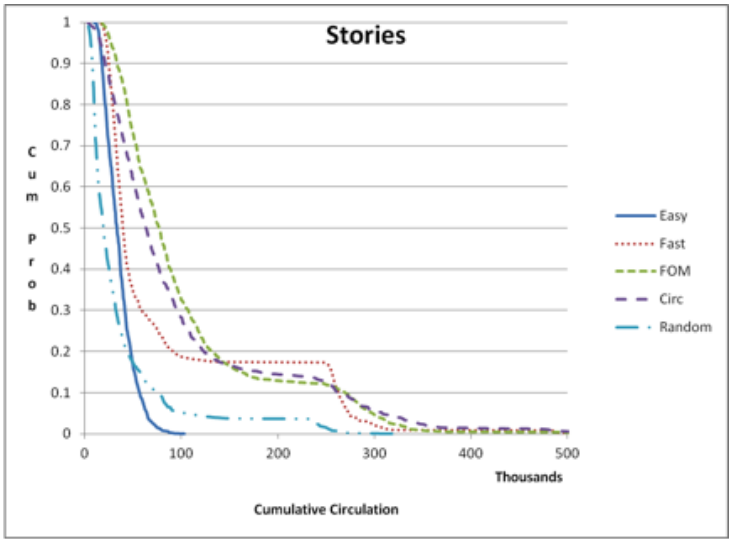


Figure 6: Cumulative Press run of Stories from Monte Carlo.

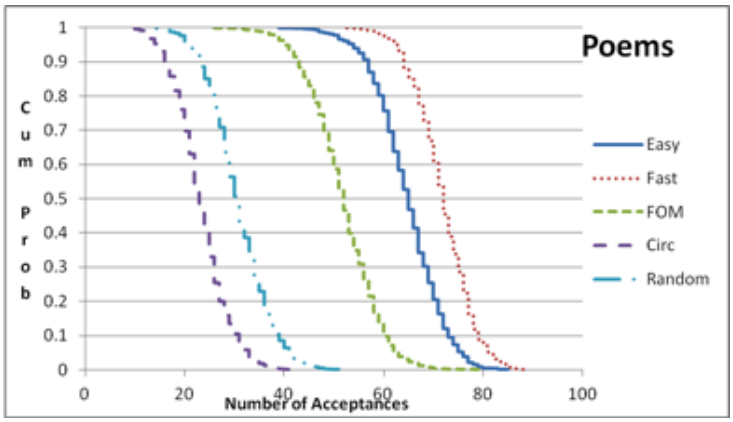


Figure 7: Number of Poem Publications from Monte Carlo.

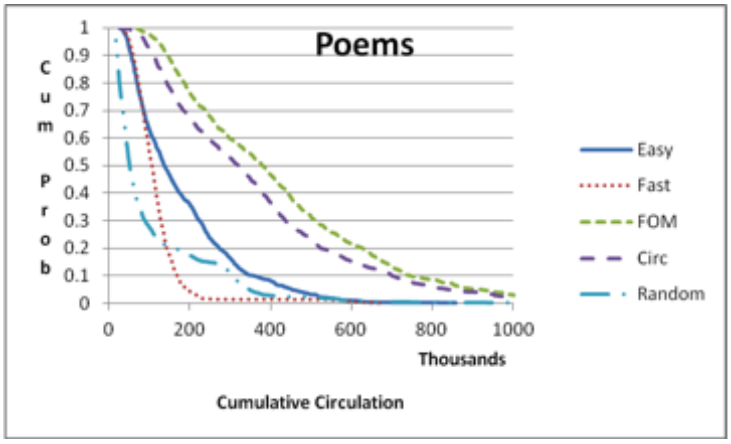


Figure 8: Cumulative Press run of Poems from Monte Carlo.

