Nguyen Tung Luong 20214913 – Applied statistics report (guided by prof. Nguyen Linh Giang)

I. Analysis of means (ANOM)

The promotion of the useful technique of analysis of means (ANOM) seems especially appropriate in this period of increased knowledge of the significance of statistical approaches for solving the economic problems of business. A good alternative to the Analysis of Variance (ANOVA) for comparing a set of treatments is the Analysis of Means (ANOM). The built-in simplicity of understanding and graphical display are appealing features of ANOM. An ANOM chart depicts decision lines, conceptually similar to a control chart, allowing magnitude differences and statistical significance of the treatments to be evaluated concurrently.

Prof. Ott created the ANOM technique in 1967 to compare a group of treatment means and determine if any of them significantly deviate from the group mean. Schilling (1973) expanded on this method and coined the term "analysis of means for treatments effects" (ANOME). The Analysis of Means approach is appropriate for components with fixed effects but is improper for factors with random effects, it is vital to mention. The model takes the mean of the factors to be constant for fixed effects. The factor level means, on the other hand, are random variables for random effects, and in such case the goal is estimation of variance rather than mean effects.

For our situation, a fixed effects model is appropriate; however, other problems may require different types of design (for instance random effect models, multiple factors), hence in this work, we only examine fixed effect models.

Balanced Design

Let us consider a single factor completely randomized design (CRD) with t treatment levels each with n observations. The corresponding fixed effects model can be written as

$$y_{ij} = \mu + \tau_{ij} + \varepsilon_{ij}\ , \text{ i=1,2,...,t, j=1,2,...,n,} \qquad (1)$$

where $N = nt$ is the total number of observations and $\tau_i = \mu - \mu_i$ is the fixed effect of the ith treatment level with respect to the overall mean μ. We also assume that the random errors - $\varepsilon_{ij} \overset{i.i.d}{\sim} N(0, \sigma^2)$.

For this model, observed treatment average $\overline{y}_{i\bullet} = \sum_{i=1}^{n} y_{ij}/n$ is an unbiased estimator for μ_i. Highly disperse $\overline{y}_{i.}$'s would indicate significant mean effects. On the other hand, if none of the effects is significantly different from the others, then the $\overline{y}_{i.}$'s must be close to the overall average $\overline{y}_{\bullet\bullet} = \sum_{i=1}^{t} \sum_{j=1}^{n} y_{ij}/N$ .In summary, the ANOM tests to see if at least one of the μ_i's is significantly different from the overall mean μ and graphically identifies exactly which treatment means cause the observed significance. That is, in ANOM we plot the $\overline{y}_{i\bullet}$'s along with $\overline{y}_{\bullet\bullet}$ in a decision chart similar in appearance to a control chart to visualize the mean effects.

Nelson (1981) indicates that the joint distribution of the absolute mean differences $|\bar{y}_{i\bullet} - \bar{y}_{\bullet\bullet}|$ becomes an equi-correlated t-dimensional *t-distribution* with correlation $\rho = -1/(t - 1)$ with degrees of freedom $N - t$. This distribution is also known as the Studentized maximum absolute deviation distribution or the h-distribution. Its probability calculations are reported in Nelson (1993) and Nelson et al. (2005).

For this single-factor model, the following upper and lower decision limits, denoted as UDL and LDL, respectively, will be used to make appropriate decisions about the mean effects:

$$UDL = \bar{y}_{\bullet\bullet} + h(\alpha; t, N-t)p(t-1)MSE/N, \qquad (2)$$

$$LDL = \bar{y}_{\bullet\bullet} - h(\alpha; t, N-t)p(t-1)MSE/N, \qquad (3)$$

where $MSE = \sum_{i=1}^{t} s_i^2/t$, s_i is the sample standard deviation of the *i*-th treatment level (or mean squared error), $N - t$ is the degrees of freedom for the MSE, and t is the number of means being compared. The critical value h(α, t, N − t) is the upper 100α percentage point of the *h*-distribution.

