How predictive are RIWI's China employment data of key headline economic indicators in China?

By Mark Kruger*

January 17, 2020

Summary: China's headline Manufacturing Purchasing Manager Index ("PMI") and the employment sub-indexes of the PMIs can be well-modeled using RIWI employment and wage data in China. The RIWI data appear to capture Chinese labour market developments as measured by the employment sub-indexes of the PMI survey. The RIWI data are able to out-perform both a naïve random walk and the Reuters' survey of analysts in predicting the headline PMI. Even if the RIWI data are limited to survey information collected in the first half of the month, 2 weeks before the official PMI is released, its predictions are still able to match those of the economists surveyed by Reuters. The RIWI data give researchers access to information which is unavailable from the PMI employment sub-components: labour market strength by disaggregated employment sector. This information could be useful in gauging the relative strength of economic activity in these sectors.

*Mark Kruger is a Shanghai-based consultant on economic and financial issues related to China. He is also a Senior Fellow at both the University of Alberta's China Institute and the Centre for International Governance Innovation. He was formerly a Senior Director in the Bank of Canada's International Department, a Senior Advisor to the Canadian Executive Director at the International Monetary Fund, and head of the Economic and Financial Section in the Canadian Embassy in Beijing. Mark most recently co-authored *Can Media and Text Analytics Provide Insights into Labour Market Conditions in China?* (International Journal of Forecasting Volume 35, Issue 3, July–September 2019, Pages 1118-1130). Mark can be reached at mark@markkrugerconsultantservices.com.

Introduction

To assess RIWI's China employment data, as reported to RIWI by randomly engaged Web users across China, we see how well RIWI's data performs in models vis-a-vis existing Chinese economic measures. If RIWI data were better at predicting wage movements than the existing data, for example, then RIWI data could be relied on even when their movements differed from those of existing measures. Previous work shows that the RIWI aggregate employment data are consistent with existing official measures, and that this correlation is broad-based and not driven only by one or two observations.[1] This work builds upon that analysis.

Can RIWI data predict PMI sub-indexes?

We start by examining how well RIWI data perform in predicting the employment sub-components of the official purchasing managers index ("PMI"). The PMI is a Chinese government survey of business conditions in the Chinese manufacturing sector that comes out ahead of official economic data. It is watched closely in financial markets as an early signal to extrapolate China's overall growth trend. PMI is usually published several days before the official activity growth indicators (e.g. industrial production, retail sales, and fixed asset investment). It is often seen as the leading indicator on China's overall growth.

Data

The PMIs are summary indices built from the Chinese National Bureau of Statistics (NBS's) monthly surveys of 3000 manufacturing and 4000 non-manufacturing firms. These are not surveys of opinion or sentiment. Rather, the questions asked pertain to objective data such as changes in production, employment, orders, prices and inventories. The indices are designed such that a reading above 50 indicates an improvement in economic conditions, while a reading below 50 indicates a deterioration. The farther the index reading from 50, in either direction, gives a sense of the breadth of the improvement or deterioration. The PMIs and their sub-components are published monthly, on the last day of the month to which they refer.

[1] Kruger, Mark. Assessment of RIWI's China employment data. January 3, 2020.

RIWI collected the data for this analysis using the company's machine-learning technology that exposed respondents representative of the Web-using population in China randomly to a range of questions on economic behavior. The specific questions used in this analysis were on employment status, employment sector, and wages questions. RIWI's randomized respondent-engagement approach gathers broad-based sentiment from across all of China, including urban and rural regions, and from across age groups. RIWI respondents remain anonymous and RIWI does not collect, process, store or transfer personally identifiable data. RIWI respondents are not incentivized to participate in RIWI surveys in any way and so respondent counts vary by question.

RIWI aggregated its survey respondents into nine employment categories:

1. Agriculture
2. Manufacturing
3. Technology/engineering
4. Healthcare
5. Education
6. Financial services
7. Construction
8. Government and
9. Other

Within each of these employment categories it distributed the respondents into five income categories:

1. Less than 1500 RMB/month
2. Between 1500 and 3000 RMB/month
3. Between 3001 and 5000 RMB/month
4. Between 5001 and 8000 RMB/month and
5. More than 8000 RMB/month

Modeling approach

To model the manufacturing employment sub-index, we relied on the share of the respondents in the manufacturing sector that reported that they were working full time. Working full and part-time were the only employment responses collected at the sectoral level. No data were collected on those who wanted to work in a sector, but who were unemployed. We also relied on the share of workers who reported that they were in the second- and third-income quintiles – i.e. those that were earning 3001-8000 RMB/month. This income range likely covers the vast majority of manufacturing workers in China. In 2018, the last year for which we currently have data, the average monthly wage in manufacturing was just over 6,000 RMB/month.