Testing the Numerical Strategy

The simulations do not say anything about how the numerical strategies compare to the traditional submission strategy, so I ran a test to find out how the numerical strategy works in the real world. Since cumulative circulation is difficult to measure, I chose to optimize the number of publications. For the numerical strategy I submitted to the publications with the shortest response time (less that three months and often quicker). I compared this to my typical method of submitting. This control strategy consisted of submitting to journals that look relevant, new journals, and those that publish authors

I like. I randomized 71 poems and 39 stories so I could submit roughly half by the numerical method and half by the traditional method.[5] I then submitted the pieces according to their chosen strategies for twelve months and tabulated the results as shown in Table 3. Here the Number Accepted are those pieces accepted during the one-year test period. If I'd submitted a piece to a publication during the test period and they accepted it after the close of testing, I did not count this in the Number Accepted.

Type of Submission	Number Submitted	Number Accepted
Poetry Numerical Strategy	34	17
Poetry Control Strategy	37	9
Fiction Numerical Strategy	18	12
Fiction Control Strategy	21	7

Table 3: Comparison of Numerical Strategy with Control.

For both poems and stories, the numerical strategy yielded roughly twice the acceptances as the control strategy. To determine whether these results are statistically significant I performed a hypothesis test of the ratios. Both results for the poems and the stories were statistically significant with p values of less than 5%. I conclude that following the numerical strategy yields more publications than my standard submission strategy.

Seasonal Variation

For years I have suspected that few of the poems and stories I submit in summer get accepted. Perhaps this is because many university-based publications take the summer off or it may be due to how I submit. I typically buy the latest *Poet's Market* and *Novel & Short Story Writer's Market* in August or September and immediately send out a flood of submissions. By summer I may have used up my good leads.

To determine whether this seasonal variation is indeed true, I examined my submission records for four years (2008 through March 2012) and plotted the percent of poems and stories accepted versus season. Since I'm comparing ratios, any seasonal difference due to the total number of pieces submitted should divide out. Figure 9 shows the results for poems.

[5] I randomized these by a coin flip with some minor adjustments.

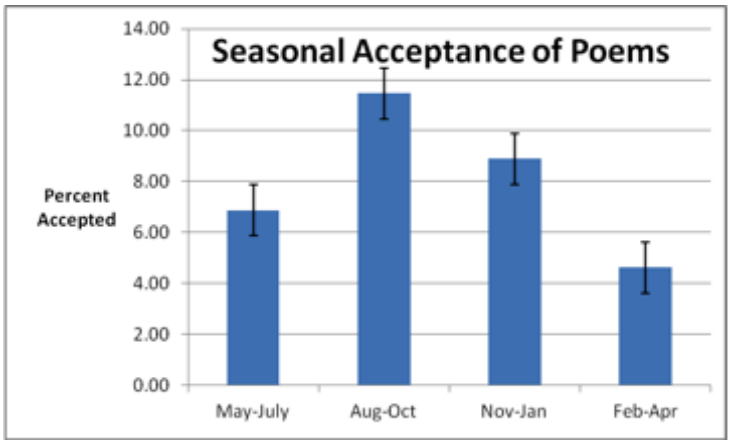


Figure 9: Seasonal Variation of Poetry Acceptances.

The data certainly seem to imply that there is a seasonal variation. To be sure they are statistically significant, I performed a ratio hypothesis test comparing the August through January data with the February through July data. The results show that the seasonal variation for poems is statistically significant ($p = 0.2$) while there is not enough data for stories to make a conclusion ($p = 0.8$). The details are shown in Table 4.

	Poems Accepted	**Poems Submitted**	**Stories Accepted**	**Stories Submitted**
Aug–Jan	44	409	26	156
Feb–July	13	261	9	131

Table 4: Seasonal Variation Hypothesis Tests.