Unbalanced Design

For sets of means each of which is based on the sample size, equations (2) and (3) will give exact results. When the means are based on unequal sample sizes their deviations from the grand mean are no longer equicorrelated and decision limits, the following equations (4) and (5) should be computed using critical values that are upper bounds on the true but not available values of h . The decisions limits in this case are

$$UDL = \bar{y}_{\bullet\bullet} + m(\alpha; t, N-t)\sqrt{(N-n_i)MSE/(N * n_i)}\,, \qquad (4)$$

$$LDL = \bar{y}_{\bullet\bullet} - m(\alpha; t, N-t)\sqrt{(N-n_i)MSE/(N * n_i)}\,, \qquad (5)$$

where $N = \sum_{i=1}^{t} n_i$, $MSE = \sum_{i=1}^{t}(n_i - 1){s^2}_i/(N - t)$ and *m(α; t, N − t)* values are given in Nelson (1989) and Nelson et al. (2005). As one expects, the limits for each treatment will depend on their sample size (n_i).

Following Sidak (1967), the necessary values of *m(α; t, N − t)* can be calculated as the upper $\alpha^*/\ 2$ percentage points of a *t*-distribution, where

$$\alpha^* = 1 - (1-\alpha)^{1/t}\,, \qquad (6)$$

α : desired significance level

t : number of groups

The upper bounds obtained by the use of equation (6) are slightly less than the factors given by L.S.Nelson (1974). Remember that in the use of equation (5) it is necessary to calculate as many pairs of decision lines as there are different sample sizes.

Robust ANOM (using Rank)

The ANOM procedure is fairly robust to the departure of the normality assumption. However, Mendes and Yigit (2013) indicates that its statistical power depends on how well the homogeneity of the variance assumption is satisfied. Data transformation is preferred when data are non-normal and heterogeneous. In the cases where transformation is infeasible or inadequate, non-parametric methods such as rank-based tests may be preferred. Analysis of means using ranks (ANOMR) proposed by Bakir (1989) is a useful candidate for such data. The ANOMR procedure is a graphical alternative for the popular Kruskal-Wallis (K-W) test.

To explain the ANOMR procedure we reconsider the single fixed factor unbalanced design discussed in Section (2.2) and assume that t populations of interest have a similar shape and at most different in their location parameters. Like in the K-W test, we first assign ranks r_{ij} (i = 1,2,...,t, and j = 1,2,...,n_i) for all the responses y_{ij} ignoring their group memberships in the combined sample of size $N = \sum_{i=1}^{t} n_i$.

Let $\overline{R}_{i\bullet}$ denotes i-th treatment rank average, that is, $\overline{R}_{i\bullet} = \sum_{j=1}^{n_i} r_{ij}/n_i$ and the overall average of the ranks is $\overline{R}_{\bullet\bullet} = \sum_{i=1}^{t}\sum_{j=1}^{n_i} r_{ij}/N = (N+1)/2$.

Then, to test the null hypothesis that all t populations have exact same location parameter, the calculated rank averages $\overline{R}_{i\bullet}$ are plotted along with the following upper and lower decision limits to conduct the ANOMR test.

$$UDL = \overline{R}_{\bullet\bullet} + C(\alpha, t, n_1, n_2, \dots, n_t), \quad (6)$$

$$LDL = \overline{R}_{\bullet\bullet} - C(\alpha, t, n_1, n_2, \dots, n_t), \quad (7)$$

where C(α,t,n_1,n_2,...,n_t) is a constant that satisfies $P\left(\max_{1\le i\le t} |\overline{R}_{i\bullet} - \overline{R}_{\bullet\bullet}| \ge C\right) = \alpha$ under the null hypothesis. For equal sample sizes, it is recommended to apply Bonferroni adjustment to obtain suitable decision limits using $C^{*}(\alpha, t;\, n) = w(\frac{\alpha}{2t}, n(t-1), n)/n - (N+1)/2$ in place of - C(α, t; n_1, n_2, ..., n_t), where $w(\frac{\alpha}{2t}, n(t-1), n)$ is the upper 100(α)% percentile point of the Wilcoxon rank sum statistic with sample sizes n(t − 1) and n. However, obtaining exact critical values C(α, t, n_1, n_2, ..., n_t) become computationally expensive for even relatively moderate t and n_i values. Bakir (1989) provided a limited set of exact critical values for a few specific significance levels.