To model the non-manufacturing employment sub-index, we chose to concentrate on three employment sectors: technology/engineering, financial services and construction. We made this choice so as to focus on the firms most likely surveyed by the NBS. While there are private health care and education providers in China, the vast majority of these services are provided by the State. We similarly ignored all those respondents who reported working for the government and those who were in the other category. As with the manufacturing sector, here too we concentrated on those earning 3001-8000 RMB/month. This wage distribution is at the low end of those in technology/engineering and financial services but about right for construction.

Simple models for each of the two sectors were estimated using Ordinary Least Squares and the fitted values (predictions) were then compared to the actual series.

The data ran from July 2018 to December 2019, with missing observations for December 2018-February 2019, 15 observations in total, with a total of 48,778 unique respondents reporting their employment status, 46,821 their wage, and 27,386 their industry (only respondents who indicated they are currently working were shown this question).

Results

The results for the manufacturing sector are shown in Table 1. The fitted values have a correlation of 0.88 with the actual series. Both the full-time employment share and the 3001-8000 RMB/month wage share are highly significant at the 1% level.

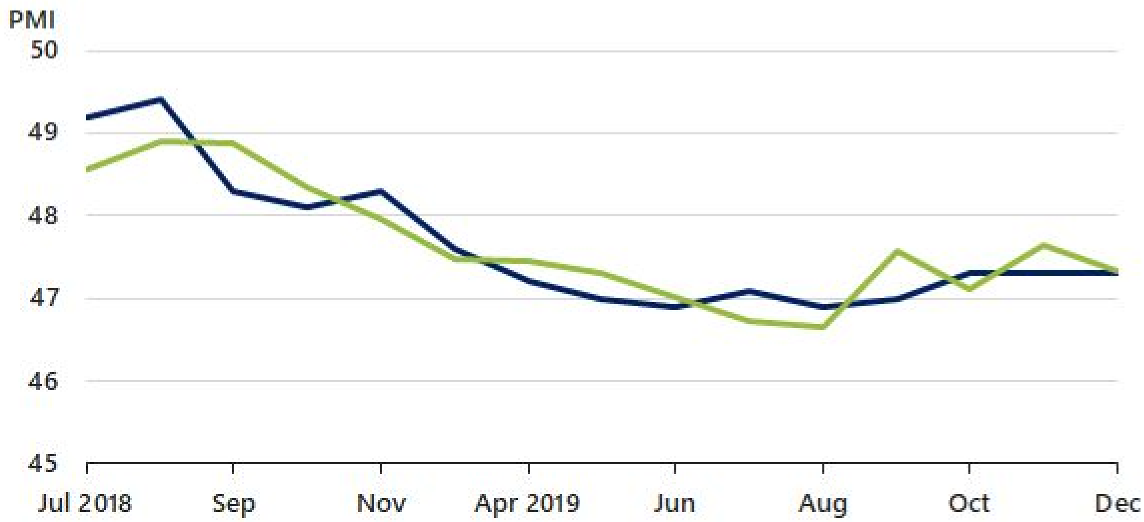
The model for the non-manufacturing sector, which is summarized in Table 2, does not work quite as well, perhaps because we fail to capture the distribution of non-manufacturing firms surveyed by the NBS. Nevertheless, the fitted values are fairly well correlated with the actual series, having a correlation coefficient of 0.67. Both the full-time employment share and the 3001-8000 RMB/month wage share are significant at the 10% level, which we believe is acceptable, given the short time series we are working with.

Graphs of both models and the actual employment sub-indices are presented below in Charts 1 and 2.

