Understanding Agricultural Insurance Loss Ratio Characteristics

A study utilizing data from select wheat-growing regions in the United States, quantifying differences in losses incurred at the county level, with the goal of augmenting information available to agricultural insurance providers

ASA Project Competition 2020
Asmi Kumar

Introduction

Wheat is the most widely grown staple crop in the world, accounting for a fifth of globally consumed calories and continuing to rise in demand. In addition to being a major source of starch and energy, wheat also provides substantial amounts of many components essential for health, notably protein, vitamins, and dietary fiber. The state of Kansas leads the United States in wheat production. Its economy relies heavily on wheat, with about one-fifth of the country's total wheat production coming from it. Washington is also another leader producer of wheat.

In this project, the experiments presented provide insight into the differences in losses and gains incurred by agricultural insurance companies associated with different wheat-growing regions. The statistics obtained suggest that, to gain optimally, it is highly important for insurance companies to consider geography when formulating policies, i.e. to tailor them to a regional or even county level. In the next section, I evaluate whether Kansas and Washington exhibit different *loss ratio* (a ratio of losses to gains) distributions. To do this, I randomly chose a county from each one; these happened to be Morton, KS and Whitman, WA. Following this, I conduct regional-level experiments. For each of Morton and Whitman, I randomly selected a neighboring county, in terms of geographical proximity. These were Grant, KS and Skagit, WA. Running statistical tests to check whether the loss ratios of pairs of same-state counties differ, I provide additional evidence to support the need for developing insurance policies at the county level, as broader levels (e.g., regional) are sufficient only in some cases.

There are several key measurements that are necessary for calculating a loss ratio, which measures the success and profitability of a company. The first is *indemnity*, the sum of money an individual or business receives for qualifying losses paid under an insurance policy. The second is *premium*, the amount of money an individual or business pays for an insurance policy for risk protection. Loss ratio is calculated as shown:

loss ratio = indemnity ($) ÷ *total premium* ($)

By combining the factors of earned money and losses, loss ratios provide an effective way to evaluate whether an insurance company performed well in a given time period. It is ideal to keep the loss ratio as close to zero as possible, since this would mean paying minimal indemnities and receiving more premium benefits from customers. If a loss ratio grows too high, it signals to insurance companies that they may need to increase premiums.

Statistical Question

Do counties in geographically separate states exhibit statistically significant differences in their loss ratios? The null and alternative hypotheses, with significance level $\alpha = 0.05$, are:

$$H_0\text{:}\ \mu_{KS} = \mu_{WA},$$

$$H_a\text{:}\ \mu_{KS} \neq \mu_{WA},$$

where μ_{KS} represents the true average of the loss ratios of Kansas counties and μ_{WA} represents the true average of the loss ratios of Washington counties. The null states that the differences in loss ratios in Washington and Kansas counties are statistically insignificant, whereas the alternative states that they are statistically significant.

Data Collection

I collected one dataset for each of the examined counties in Washington and Kansas from the

USDA Agricultural Report Generator. The portal allows users to isolate a specific commodity (i.e., crop, in this case, wheat) and location down to the county level to analyze many different factors of that commodity. I used data from the USDA because it would be the most accurate and reliable source of data that covered locations across the country with consistency. First, I navigated to the “Summary of Business” page on the USDA’s Risk Management Agency page in order to access the insurance claims for all crop policies, not just ones with a loss. I then headed to the “Report Generator” tab, where I sorted the data by my desired commodity of wheat, selected the entire available time frame (1989 to 2019), and narrowed my datasets down to the aforementioned Kansas and Washington counties. Because the datasets obtained were separate, I directly merged them in Excel after downloading.