Due to unavailability of exact critical values for moderate sample sizes, Bakir (1989) suggested using asymptotic procedures. Further, they clarified that the asymptotic joint distribution of the $|\overline{R}_{i\bullet} - \overline{R}_{\bullet\bullet}|$ is the same as that of the $\left|\overline{y}_{i\bullet} - \overline{y}_{\bullet\bullet}\right|$. Therefore, it is recommended using the ANOM decision limits given in Equations (2), (3), (4), and (5) by replacing the observed data y_{ij} by their ranks r_{ij}.

II. Approach and Implementation of ANOM

We will apply it on Financial Ratio of

- Group of stock split by field.
- A group of stock in one field.

We believe that we can find some significant difference in stock so that we can analyz it.

1. Data

Data was got from vnStock API which take data from SSI and TCBS two securities company and make it to a dataframe form so we can believe about the right of source data.
For the Group of stock each dataframe will contain ratio of all the company in a field

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1		AAA	GVR	PHR	DPR	BRR	RTB	HRC	NNG	DRG	NHH	TRC	DRI	SBR	RDP	HII	PLP	VTZ	DAG
2	marcap	4434	86400	6951	2659	2228	1574	1538	1177	1153	1152	865	641	587	438	398	398	387	332
3	price	11600	21600	51300	61200	19800	18882	50900	14429	7400	15800	25500	8757	7207	8920	5400	5680	9000	5500
4	mberOfDa	-1	-1	0	1	0	1	1	1	0	2	0	0	0	1	-1	1	1	-1
5	iceToEarni	36,5	26,1	8,5	10,4	16	8,5	159,8	43,1	39,1	10	10,6	11,9	16,8	63,5	-9,4	25,3	13,5	34,5
6	peg	-0,6	-1	1	-0,2	-29,7	-0,2	-3	-0,6	-0,6	0,4	-8,8	-0,4	-0,4	-0,8	0,1	-0,3	0,1	1,2
7	riceToBoo	0,8	1,7	2	1,1	1,6	1,1	2,8	0,9	0,8	0,9	0,5	1,1	0,7	1	0,5	0,4	1,4	0,5
8	eBeforeEb	15,6	16,5	19,6	8,3	37,1	8,9	-395,2	9,8	7,2	6,2	10,3	3,9	214,9	9,1	-16,3	8,9	9,4	10,7
9	dividend	0	0	0,102	0,038	0	0	0	0	0	0,039	0,033	0	0	0	0	0	0	0
10	roe	0,023	0,067	0,241	0,112	0,103	0,128	0,017	0,022	0,019	0,114	0,044	0,086	0,043	0,015	-0,064	0,019	0,109	0,014
11	roa	0,011	0,042	0,133	0,064	0,091	0,07	0,012	0,014	0,011	0,054	0,036	0,066	0,035	0,003	-0,017	0,005	0,025	0,004
12	itOnIntere	1,6	4,4	3,7	1589,5		0,2	-0,3	5,6	0,6	4	0,9	11		1,1	3,9	1,7	1,8	1
13	rentPaym	1,7	2,7	3	8,6	1,4	3	0,3	1,3	0,6	1,3	1,9	2,2	2,5	1,2	1,2	1,2	1	1,1
14	ickPayme	1,4	2,2	2,8	7,7	1,1	2,1	0,3	1	0,4	0,9	1,7	1,4	2,1	0,6	1	0,9	0,5	0,5
15	sProfitMa	0,068	0,243	0,142	0,381	0,213	0,125	0,03	0,186	0,129	0,17	0,125	0,272	0,792	0,101	0,047	0,132	0,062	0,062
16	stTaxMarg	0,015	0,133	0,699	0,306	0,101	0,363	0,014	0,068	0,038	0,06	0,024	0,117	0,608	0,001	0,009	0,008	0,016	0
17	ebtOnEqui	0,5	0,1	0,1	0	0	0,1	0,4	0,3	0,2	0,5	0,1	0,2	0	1,9	0,7	1	2,5	1,8
18	ebtOnEbitc	8,1	1,4	1	0,1	0	1,1	-48,6	3,1	2	2,6	2,4	0,8	0	5,8	-12,5	6,2	5,3	7,8
19	ncome5yea	-0,073	0,02	0,222	0,031	0,098	-0,029	0,035	0,082		0,141	-0,123	-0,11				-0,085		-0,341
20	sale5year	0,303	0,049	0,007	0,025	0,017	0,169	0,011	0,065		0,218	0,052	-0,031		0,165	0,516	0,555		0,064
21	ome1quarter		-0,397	-0,456	0,048	-0,924	0,254	-0,932		0,941	-0,133	1,748	-0,482	-0,97	-0,007			0,627	0,047
22	ale1quarte	0,062	-0,541	-0,434	-0,529	-0,785	0,282	-0,651	-0,159	-0,336	-0,176	-0,351	-0,185	-0,986	-0,242	-0,113	-0,143	0,043	0,162
23	nextIncome	1,644	-0,027	-0,21	0,222	-0,088	0,032	0,312	0,668	0,164	0,367	0,043	-0,025	0,379	-0,787	-0,071	-0,503		0,491