Chart 1: Purchasing Managers' Manufacturing Employment Sub-Index

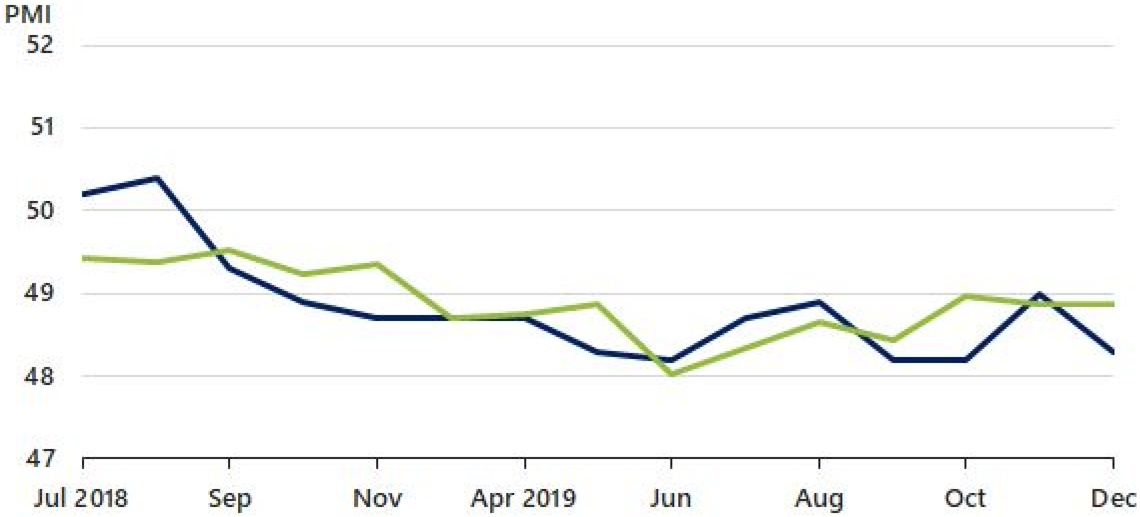
RIWI predicted PMI | Official PMI

Correlation: 0.88



Source: National Bureau of Statistics of China & RIWI data (China tracking, July 1-November 30, 2018 & March 1-December 31, 2019). Analysis by Mark Kruger.

Chart 2: Purchasing Managers' Non-Manufacturing Employment Sub-Index
RIWI predicted PMI | Official PMI
Correlation: 0.67



Source: National Bureau of Statistics of China & RIWI data (China tracking, July 1-November 30, 2018 & March 1-December 31, 2019). Analysis by Mark Kruger.

Can RIWI predict headline PMI?

Next, we assess the relationship between RIWI's survey data and the headline Manufacturing PMI published by the NBS.

Modeling approach

To model the manufacturing PMI, we relied on the share of the respondents working in the manufacturing sector that reported that they were working full time. Working full- and part-time were the only employment responses collected at the sectoral level. No data were collected on those who wanted to work in a sector, but who were unemployed.

We also relied on the share of workers who reported that they were in the second- and third-income quintiles – i.e. those that were earning 3001-8000 RMB/month. This income range likely covers the vast majority of manufacturing workers in China. In 2018, the last year for which we currently have data, the average monthly wage in manufacturing was just over 6,000 RMB/month.

A simple model was estimated using Ordinary Least Squares. The model's fitted values (predictions) were then compared to the actual series, a random walk series – where the current value of the PMI is assumed to be equal to the previous month's value – and to a survey of analysts predictions compiled by Reuters. We also computed fitted values for data collected over the first half of each month. The full-month model was not re-estimated when applied to the half month data to avoid over-fitting. Rather, the coefficients from the full-month model were applied to the partial-month data.

The data ran from July 2018 to December 2019, with missing observations for December 2018-February 2019, 15 observations in total, for a total of 48,778 unique observations for employment status, 46,821 for wage, and 27,386 for industry (only respondents who indicated they are currently working were shown this question).