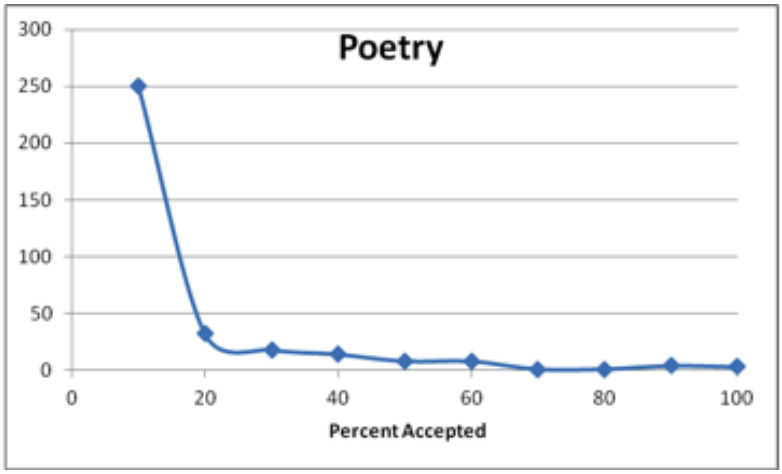
Conclusion

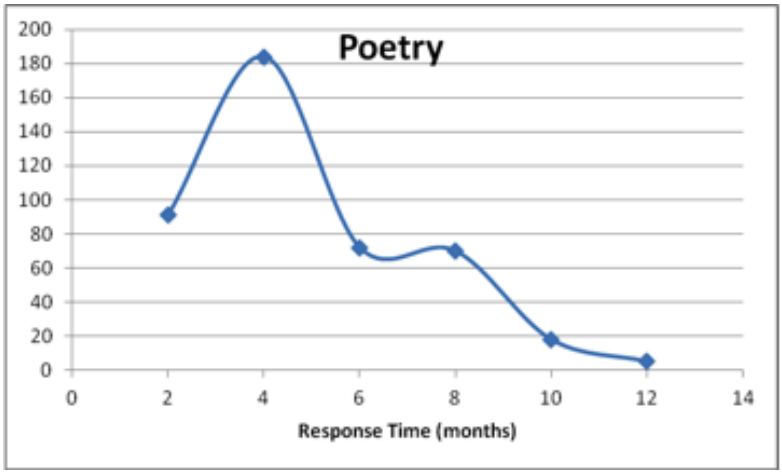
The relevant questions for a writer submitting poems or stories are not just what and where but also how many. Tracking publication metrics and using simple statistics has helped me focus my submissions to maximize my publications. It has also been a potent weapon against despair by allowing me to see how common rejection is for all writers. In conclusion I recommend writers keep the Japanese martial artists' motto in mind. *Ganbatte!* Keep trying!

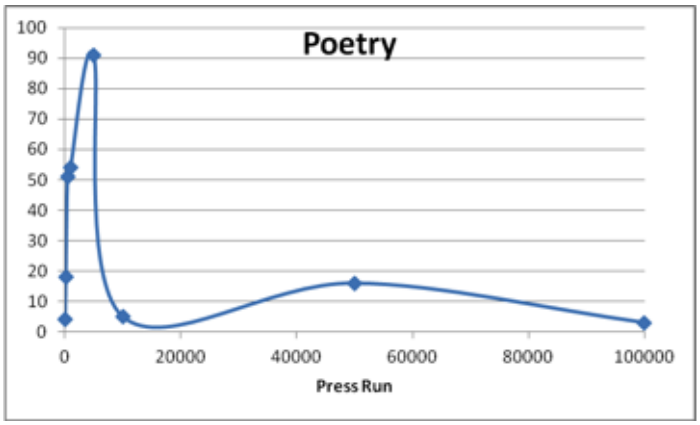
A. Distributions from *Poet's Market* and *Novel & Short Story Writer's Market*