Data will look like this, and each have about 20-25 ratio but we only need some important Ratio that can mostly affect the Short or Long term price of Stock. Luckily, we don't have to deal with None Value of Data because the data we got always contain data as it is the data that can be calculated from the financial statement and can't not be hidden or missing each quarter. And about what and why we use we will explain in next part.
For the group stock of each field each file contains a company. For example, below is AGR for AGR.

	Q1-2023	Q4-2022	Q3-2022	Q2-2022	Q1-2022	Q4-2021	Q3-2021	Q2-2021	Q1-2021	Q4-2020	Q3-2020	Q2-2020
ticker	AGR	AGR	AGR	AGR	AGR	AGR	AGR	AGR	AGR	AGR	AGR	AGR
quarter	Q1	Q4	Q3	Q2	Q1	Q4	Q3	Q2	Q1	Q4	Q3	Q2
year	2023	2022	2022	2022	2022	2021	2021	2021	2021	2020	2020	2020
priceToEarning	14.4	11.0	11.5	12.4	10.6	14.1	10.9	8.5	20.9	19.1	10.8	9.6
priceToBook	0.8	0.6	0.8	0.8	1.7	2.2	1.5	1.3	1.2	0.9	0.5	0.4
valueBeforeEbitda	19.4	14.2	8.4	8.5	7.1	11.9	21.5	15.2	19.5	17.6	19.4	9.0
roe	56	59	75	68	173	168	146	163	61	48	51	39
roa	49	53	0.07	65	156	153	135	154	56	44	47	37
daysReceivable	2518.0	2363.0	2086.0	2085.0	1883.0	1849.0	2009.0	2063.0	2102.0	2661.0	2796.0	2878.0
daysInventory												
daysPayable	14.0	0.0	1.0	7.0	71.0	76.0	81.0	-28.0	20.0	55.0	26.0	0.0
earningPerShare	650	679	840	746	1840	1779	1488	1617	560	449	474	362
bookValuePerShare	11792.0	11593.0	11613.0	11431.0	11644.0	11406.0	10856.0	10608.0	9584.0	9733.0	9467.0	9247.0
equityOnTotalAsset	897	0.89	0.94	967	884	897	0.92	959	0.92	917	0.92	924
equityOnLiability	8.7	8.1	15.6	29.2	7.6	8.7	11.6	23.3	11.5	11.0	11.5	12.1
currentPayment	9.4	8.8	17.0	32.7	8.6	9.8	11.8	24.1	10.6	10.0	10.3	10.7
quickPayment	9.4	8.8	17.0	32.7	8.6	9.8	11.8	24.1	10.6	10.0	10.3	10.7
epsChange	-43	-191	125	-594	34	195	-0.08	1.884	248	-53	308	241
ebitdaOnStock	841.0	799.0	1000.0	890.0	1337.0	1327.0	959.0	1112.0	711.0	585.0	612.0	469.0