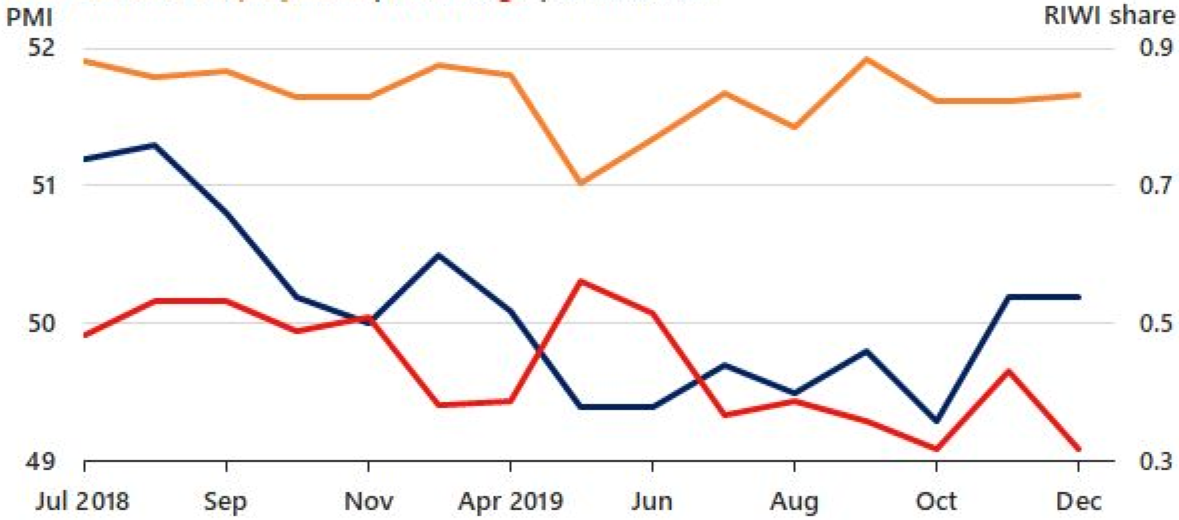
Results

The model for the headline manufacturing PMI is shown in Table 3. The fitted values have a correlation of 0.87 with the actual series. The coefficients on both the full-time employment share and the 3001-8000 RMB/month wage share are significant at the 1% level. This model benefits from the fact that the two component series are both well correlated with the PMI, but poorly correlated with each other, suggesting that they are picking up different types of information.

The actual series (blue) and the two components (orange and red) are shown in Chart 3 below. The actual series is graphed against the model (green) in Chart 4.

Chart 3: **Headline Manufacturing Purchasing Managers' Index**

RIWI full time employment | RIWI wage | Official PMI

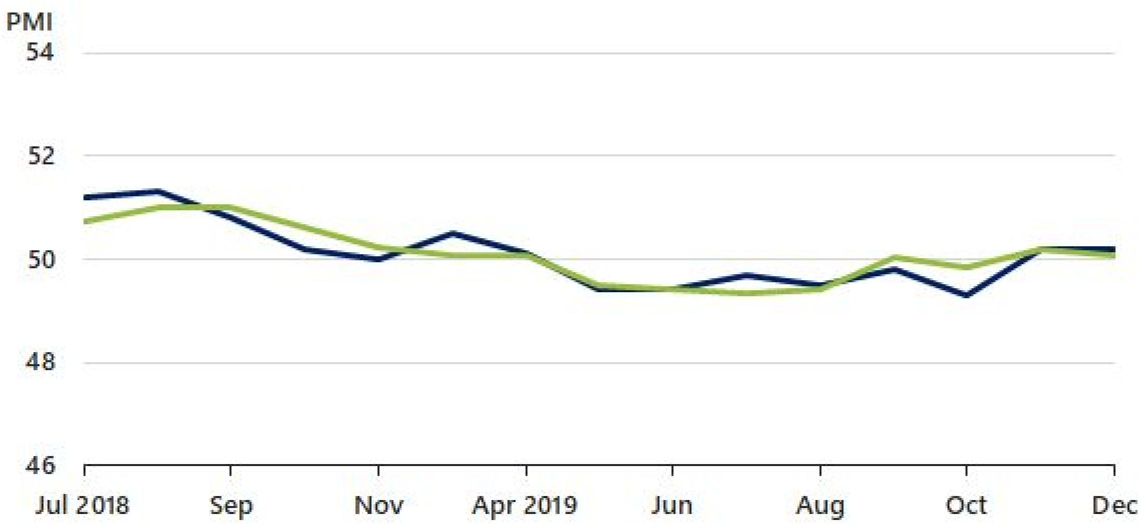


Source: National Bureau of Statistics of China & RIWI data (China tracking, July 1-November 30, 2018 & March 1-December 31, 2019). Analysis by Mark Kruger.