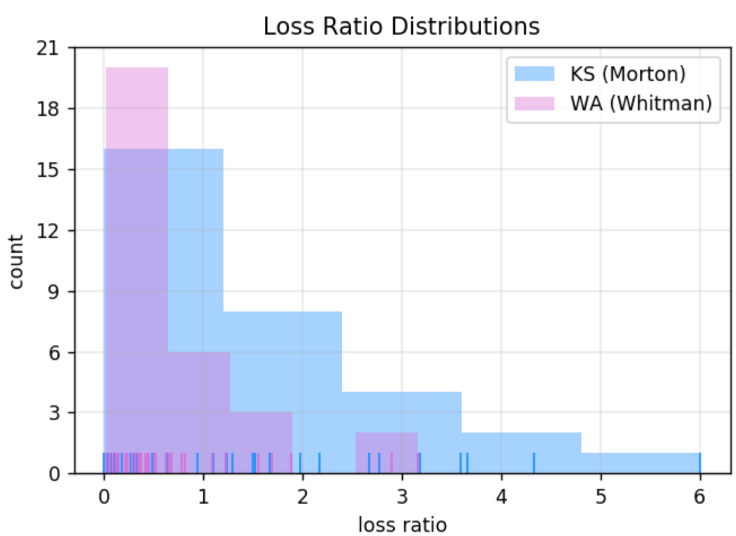
The valuable information gained from the Report Generator application includes many money-based measurements such as insurance premiums, indemnities, and ultimately, loss ratios. This information is needed to explore the trends of loss ratio in relation to location. The table below displays the raw data utilized in this project. Data is organized by county and year, and each cell displays the appropriate loss ratio for the respective county and year.

Commodity Year	Grant, KS	Morton, KS	Skagit, WA	Whitman, WA
1989	2.216181	3.178266	0	1.233943
1990	0.8195	0.001131		0.088986
1991	0.02655	0.098902		0.526554
1992	0.160988	0.145243		1.555545
1993	0.156027	1.669449		0.126128
1994	3.751074	3.655929		1.693317
1995	1.767409	4.335037	0	0.657643
1996	4.311336	6.007095	0.118555	0.062138
1997	0.10953	0.307944	0.222476	0.025961
1998	0.193834	0.152911	0.515913	0.284984
1999	0	0.063682	1.53755	0.816234
2000	0.562748	0.27355	2.490043	0.363594
2001	0.797164	0.642943	0	0.517444
2002	2.19443	3.595545	0.26381	0.424758
2003	1.688927	1.105589	0.149766	0.453011

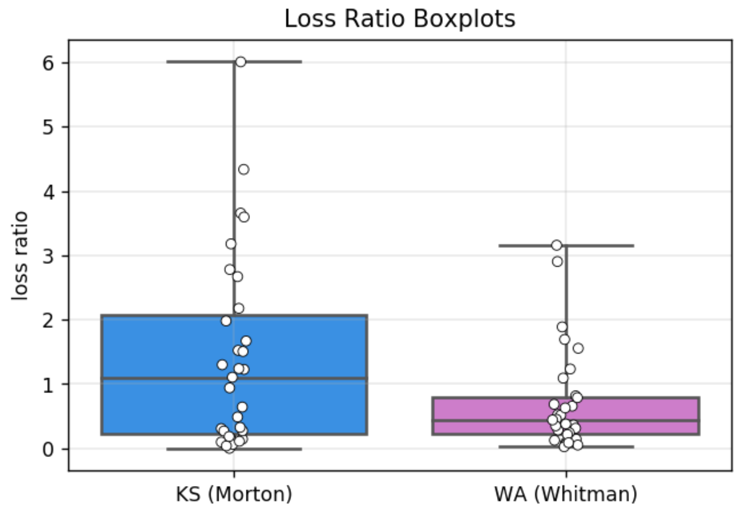
2004	1.546669	1.227428	0.787866	0.14495
2005	0.187529	0.326681	0.767894	0.788827
2006	3.785357	1.981035	0	0.350659
2007	0.071094	0.03953	0.542462	0.313869
2008	1.427966	2.176164	2.080537	1.092137
2009	1.038	1.523669	2.196538	3.15679
2010	0.049296	0.270633	1.033007	0.442402
2011	2.410642	2.779944	2.498145	0.243944
2012	1.238957	1.239923	2.935736	0.052243
2013	2.413	2.669092	0.859945	0.216395
2014	2.023561	1.300229	2.528034	1.888145
2015	1.524709	1.506515	5.827459	2.903816
2016	0.112038	0.185236	2.190602	0.379717
2017	1.133707	0.940257	2.821362	0.685765
2018	0.130522	0.487913	2.076544	0.088343
2019	0.21751	0.111321	0	0.628454

Data Display

Below, I compare the loss ratio distributions of Morton, KS and Whitman, WA.