	Q1-2020	Q4-2019	Q3-2019	Q2-2019	Q1-2019	Q4-2018	Q3-2018	Q2-2018	Q1-2018	Q4-2017	Q3-2017
ticker	AGR	AGR	AGR	AGR	AGR	AGR	AGR	AGR	AGR	AGR	AGR
quarter	Q1	Q4	Q3	Q2	Q1	Q4	Q3	Q2	Q1	Q4	Q3
year	2020	2019	2019	2019	2019	2018	2018	2018	2018	2017	2017
priceToEarning	7.5	11.2	13.2	10.4	11.7	10.7	22.6	23.1	49.0	20.5	15.2
priceToBook	0.2	0.4	0.5	0.4	0.5	0.4	0.5	0.4	0.6	0.7	0.7
valueBeforeEbitda	9.3	4.5	5.1	8.2	8.4	8.1	9.4	12.2	14.4	10.5	13.9
roe	33	36	38	42	0.04	37	21	18	13	37	47
roa	0.03	34	37	41	37	37	0.02	18	13	36	47
daysReceivable	3012.0										
daysInventory											
daysPayable	0.0	0.0	1.0	1.0	2.0	0.0	0.0	0.0	0.0	0.0	0.0
earningPerShare	292	321	343	374	349	319	175	149	106	294	377
bookValuePerShare	8956.0	9277.0	9413.0	9652.0	9110.0	8822.0	8699.0	8377.0	8351.0	8323.0	8312.0
equityOnTotalAsset	0.94	935	954	964	887	976	992	992	994	993	983
equityOnLiability	15.6	14.5	20.6	26.8	7.9	40.1	117.6	129.4	162.5	137.7	56.3
currentPayment	15.6	14.7	21.9	28.7	8.7	44.6	104.1	121.3	151.7	106.1	43.4
quickPayment	15.6	14.7	21.9	28.7	8.7	44.6	104.1	121.3	151.7	106.1	43.4
epsChange	-92	-64	-81	71	92	823	0.17	406	-639	-218	0.31
ebitdaOnStock	376.0	410.0	464.0	477.0	465.0	412.0	285.0	253.0	184.0	371.0	376.0

Data taken from each quarter some field might be missing but we don't use it because most of it doesn't affect much on stock price. Therefore, we only take some Ratio of each company.

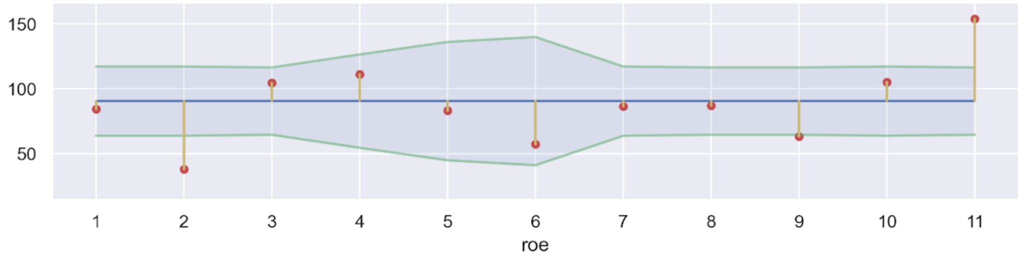
2. Group of fields.

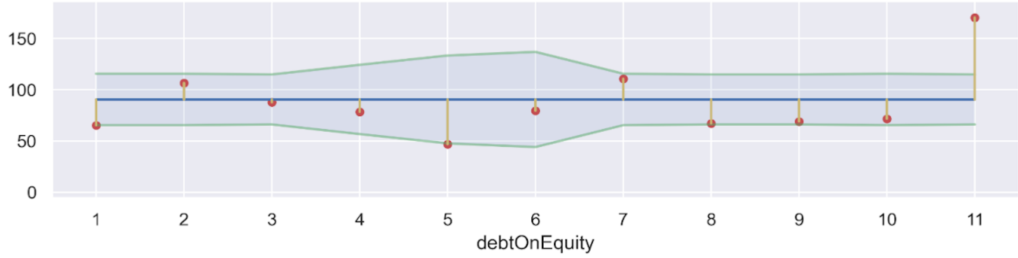
We have a group of fields that you can see by following our code. We take the ratio of stock in the first quarter of 2023 and then group it by field. We use 11 groups that have the highest liquidity on the Vietnamese Stock Market.