Chart 4: **Headline Manufacturing Purchasing Managers' Index**

RIWI predicted PMI | Official PMI

Correlation: 0.88



Source: National Bureau of Statistics of China & RIWI data (China tracking, July 1-November 30, 2018 & March 1-December 31, 2019). Analysis by Mark Kruger.

The actual PMI series, the fitted values for both the full- and half- month models are shown in Table 4, along with the random walk and Reuters analysts' forecasts. The full month model has the highest correlation with the actual series. The correlation with the predicted values for the half-month model are somewhat lower but still higher than those of the random walk and the Reuters analysts.

Table 4 also shows the monthly errors for each series with respect to the actual series. The mean absolute and mean squared errors are lowest for the full- and half-month series, respectively. The half-month series show a slight negative bias, indicating that it tends to under-predict.

We undertake a series of tests of means to determine if the lower errors made by the full- and half-month models are statistically different from those made by the random walk and the Reuters analysts. These tests are reported in Table 5.

The first sub-table shows that the difference in the mean absolute error made by the full-month model and the random walk is statistically significant at the 4% level (i.e. we can be 96% sure the errors are not the same). The second sub-table shows that the difference in the mean absolute error made by the full-month model and the Reuters analysts is statistically significant at the 8% level. The third sub-table shows that while the mean absolute error made by the half-month model is smaller than that made by the Reuters analysts, the difference is not statistically significant (i.e. the P-value exceeds the 10% critical value). Nevertheless, we can confidently say that even the half-month model performs no worse than the Reuters analysts!

Caveats

The time series that we are working with is short, so there is a possibility that the correlations we have will not be robust over time. Mitigating this risk is a fairly simple modeling strategy that does not overly rely on data mining. This analysis was conducted using the unweighted data. These are the data as reported to RIWI, not reweighted according to age and gender weights of the general population. (It is not clear if the correlations will hold up if the weighted data are used.)

Summary

To sum up, the headline manufacturing PMI can be well-modelled using the RIWI data. The RIWI data are able to out-perform a naïve random walk in predicting the PMI. The RIWI data are also able to out-perform Reuters' survey of analysts in predicting the PMI. Even if the RIWI data are limited to survey information collected in the first half of the month, its predictions are still able to match those of the economists surveyed by Reuters. A simple model using RIWI data is also correlated at a statistically significant level with the PMI’s employment sub-indexes.

The RIWI data also give researchers access to information which is unavailable from the PMI employment sub-components – labour market strength by disaggregated employment sector. This information could be useful in gauging the relative strength of economic activity in these sectors.

Appendix

Table 1: Manufacturing PMI Sub-Index, Summary Output

Regression Statistics	
Multiple R	0.882357
R Square	0.778553
Adjusted R Square	0.741646
Standard Error	0.415517
Observations	15

ANOVA					
	df	SS	MS	F	Significance F
Regression	2	7.284145	3.642072	21.09455	0.000118
Residual	12	2.071855	0.172655		
Total	14	9.356			

	Co-efficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	34.84423	2.343574	14.86799	4.3E-09	29.73802	39.95044	29.73802	39.95044
Full time	9.696179	2.656849	3.649504	0.00333	3.907404	15.48495	3.907404	15.48495
Wage share	10.42787	2.014446	5.176546	0.00023	6.038773	14.81698	6.038773	14.81698

Table 2: Non-Manufacturing PMI Employment Sub- Index, Summary Output

Regression Statistics	
Multiple R	0.665053
R Square	0.442296
Adjusted R Square	0.349345
Standard Error	0.532854
Observations	15

ANOVA					
	df	SS	MS	F	Significance F
Regression	2	2.70213248	1.35106624	4.758391292	0.03009
Residual	12	3.40720085	0.283933405		
Total	14	6.10933333			

	Co-efficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	36.88555	5.05381034	7.298561522	9.49536E-06	25.87424	47.89685	25.87424	47.89685
Full time	11.56894	6.13040918	1.887139917	0.083561355	-1.78807	24.92595	-1.78807	24.92595
Wage share	6.410122	3.10873095	2.061973997	0.061555689	-0.36322	13.18347	-0.36322	13.18347