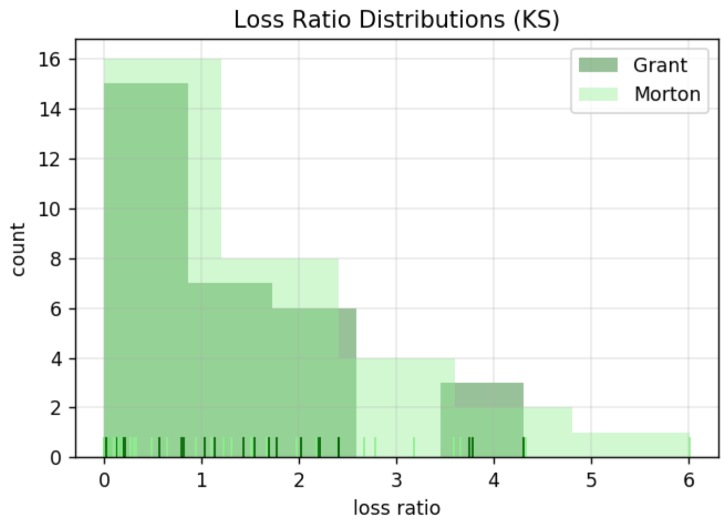


County	n	Mean	SD (S)	Min	Q_1	Median	Q_3	Max
Morton	31	1.42	1.50	0.01	0.23	1.11	2.08	6.01
Whitman	31	0.72	0.79	0.03	0.23	0.44	0.80	3.16

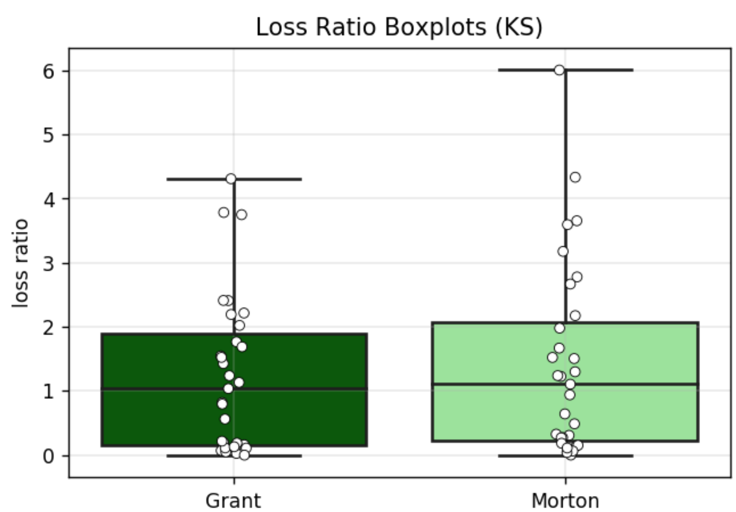


By the boxplot, table, and histogram above, it is evident that both Morton's and Whitman's loss ratios are more concentrated on the lower end; they even have the same first quartile value of 0.23. However, their distributions are vastly different. Whitman's data is extremely skewed to the left, while Morton's loss ratios decrease much more gradually. Morton seems to suffer relatively higher loss ratios, as displayed in the histogram. Their ranges are remarkably contrasting. Morton has a range of 6.00, whereas Whitman's is only 3.13, and as loss ratios increase, the third quartile and median values of Morton are much greater than those of Whitman. The boxplot, containing white dots (jittered for visibility) denoting individual years' loss ratios, supports that Morton's distribution is more widespread. Morton indicates clustering in higher loss ratios, while Whitman's data is only significantly clustered towards the bottom.

Next, I investigated two more counties that were in the same geographic location as Morton and Whitman: Grant, KS and Skagit, WA, respectively. I analyzed loss ratios at the county level between the pairs Grant and Morton in Kansas and Skagit and Whitman in Washington to discover whether they were similar or not. It is reasonable to predict that they would be, since they are located directly next to one another. Below, I display loss ratio distributions for counties in Kansas.