$\{1: 'real\_estate', 2: 'mineral', 3: 'production', 4: 'retail',$
$5: 'Technology', 6: 'tourism', 7: 'transportation',$
$8: 'chemistry', 9: 'security', 10: 'Drink\_food', 11: 'Bank\_ind'\}$

To apply ANOM we work on 5 ratios.

- Price To Earning (P/E) relates a company's share price to its earnings per share.
- Price To Book (P/B) measures the market's valuation of a company relative to its book value.
- Return on common equity (ROE) measure of a company's net income divided by its shareholders' equity.
- Debt On Equity (D/E) compares a company's total liabilities with its shareholder equity and can be used to assess the extent of its reliance on debt.





As we can see from P/E and P/B transportation and minerals had significant value. Transportation is out of the upper bound where minerals is under the lower bound for P/E and the lower bound for minerals at P/B. With high P/E maybe mean that company are overvalued or investor are expected on the grow of that company while the P/B have different mean that is based on the ratio between market price and book price but we can use as the same with P/E where if P/B small mean that company maybe are undervalued and that will be a good invest. Therefore, if we follow that rule minerals have high potential to invest in while transportation is not recommended. The data is taken from the first quarter of 2023 and luckily, we have time to see it after half of the year 2023 passes. Since it grows slowly in the first 4 months of 2023 it actually grows dramatically in June and July. We took HPG as an example because it has the highest market cap in the mineral group.

If we investigate in April when all the financial statements are done, we return about 40%. It is much higher than most stocks on transportation.

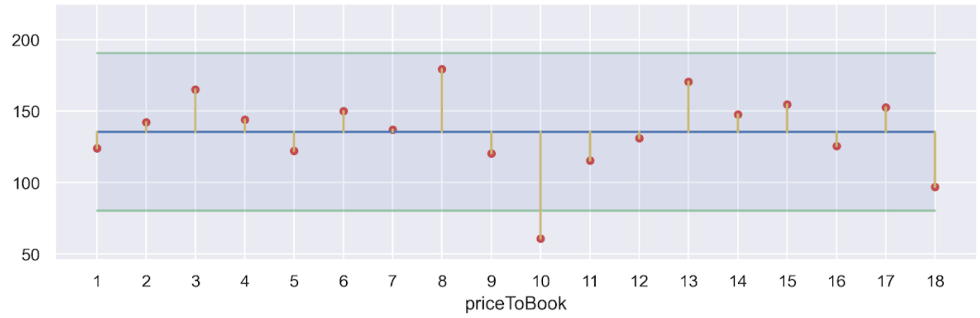
On ROE and D/E it is not used as much as P/E and P/B to find good stock, investors often use it to see how that company is using money. It is not a good use for comparing between different groups as each group has different D/E, or ROE, ROA. Why high [leverage is optimal for banks](#) can be found in this article to understand more why banks have high D/E.