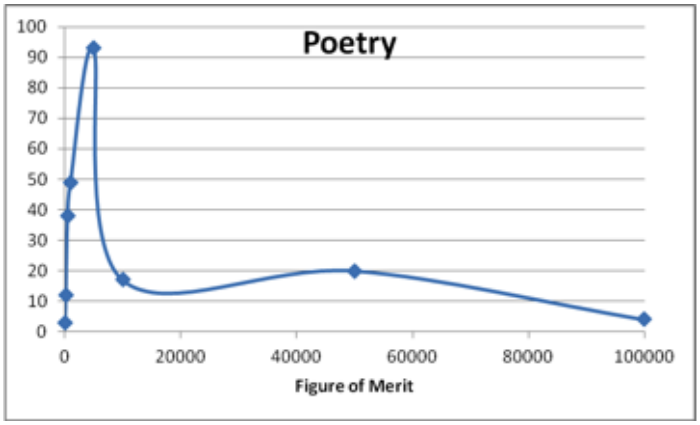
In this appendix I present plots for the distributions of acceptance rates, response time, press run, and FOM for poetry and story submissions that I have computed using the numerical data I extracted from *Poet's Market* and *Novel & Short Story Writer's Market* (about 500 journals with complete numerical data in each).

Distributions for poetry submissions:

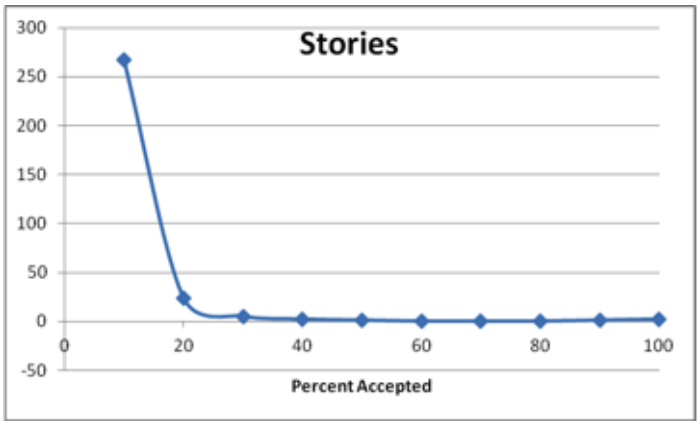


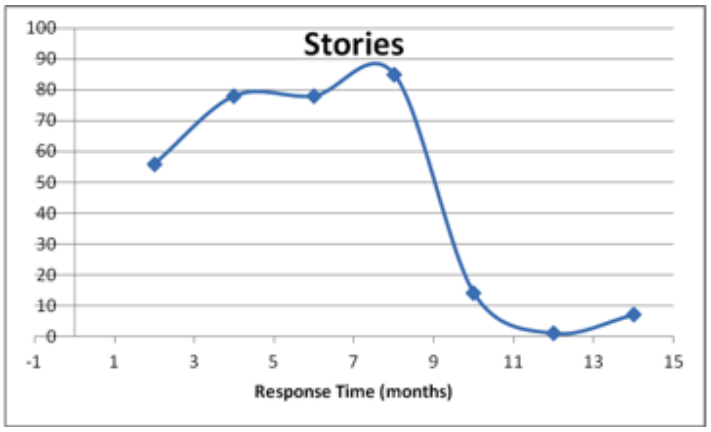


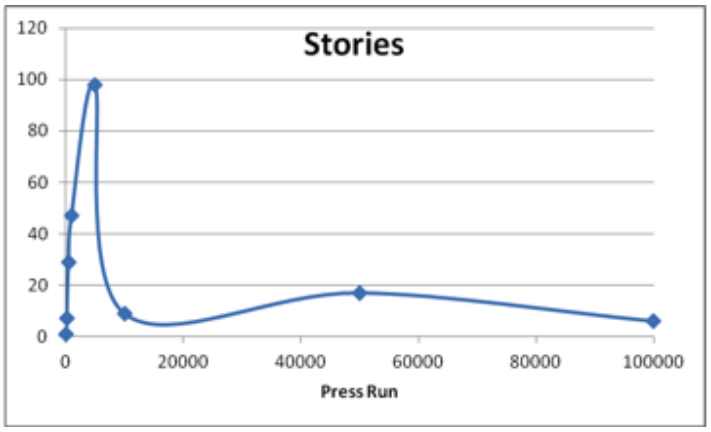


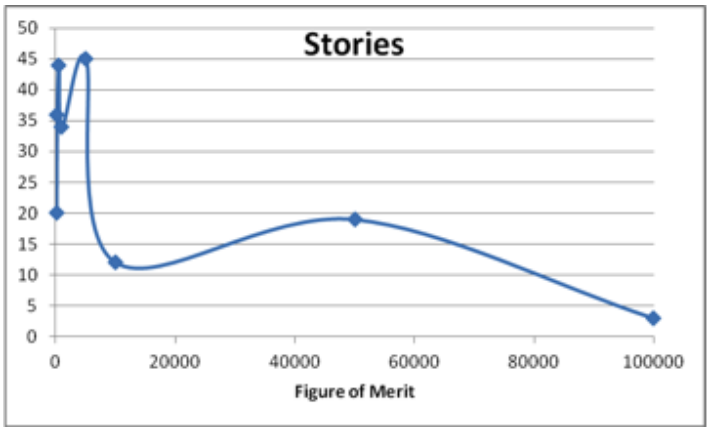


Distributions for story submissions:




Stories
100
90
80
70
60
50
40
30
20
10
0
-1
1
3
5
7
9
11
13
15
Response Time (months)


Stories
120
100
80
60
40
20
0
0
20000
40000
60000
80000
100000
Press Run


Stories
50
45
40
35
30
25
20
15
10
5
0
0
20000
40000
60000
80000
100000
Figure of Merit

B. Details of Monte Carlo Calculations

Algorithm:

Response time, percent acceptance, circulation, and figure of merit data from the 2011 *Poet's Market* or *Novel & Short Story Writer's Market* were stored in an Excel spreadsheet with rows corresponding to different journals.

For fiction, the software started with a given number of stories to submit. For each story, it chose a journal from the spreadsheet using a random number to pick the row. For the random strategy, all rows were available. For Easy strategy, the rows were ordered from highest percent accepted to lowest, and only the first 50 rows were available. Other strategies followed a similar algorithm. Once the story was "submitted", the software kept track of when the journal would respond (current month plus response time in months). When the month counter said a response was due, the software used the results of a random number generator to determine if the story was accepted. For example, if a journal's acceptance rate is 10%, the random number generator would have to return a number less than 0.10 for the story to be accepted. If the story was accepted, the number of published stories was incremented as well as the cumulative circulation. If the story was rejected, it was returned to the pool of stories available for submission. Each trial lasted for a submission period of 36 months and there were 1000 trials.

A similar Monte Carlo calculation was performed for submitting poems. It was a bit more complicated because it allowed submission of more than one poem to the same journal (typically 3 to 5). Binomial statistics were used to determine how many of the submitted poems a given journal accepted. In the simulations poems sent out in batches of three get more publications than when sent out in batches of five. This is an artifact that comes about from waiting until there are enough free poems to make a batch size before submitting. As such I have only discussed poems submitted in batches of five in this article.

Neither version of my simulations included simultaneous submissions. Once published a piece was not submitted again.

Short Stories

Number of Stories = 50

Length of Submission Period = 36 months

Number of Trials = 1000

Strategies:

Random: all 177 publications used
Easy: 50 publications with biggest percent accepted used
Fast: 50 publications with shortest response time used
Press run: 50 publications with biggest press run used
FOM: 50 publications with biggest Figure of Merit used

Poems

Number of Poems = 100

Poems per Submission = 5, poems sent to same journal in groups of 5

Length of Submission Period = 24 months

Number of Trials = 1000

Strategies:

Random: all 215 publications used
Easy: 50 publications with biggest percent accepted used
Fast: 50 publications with shortest response time used
Press run: 50 publications with biggest press run used
FOM: 50 publications with biggest Figure of Merit used