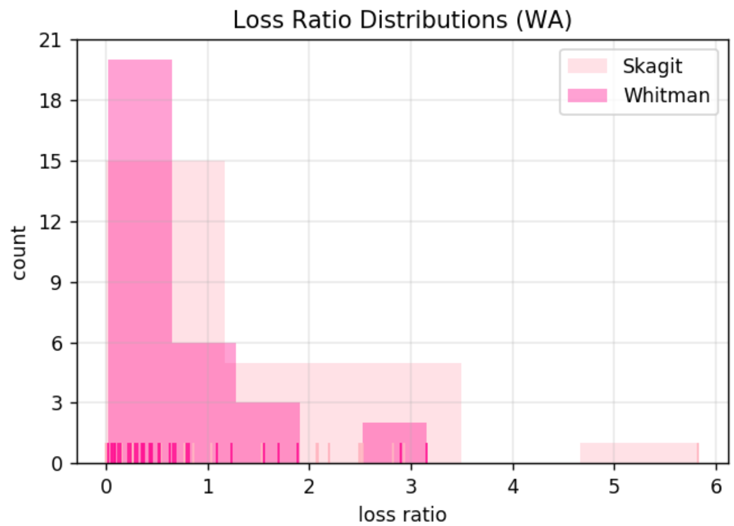


County	*n*	Mean	SD (*S*)	Min	Q_1	Median	Q_3	Max
Grant	31	1.23	1.21	0.00	0.16	1.04	1.90	4.31
Morton	31	1.42	1.50	0.01	0.23	1.11	2.08	6.01

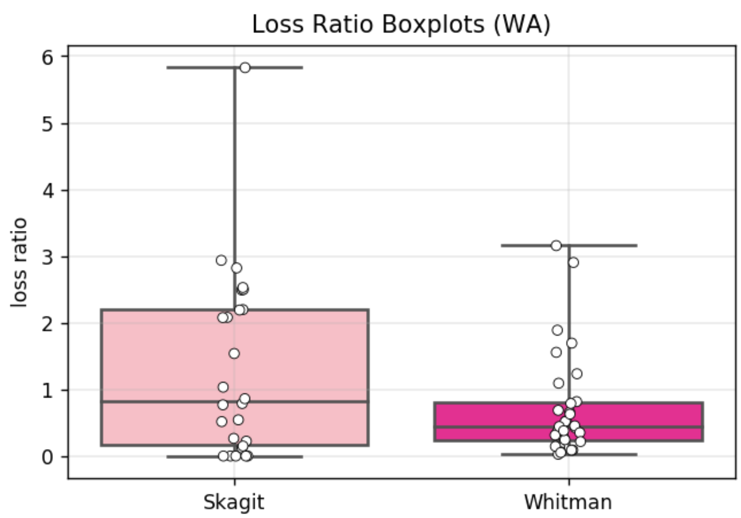


From the diagrams above, it is evident that most loss ratios in Kansas are heavily concentrated between 0 and 3, although there are also a few loss ratios above 3 present for both counties. Gradually, for both Grant and Morton, the frequency of the loss ratios constantly decreases over increasing loss ratio, as shown in the histogram. Other than the high outlier in Morton, the counties have very similar ranges, quartile and median values, and averages; each of these calculations reveal less than 0.2 of a loss ratio difference. The counties exhibit similar spread statistics, indicating there is minimal variation between them.

Next, I display loss ratio distributions for counties in Washington.



County	n	Mean	SD (S)	Min	Q_1	Median	Q_3	Max
Skagit	26	1.32	1.39	0.00	0.17	0.82	2.20	5.83
Whitman	31	0.72	0.79	0.03	0.23	0.44	0.80	3.16



The Washington results are surprising. As opposed to the Kansas counties' distributions, the Washington counties have much different clustering. Whitman data is centered around a low loss ratio, between 0 and 1, and has a range of just 3.13, compared to the Skagit range of 5.83. The averages of the counties also differ by 0.5. Overall, Skagit has a wider distribution and a greater variability than Whitman. These counties, unlike the Kansas ones, have significantly different ranges, quartile and median values, and averages, and they do not follow similar trends.

Data Analysis

To analyze if the observations and hypotheses made above are valid, I performed significance tests, which determine how significantly different datasets are. The condition of randomness is satisfied as the counties associated with Kansas and Washington were randomly chosen. While the data is not normally distributed, the dataset used contains enough entries to conduct t-tests validly. The data entries are independent, in that each year's loss ratio values for each county were recorded separately.