Table 3: Manufacturing PMI Model, Summary Output

Regression Statistics	
Multiple R	0.87276
R Square	0.76171
Adjusted R Square	0.722
Standard Error	0.33255
Observations	15

ANOVA

	df	SS	MS	F	Significance F
Regression	2	4.242223062	2.121112	19.17952	0.000183
Residual	12	1.327110271	0.110593		
Total	14	5.569333333			

	Co-efficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	39.0518	1.87565312	20.82036	8.72E-11	34.96507	43.13847	34.96507	43.13847
Full time	9.94569	2.126378528	4.677291	0.000535	5.312711	14.57867	5.312711	14.57867
Wage share	6.14537	1.612239168	3.811696	0.002477	2.632598	9.658133	2.632598	9.658133

Table 4: Manufacturing PMI, Summary of Predictions and Errors

	Predictions					Errors			
	Actual PMI	Full-Month	Half-Month	Random Walk	Reuters Survey	Full-Month	Half-Month	Random Walk	Reuters Survey
Jul-18	51.2	50.72	50.79	51.5	51.3	-0.48	-0.41	0.3	0.1
Aug-18	51.3	50.99	50.86	51.2	51	-0.31	-0.44	-0.1	-0.3
Sep-18	50.8	51.02	50.96	51.3	51.2	0.22	0.16	0.5	0.4
Oct-18	50.2	50.63	50.3	50.8	50.6	0.43	0.1	0.6	0.4
Nov-18	50	50.22	50.42	50.2	50.2	0.22	0.42	0.2	0.2
Mar-19	50.5	50.08	50.11	49.2	49.5	-0.42	-0.39	-1.3	-1
Apr-19	50.1	50.07	50	50.5	50.5	-0.03	-0.1	0.4	0.4
May-19	49.4	49.51	49.5	50.1	49.9	0.11	0.1	0.7	0.5
Jun-19	49.4	49.43	49.88	49.4	49.5	0.03	0.48	0	0.1
Jul-19	49.7	49.34	49.62	49.4	49.6	-0.36	-0.08	-0.3	-0.1
Aug-19	49.5	49.41	49.26	49.7	49.7	-0.09	-0.24	0.2	0.2
Sep-19	49.8	50.04	50.07	49.5	49.5	0.24	0.27	-0.3	-0.3
Oct-19	49.3	49.87	49.18	49.8	49.8	0.57	-0.12	0.5	0.5
Nov-19	50.2	50.19	49.91	49.3	49.5	-0.01	-0.29	-0.9	-0.7
Dec-19	50.2	50.09	49.29	50.2	50.1	-0.11	-0.91	0	-0.1
Correlation	1	0.87	0.82	0.7	0.76				
Mean						0	-0.1	0.03	0.02
Mean absolute						0.24	0.3	0.42	0.35
Root mean squared						0.3	0.37	0.54	0.43

Table 5: Test for Equivalence of Mean Absolute Errors

t-Test: Paired Two Sample for Means		
	Full-Month	Random Walk
Mean	0.242581	0.42
Variance	0.031745	0.123143
Observations	15	15
Pearson Correlation	0.222092	
Hypothesized Mean Difference	0	
df	14	
t Stat	-1.92728	
P(T<=t) one-tail	0.037243	
t Critical one-tail	1.76131	
P(T<=t) two-tail	0.074485	
t Critical two-tail	2.144787	

t-Test: Paired Two Sample for Means		
	Full-Month	Reuters Survey
Mean	0.242581	0.353333
Variance	0.031745	0.064095
Observations	15	15
Pearson Correlation	0.135873	
Hypothesized Mean Difference	0	
df	14	
t Stat	-1.48368	
P(T<=t) one-tail	0.080028	
t Critical one-tail	1.76131	
P(T<=t) two-tail	0.160056	
t Critical two-tail	2.144787	

t-Test: Paired Two Sample for Means		
	Half-Month	Reuters Survey
Mean	0.300066	0.353333
Variance	0.049162	0.064095
Observations	15	15
Pearson Correlation	-0.27086	
Hypothesized Mean Difference	0	
df	14	
t Stat	-0.54429	
P(T<=t) one-tail	0.297402	
t Critical one-tail	1.76131	
P(T<=t) two-tail	0.594804	