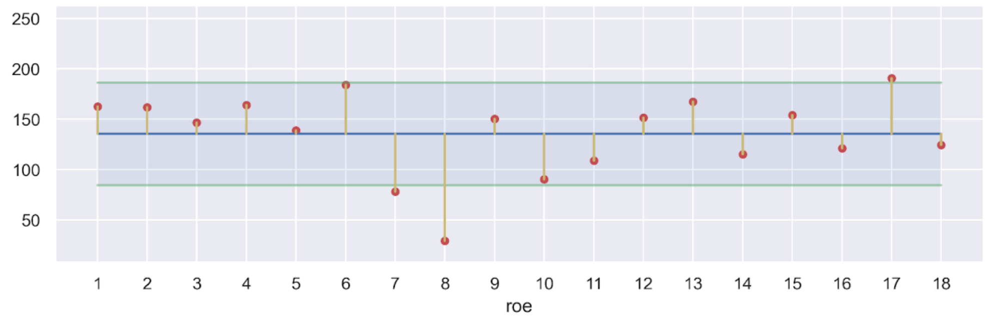
3. Group of stock

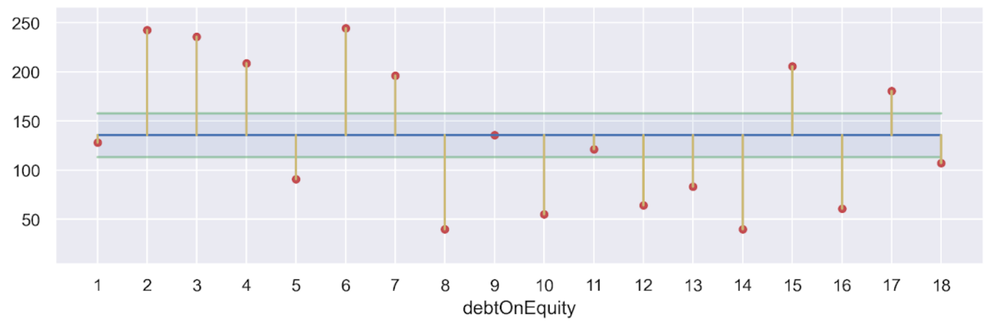
To see deeper we look into each stock, we will use the ratio on each quarter of a stock. We consider the security group which contains 19 companies with their ratio from each quarter that we could get.

$\{1: ratio\_SHS, 2: ratio\_TVS, 3: ratio\_SSI, 4: ratio\_MBS, 5: ratio\_TVB, 6: ratio\_VND,$
$7: ratio\_PHS, 8: ratio\_SBS, 9: ratio\_FTS, 10: ratio\_AGR, 11: ratio\_VDS, 12: ratio\_VIX,$

$13: ratio\_ORS, 14: ratio\_APG, 15: ratio\_HCM, 16: ratio\_EVS, 17: ratio\_VCI, 18: ratio\_TCI\}$







Similarly, we will look for the Significant value of P/E and P/B, that is PHS and AGR. Like the last analysis PHS will not be recommended while AGR will be the stock that might bring us good profit.

It actually does, if we invest when the financial statements often are in the middle or at the end of April are public, we could get 25% revenue. But it doesn't mean this is the best stock ROE can show us how they profit based on their money. All the stocks that have a higher ROE can be considered as making money better or in other words they are better off investigating long term.

With VCI and VND, which have a Higher ROE than others, they are very good to invest even though VND got some bad news in early July, but we still got good profit from it.

We can see that ANOM can be used on stock to detect its value between others. It can find significant Ratio between stock in a field or between some fields. You should combine ANOM with others to have a better view and to have better investigation it may need more knowledge in the Financial.

Investing or any other work is not easy. I recommend consulting financial news outlets, stock market analysis websites, or seeking advice from a qualified financial advisor who can provide you with up-to-date information and insights on the stock's performance.

References

[1] K. V. Narasimha Murthy, S. Ismail: "Classification and a Critical Comparison of Different Company Shares under Consideration through Analysis of Means Technique". Journal of Economics and Finance, pp.26- 32

[2] Kalanka P. Jayalath, Jacob Turner: "Analysis of Means (ANOM) Concepts and Computations" Volume 16, Issue 1, Article 5 (AAM)

[3] SCHILLING, E.G. (1973). "A Systematic Approach to the Analysis of Means". Journal of Quality Technology 5, pp.93-108, 147- 159.

[4] Nelson, L. (1983). Exact critical-values for use with the analysis of means, Journal of Quality Technology, Vol. 15, No. 1, pp. 40–44.

[5] Nelson, P.R., Wludyka, P.S. and Copeland, K.A.F. (2005). *The Analysis of Means: A Graphical Method for Comparing Means, Rates, and Proportions*, Vol. 18, Philadelphia, PA: Society for Industrial and Applied Mathematics.

[6]Nelson, L.S. (1974). "Factors for the Analysis of Means". Journal of Quality Technology 6, pp. 175-181

[7] Ott, E. R. (1967). Analysis of means – a graphical procedure, Industrial Quality Control, Vol. 24, No. 2, pp. 101–109.

[8] SIDAK, Z. (1967). “Rectangular Confidence Regions for the Means of Multivariate Normal Distributions”. Journal of the American Statistical Association 62, pp. 626-633.

[9] Bakir, S. T. (1989). Analysis of means using ranks, Communications in Statistics–Simulation and Computation, Vol. 18, No. 2, pp. 757–776.