After setting the significance level α equal to 0.05, I first ran a two-sample, two-tailed t-test on the Morton and Whitman datasets, as proposed previously. The calculation is as follows:

$$t = \frac{\mu_{KS} - \mu_{WA}}{\sqrt{\frac{{S_p}^2}{n_{KS}} + \frac{{S_p}^2}{n_{WA}}}}, \ {S_p}^2 = \frac{(n_{KS} - 1)S_{KS} + (n_{WA} - 1)S_{WA}}{n_{KS} + n_{WA} - 2}$$

$${S_p}^2 = \frac{30(1.50)^2 + 30(0.79)^2}{31 + 31 - 2} = 1.44 \ \rightarrow \ t = \frac{1.42 - 0.72}{\sqrt{1.44\left(\frac{2}{31}\right)}} = 2.309286 \ \rightarrow \ p \approx 0.024387$$

Now I perform a similar two-sample t-test with Grant, KS and Morton, KS to confirm possible similarities between them:

$${S_p}^2 = \frac{30(1.21)^2 + 30(1.50)^2}{31 + 31 - 2} = 1.86 \ \rightarrow \ t = \frac{1.42 - 1.23}{\sqrt{1.86\left(\frac{2}{31}\right)}} = 0.551857 \ \rightarrow \ p \approx 0.583096$$

The final two-sample t-test involves Skagit, WA and Whitman, WA:

$$S_p{}^2 = \frac{30(1.39)^2 + 30(0.79)^2}{31 + 31 - 2} = 1.28 \rightarrow t = \frac{1.32 - 0.72}{\sqrt{1.28\left(\frac{2}{31}\right)}} = 2.078918 \rightarrow p \approx 0.043402$$

Below lies a table of all experiments, t-test statistics, and *p*-values.

	t-Test Statistic	***p*-Value**
Morton, KS; Whitman, WA	2.309285776095369 ≈ 2.31	0.02438716257658269 ≈ 0.02
Grant, KS; Morton, KS	0.5518573687360557 ≈ 0.55	0.5830957847768103 ≈ 0.58
Skagit, WA; Whitman, WA	2.0789179471956665 ≈ 2.08	0.04230190046540257 ≈ 0.04

Conclusion

The t-test with the average Morton and Whitman datasets produced a *p*-value of 0.0244, so I reject the null hypothesis and determine that these counties have statistically significant loss ratio differences, probably due to the fact that wheat crop yields vary under unique environmental circumstances. I then confirmed these results and evaluated whether I could uncover more information about loss ratio consistency among counties. Since the t-test between the Kansas counties of Grant and Morton yielded a *p*-value of 0.5831, I fail to reject the null hypothesis and conclude that their loss ratios are not statistically different. On the other hand, the t-test between the Washington counties of Skagit and Whitman yielded a *p*-value of 0.0423, which means that I reject the null hypothesis and determine that their loss ratio values are statistically different. This implies that they have more variance and less consistency than their counterparts in Kansas. Given that loss ratio is the ratio of indemnity to total premium, these results suggest that insurance companies associated with Grant and Morton experienced similar balances between indemnity and total premium values. It is also possible that the environmental conditions in Grant and Morton were similar, but that those in Skagit and Whitman in Washington were not.

It is evident that Washington and Kansas, and even counties within them in the Washington case, require distinct insurance plans, recommendations, and insurance policies. Therefore, the results advise to insurance companies that a single policy could be applied in the southwestern region of Kansas, whereas more specific, granular policies must be developed for counties in Washington. By creating a way to specify loss ratios based on location, insurance companies will be able to better tailor their policies to various locations. Insurance companies will be able to utilize these predictions to improve their decisions, premium pricing rates, and more.

Reflection on Process

I believe this project went well. I effectively tested and proved that nationwide insurance policies would not be the best fit for multiple locations, and that different areas have varying degrees of consistency in loss ratio. These findings are important, because the agricultural industry plays an integral role in the world. I developed an experimental design that was effective in carrying out and evaluating my hypotheses. It is interesting that insurance companies must tailor their policies to specific regions and counties.

One possibility for further enhancement is working with monthly data, which would increase the amount of data by a factor of twelve in addition to reducing potential variance. Unfortunately, loss ratio data via the Report Generator application is currently not available at the monthly level. As I continue to build on this project, I would like to test several more counties around other regions of Kansas and Washington and see if the results are consistent with the presented ones within this report, as well as experiment with various other wheat-growing states. This would allow me to expand the goals and findings of the study.

Intrigued by this topic, I conducted further analyses on how climate change impacts wheat production in the United States, and the resulting effects insurance companies face over time. A link to this work is given here.